3-2003

A practical procedure to estimate the quality and the adverse impact of single-stage selection decisions

Wilfried DE CORTE
Ghent University

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

DOI: https://doi.org/10.1111/1468-2389.00229

Citation

DE CORTE, Wilfried and LIEVENS, Filip. A practical procedure to estimate the quality and the adverse impact of single-stage selection decisions. (2003). *International Journal of Selection and Assessment*. 11, (1), 89-97. Research Collection Lee Kong Chian School Of Business.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/5638

Published in International Journal of Selection
and Assessment, 2003 March, Volume 11,
Issue 1, Pages 89-97

A Practical Procedure to Estimate the Quality and the Adverse Impact of Single-Stage Selection Decisions

Wilfried De Corte* and Filip Lievens

The quandary posed by the conflicting goals of valid selection and a diverse workforce is one of the most perplexing problems facing the practice of personnel selection today. To help address the issue, the article presents a comprehensive method and a related computer program to estimate the expected adverse impact and the expected quality of the majority, the minority and the total selected work force. Compared to previous related procedures, the present method is much more general as it can address situations with both multiple predictor and multiple criterion dimensions. In addition, the expected effects can be computed given the overall selection ratio and the estimates are derived analytically and, hence, are accurate. To assist the selection practitioner, the method is made available as a free download from the Internet.

Introduction

In the past few years, considerable attention has been given to the dilemma faced by employers who intend to obtain an optimally qualified workforce, but at the same time want to eliminate adverse impact (e.g., Campion, Outtz, Zedeck, Schmidt, Kehoe, Murphy and Guion 2001; Hough, Oswald and Ployhart 2001 and Sackett, Schmitt, Ellingson and Kabin 2001). The conflict arises because many valid selection procedures, and cognitive ability tests in particular, consistently show marked racial subgroup differences (e.g., Hough *et al.* 2001). These differences are typically in favour of the White applicant group (often referred to as the majority group) as compared to Black, Hispanic or immigrant applicant groups (e.g., te Nijenhuis and Van der Flier 1997), but in some cases White applicants underscore in cognitive ability in comparison to East Asian applicants (Hough *et al.* 2001). As a consequence of these differences, the optimal hiring policy of top-down selection usually leads to the phenomenon of adverse impact in that the selection rate in the White applicant group is substantially higher than that in the Black and Hispanic applicant groups.

To address the above described problem, that is henceforth referred to as the quality-adverse impact trade-off problem, several remedial procedures have been proposed. By and large, these procedures can be regrouped in two major categories. In the first category, as represented by the procedures of within group norming, separate cutoffs and fixed, sliding and criterion-related banding (Aguinis, Cortina and Goldberg 1998; Cascio, Outtz, Zedeck and Goldstein 1991), the optimal method of referral (i.e., top-down selection on the basis of the unadjusted predictor scores) is replaced by a method for selecting applicants that takes into account predicted performance as well as the racial composition of the total applicant group. However, within group norming and the usage of separate cutoffs involve some form of minority group preference that seems unacceptable to the courts (Sackett *et al.* 2001), whereas banding without such minority preferences within the bands does little to reduce adverse impact (cf. Campion *et al.* 2001, p. 167; Sackett and Wilk 1994). Thus, Schmidt concludes that 'the current stance of the courts appears to block achievement of what is perhaps the major objective of SED banding' (cf. Campion *et al.* 2001, p. 167); a conclusion that seems also applicable to the newer form of banding outlined by Aguinis *et al.* (1998).

The procedures of the second major category aim for a better balance between quality and workforce diversity through the development or usage of predictors or predictor composites that exhibit smaller mean differences between the applicant groups. Thus, alternate selection predictors with lower adverse impact levels such

---

* Address for correspondence: Wilfried De Corte, Faculty of Psychology, Ghent University, H. Dunantlaan 1, B-9000 Ghent, Belgium. E-mail: wilfried.decorte@rug.ac.be

as simulations, work sample tests and tests with a different mode of presenting the test stimuli (e.g., video-based testing) have been proposed. As observed by Sackett *et al.* (2001), the approach is often linked to the emerging recognition that the job performance criterion is multidimensional (cf. Murphy and Shiarella 1987) or hierarchically ordered (cf. Viswesvaran and Ones 2000) and that an appropriate weighting of the predictor and criterion components can reduce the extent of adverse impact. However, replacing adverse impact predictors by less offending but typically less valid ones or using different predictor and criterion weighing schemes may seriously affect the expected quality of the selected employees.

Given the practical importance of the quality-adverse impact trade-off problem and the absence of a generally adequate solution method, it is not surprising that several studies converged on this topic (e.g., De Corte 1999a; Hattrup, Rock and Scalia 1997; Sackett and Ellingson 1997 and Schmitt, Rogers, Chan, Sheppard and Jennings 1997). Yet, despite this interest no comprehensive, practical procedure to estimate both the expected quality and the expected adverse impact of an intended selection decision has thus far been made widely available (see the next section). To remove this deficiency, this study presents both such a procedure and an easily applicable computer program to implement the method. In addition, the article discusses a partial solution to the quality-adverse impact dilemma. The solution applies to situations in which several selection procedures, that differ in their level of adverse impact, are used and in which the overall selection rate can be chosen freely.

The next section introduces the proposed procedure. The nature of the selection decision problem that can be addressed by the procedure as well as the pursued objectives are discussed first. Next, the assumptional basis and the data that are required for the application of the method are detailed. To keep the exposition easily accessible for the reader, the derivation of the procedure is relegated to the Appendix and the exposition proceeds with an illustrative application. This example application highlights the distinctive features of the procedure and shows how the accompanying computer program can be invoked to obtain the desired results. The merits of the method, its limitations as well as its potential for further development are reviewed in the final section.

Estimation Method

Nature of the Selection Decision Problem under Investigation

The adverse impact issue arises when the applicants come from at least two different populations. Following the convential usage it is henceforth understood that the candidates belong to either the so-called majority or the minority applicant population and that these two populations have a different average score on the available selection predictor variables. In addition, and in accordance with the repeated plea to conceive the criterion behaviour (overall job performance) as multidimensional (Borman and Motowidlo 1997; Murphy and Shiarella 1997), it is also proposed that the two applicant groups may on average differ with respect to the relevant performance dimensions as well. In line with current practice, the average group differences with respect to both the predictor and the criterion variables will be expressed by the corresponding effect sizes which indicate the standardized mean score difference between the two groups (cf. Hough *et al.* 2001).

Apart from the heterogeneous nature of the applicant population, the selection situation is further characterized by the fact that the new employees will be selected in a single stage, even though several predictors may be available to perform the selection. Obviously, focusing on only single-stage selections limits the applicability of the procedure because many selections are hurdle-based in that only candidates who score above a threshold on the initial screen proceed to the subsequent stages of the selection process (Sackett and Roth 1996). The calculation of the expected quality and the expected adverse impact for these multistage scenarios poses computational problems, however. These problems have not been resolved yet and only preliminary and approximate results, based on Monte Carlo simulation methods, are available (cf. Sackett and Roth 1996). In fact, even for single-stage selections with a pre-specified overall selection rate, it is not well documented how the expected quality can be determined analytically when the applicant group is heterogeneous. So, by providing such a solution framework and an easily applicable tool for its implementation, the present method offers a distinct contribution, despite its limitation to single-stage selection scenarios.

Objectives of the Procedure

The present proposal intends to resolve three issues. First, given the weights with which the predictor and the criterion variables are combined to the predictor and the criterion composite, the method computes the expected adverse impact of the selection and the expected quality of the selected employees for varying levels of the overall selection rate. To express the expected quality three related measures will be used: the expected standard score on the composite criterion of the selected employees from the majority, the minority and the total applicant group. These measures will be referred to as $\bar{Z}_a$, $\bar{Z}_i$ and $\bar{Z}_t$, respectively. Alternatively, the expected adverse impact, which is equal to the ratio between the expected selection rate in the minority group, s_i, and the expected selection rate in the majority group, s_a, will be denoted as i.

Compared to previous related research, the present method enables the estimation of the expected adverse impact for a given *overall* selection ratio instead of for a given selection rate with respect to one of the applicant groups (e.g., the Tables 1 and 2 in Sackett and Ellingson 1997, pp. 710 and 712, respectively). From a practical perspective, this is a clear advantage because practitioners are typically confronted with overall selection rate decisions and not with decisions framed in terms of the selection rate in, for example, the majority group. For a similar reason (i.e., explicit reference to the overall selection ratio instead of to the predictor cutoff value), the present method improves also on the tabled information provided by Bartram (1995). As to the estimation of the expected criterion scores of the majority, minority and total applicant group selected, no accurate computation procedure is available in the literature. Indeed, the existing studies that address this or directly related issues (e.g., Hattrup *et al.* 1997; Roth, Bobko, Switzer and Dean 2001) all use Monte Carlo simulation methods and, hence, obtain only approximate results.

The second objective of the present procedure is to determine the lowest possible overall selection rate that is still consistent with the four-fifths rule of thumb used by the Equal Employment Opportunity Commission to determine a *prima facie* case of adverse impact (cf. Sackett and Wilk 1994). According to the rule a selection does not reflect disparate impact when the value of the corresponding adverse impact ratio is at least equal to 0.80. As the latter ratio is typically higher for higher overall selection rates, disparate impact can be prevented through the implementation of a sufficiently high overall selection rate. At present, only Bartram (1995) has provided some results on the selection rates that comply with the four-fifths rule of thumb, but his results focus on situations in which a single predictor is available to perform the selection. Also, because of the tabular presentation format, only a limited number of situations in terms of the magnitude of the predictor subgroup differences are dealt with. In contrast, the present method is much more general in that it can address situations with both multiple predictor and multiple criterion dimensions. Also, the procedure starts from data supplied by the user and it can therefore always be tuned to the specific needs of any particular application. Finally, observe that the determination of the overall selection ratio that corresponds to an expected adverse impact of 0.80 (denoted as the selection rate $s^{(0)}$) effectively deals with the qualityadverse impact issue in case that the selection rate can be chosen freely. In fact, as the average predictor and criterion composite scores of the minority group applicants will typically be lower than the corresponding averages for the majority group, all selection ratios above $s^{(0)}$ will also preclude adverse impact; whereas smaller values are associated with adverse impact as defined by the four-fifths rule.

The final objective relates to the study of Roth *et al.* (2001) who investigate the effect of prior selection on the assessment of standardized ethnic group differences with respect to a related variable. When a predictor is used to screen a heterogeneous population and the remaining applicants are subsequently compared on the related variable, the difference between the subgroup averages will be smaller than the corresponding difference in the original unselected populations. To gauge this downward bias, the authors use Monte Carlo simulation methods. As a consequence, the resulting bias estimates are not very precise. This shortcoming is easily corrected by the present method, however, because the procedure provides an analytical (and, hence, accurate) estimate of the expected criterion (cf. the second, so-called related variable in the Roth *et al.* study) score in both the selected minority and the selected majority group, resulting in an accurate assessment of the bias.

Assumptional Basis and Data Requirements

Similar to previous studies on the subject (e.g., Hattrup *et al.* 1997 and Sackett and Roth 1996) it will henceforth be assumed that the predictor variables and the criterion dimensions have a joint multivariate normal distribution, with the same variance/covariance matrix but a different mean vector, in both the minority and the majority applicant population. So, whereas we assume that the correlations between the predictor and the performance dimensions, as well as the predictor validities are the same in the two applicant populations, it is recognized that the groups have a different mean score on the predictor and the criterion aspects. For convenience, and without loss of generality, it is further assumed that the distribution of the predictors and the criterion dimensions is *standard* multivariate normal in the majority applicant group, implying that the predictor and the criterion dimensions have a mean score equal to zero in the majority population.

The first assumption implies that the predictor and the criterion composites that result from an arbitrary linear combination of the predictors and the criterion aspects, respectively, have a bivariate normal distribution in the two applicant populations. Also, from both assumptions it follows that the elements of the mean vector of the predictor and criterion variables in the minority applicant group correspond to the standardized mean score difference between the two groups and are, therefore, equal to the effect size of these predictor and criterion variables. Because of the above convention to equate the average predictor and criterion scores to zero in the majority applicant population, and as the latter average values are typically higher than the corresponding averages in the minority applicant group, the effect sizes will generally have a negative value. These effect sizes will henceforth be denoted as $d_{x_1} \ldots d_{x_p}$ for the available predictors $X_1 \ldots X_p$ and as $d_{Y_1} \ldots d_{Y_c}$ for the criterion dimensions $Y_1 \ldots Y_C$. Alternatively, the symbol U will be used to refer to an arbitrary weighted combination of the predictors; whereas the symbol Z will indicate an arbitrary weighted combination of the criterion aspects. Thus, $U = a_1X_1 + \ldots + a_pX_p$ and $Z = b_1Y_1 + \ldots + b_cY_c$ where the symbols $a_1 \ldots a_p$- and $b_1 \ldots b_c$ correspond to the weights with which the predictor variables and the criterion aspects are combined to the composite predictor U and the composite criterion Z, respectively.

Given the above assumptions and suitable data on the predictors and the target criteria as well as on the selection scenario characteristics (e.g., the overall selection ratio), it is shown in the Appendix how the minority/majority selection rates (s_i and s_a), the adverse impact (i) and the expected (composite) criterion score of a minority, majority or total applicant group selected applicant ($\bar{Z}_i\bar{Z}_a$,and $\bar{Z}_t$, respectively) can be calculated. The required data are (a) the predictor validities, intercorrelations and effect sizes; (b) the intercorrelations and effect sizes of the criterion aspects; and (c) the weights which are used to construct the predictor and criterion composites. As a final datum, the ratio of majority vs. minority applicants in the total applicant population must also be available.

The data demands are quite realistic as accurate estimates of many predictor and criterion variable characteristics can be obtained from the numerous meta-analytic studies on the validity of personnel selection procedures (e.g., Bobko, Roth and Potosky 1999; Hough *et al.* 2001 and Schmidt and Hunter 1998). These meta-analytic studies often use different correction orientations for estimating the intercorrelations and validities of the selection predictors, however. Either uncorrected correlations, correlations corrected for either range restriction or attenuation, or fully corrected correlations may be provided. For the present purposes, both corrected or uncorrected estimates are adequate, provided that the correlation/validity data used in any one particular computation all reflect the same correction orientation and, hence, are gauged in a common metric. Also, as the estimates, computed for the adverse impact and the average quality of the majority and minority selected applicants, critically depend on the value of the input correlations, it is necessary that these input values are as accurate as possible.

The choice of the criterion weights $b_1 \ldots b_c$ is also quite straightforward because these weights can primarily be 'determined on the basis of an analysis of the performance domain of interest and the values that an institution places on various criterion dimensions' (cf. Sackett *et al.* 2001, p. 306). As to the specification of the predictor weights, it is emphasized that the main purpose of the present procedure relates to its potential to estimate, easily and accurately, the effect of different weighing schemes on the resulting adverse impact and quality of the selected applicants. The method is therefore ideally suited to explore the relative merits of such alternative weighing scenarios; a feature that is further illustrated in the example application reported below.

Method

The details on the derivation of the expected adverse impact and the expected quality of any particular selection decision are outlined in the Appendix. To implement the calculations, a computer program was written and compiled to an executable source. This program, called CAIQS, can be run on a personal computer under the Windows 95/98, NT, XP and 2000 Professional operating systems. To execute the program an input file, that details the nature of the studied selection decision and summarizes the characteristics of the predictor and the criterion variables, must be specified. The output of the program reports the effect size of the predictor and criterion composite, the expected adverse impact and the expected, standardized (composite) criterion score of the majority, minority and total group selected candidates for varying levels of the overall selection ratio. As indicated above, the program computes also the minimum value of the selection rate that is still consistent with the four-fifths rule of no adverse impact.

The computer program, as well as a document that describes the preparation of the input file and the actual usage of the program can freely be downloaded from the Internet at http://www.allserv.rug.ac.be/˜wdecorte/software.html. The documentation contains also an example application and further details on the generated output. One particular aspect of this output deserves additional comment. The aspect relates to the way in which the expected criterion score is standardized. Actually two, differently standardized, estimates are reported for the expected (composite) criterion score of the majority, minority and total group selected applicants. In the first standardization, the expected scores are standardized relative to the distribution of the composite criterion in the majority applicant group. In what follows, $\bar{Z}_i$ $\bar{Z}_a$, and $\bar{Z}_t$ the earlier introduced notation is maintained for the thus standardized expected (composite) criterion score of the majority, the minority and the total group selected, respectively.

In the second standardization, the expected criterion scores are standardized relative to the distribution of the composite criterion in the *total* applicant group and the resulting estimates are henceforth reported under the labels, $\bar{Z}_i^{(g)}$ $\bar{Z}_a^{(g)}$, and $\bar{Z}_t^{(g)}$ for the expected (composite) criterion score of the majority, the minority and the total group selected, respectively. This second standardization is introduced to obtain values that can be compared to those that are obtained when Monte Carlo simulation methods are invoked to estimate the expected quality of the selected work force (cf. Hattrup *et al.* 1997).

Example Applications

This section discusses two example applications of the above presented method as implemented in the CAIQS program. The main purpose of the examples is to demonstrate the instrumental value of the method and program to study (a) the impact of predictor weighing schemes on the expected quality and adverse impact of a selection and (b) the effect of prior selection, using a first predictor, on the standardized ethnic group differences with respect to a second, related variable. As both issues have been addressed elsewhere (cf. Schmitt *et al.* 1997 and Roth *et al.* 2001), the examples do not focus on an incremental substantial contribution. Instead, they first and foremost serve to demonstrate that the present method enables a more versatile and more accurate analysis of these issues.

Impact of Predictor Weighing

To illustrate the potential of the present procedure to study the impact of different predictor weighing, it is applied to a typical selection decision problem. The example considers the situation where four predictors are available to select a pre-specified proportion of selectees from a heterogeneous applicant group in which 75% of the candidates are members of the majority group and the remaining 25% belong to the minority group (cf. Hattrup *et al.* 1997). The four predictors are cognitive ability (CA), structured interview (SI), conscientiousness (CO) and biodata (BI). Alternatively, the overall performance criterion is perceived as a weighted sum of the two constituting dimensions, with weights 2 and 1 for the task (job) performance (TP) and the contextual performance (CP) dimensions (cf. Motowidlo, Borman and Schmitt 1997; Borman, Penner, Allen and Motowidlo 2001), respectively.

Table 1 summarizes the values used for the different characteristics of both the predictors and the criterion dimensions. The majority of these values derive from the meta-analytic study of Bobko *et al.* (1999), whereas the remaining numbers correspond to results presented by Hattrup *et al.* (1997), McManus and Kelly (1999) and Murphy and Shiarella (1997). It is important to emphasize that the figures included in Table 1 are first and foremost used for illustrative purposes. Although the values that are borrowed from the study of Bobko *et al.* (1999) are based on extensive meta-analytic research and, hence, reflect a fairly accurate summary, it is recognized that this is less the case for the other numbers in the table. For example, Borman *et al.* (2001) report a value of .24, instead of the presently used value of .20, for the correlation between conscientiousness and contextual performance; whereas some studies report substantially higher correlation values between task and contextual performance than the Table 1 value of .17. Alternatively, and as argued above, the Table 1 correlation data satisfy the requirement of a common metric in the sense that they all correspond to uncorrected correlation values.

Given the Table 1 data, the CAIQS program is used to compute the expected adverse impact and the expected quality of the selected work force for four values of the overall selection ratio, s_t: $s_t = .1, .3, .5, .7$. Also, to illustrate the effect of different predictor weighing scenarios, the calculations are repeated for three alternative scenarios: equal weighting of the predictors, regression based weights and a weighing scheme that reduces the effect size of the composite predictor and, hence, the expected level of adverse impact. In the regression condition the weights are equal to 0.216, 0.264, 0.098 and 0.207 for the CA, SI, CO and BI

Table 1: Effect sizes and intercorrelations of the performance predictors and the performance criteria

Variable	Effect size	Intercorrelation Matrix					
	d	1	2	3	4	5	6
Predictors							
1. Cognitive ability	−1.00						
2. Structured Interview	−0.23	.24					
3. Conscientiousness	−0.09	.00	.12				
4. Biodata	−0.33	.19	.16	.51			
Criterion dimensions							
5. Task (Job) Performance	−0.45	.30	.30	.18	.28		
6. Contextual Performance	0.00	.16	.26	.20	.25	.17	

Table 2: Expected quality and expected adverse impact for different predictor weighing scenarios

	Equal Predictor Weights				Regression Based Weights				AI Reducing Weights			
SR	*i*	$\bar{Z}_t^{(g)}$	$\bar{Z}_a^{(g)}$	$\bar{Z}_i^{(g)}$	*i*	$\bar{Z}_t^{(g)}$	$\bar{Z}_a^{(g)}$	$\bar{Z}_i^{(g)}$	*i*	$\bar{Z}_t^{(g)}$	$\bar{Z}_a^{(g)}$	$\bar{Z}_i^{(g)}$
.1	.285	.811	.820	.715	.257	.834	.840	.763	.758	.653	.719	.395
.3	.429	.507	.523	.391	.400	.522	.534	.427	.832	.400	.471	.142
.5	.553	.321	.346	.184	.526	.331	.352	.212	.880	.246	.322	−.014
.7	.687	.165	.203	.000	.666	.171	.205	.020	.923	.118	.199	−.146

Note: SR indicates the overall selection rate.

predictor, respectively; whereas the corresponding weights in the adverse impact reducing scheme are equal to −0.077, 0.306, 0.102 and 0.188, respectively. For all the analyses, the effect size of the criterion composite, d_Z, has a constant value equal to −0.378, indicating that the average composite criterion score of the minority applicant group is 0.378 standard units lower than the corresponding average in the majority applicant population. The three predictor weighing scenarios differ, however, in terms of the effect size, d_U, and the validity, ρ_{UZ}, of the resulting predictor composite. More specifically, the values for d_U and ρ_{UZ} are equal to −0.650 and .494, -0.698 and .507, and −0.155 and .417 for the equal, the regression based and the adverse impact reducing weighing schemes, respectively.

Table 2 displays the main results of the calculations. For each combination of the overall selection ratio value and the type of predictor weighing scenario, the table indicates the expected adverse impact and the globally standardized, expected criterion scores for the minority, the majority and the total group selected individuals.

As expected from the composite predictor effect sizes and validities that are associated with the three predictor weighing schemes, the results show that the regression-based predictor composite leads to the highest expected criterion scores, but also to the lowest values for the expected adverse impact ratio. In contrast, the composite predictor based on the adverse impact reducing weights generates the lowest expected criterion scores. Taken together, these findings nicely illustrate the trade-off faced by the selection decision maker when addressing the qualityadverse impact issue.

As explained above, the CAIQS program can also be used to determine the overall selection rate that is consistent with the four-fifths rule of no adverse impact. Application of this feature to the present example data results in a minimum acceptable overall selection ratio of .840, .855 and .197 for the equal, the regression based and the adverse impact reducing weighing scenarios, respectively. The expected (again globally standardized) criterion scores (i.e., the values of $\bar{Z}_i^{(g)}$ $\bar{Z}_a^{(g)}$, and $\bar{Z}_t^{(g)}$ in that order) that correspond to these selection ratios are −0.137, 0.109, and 0.057 when the predictors are weighed equally, −0.140, 0.099 and 0.049 for the regression based predictors, and 0.248, 0.575 and 0.506 for the adverse reducing weighing scheme. So, adding the constraint of no adverse impact (i.e., the adverse impact ratio, *i*, equal to 0.80), has a dramatic effect on the relative performance of the three predictor weighing schemes. Obviously, the effect is immediately related to the minimum acceptable selection rate that can be achieved under the different weighing approaches. Also, although the effect is quite familiar, the present method has, compared to existing approaches, the distinct advantage that it permits an accurate estimate of the magnitude of this effect.

Effects of Prior Selection on Ethnic Group Differences

The second example relates to some of the results reported by Roth *et al.* (2001) on the effects of prior selection using a first predictor, X_1, on the ethnic group differences with respect to a second, related predictor, X_2. Given the intercorrelation and the population effect sizes of the predictors, these authors use Monte Carlo simulation methods to estimate the standardized group differences that will be found with respect to X_2 after top-down selection using the first predictor X_1. As the method, discussed in the present article, provides accurate estimates of the expected performance of the majority and the minority selected applicants, it can be applied to replicate and, eventually correct, the Roth *et al.* analyses. Also, for the sake of comparability with the Roth *et al.* results, the majority group standardized values $\bar{Z}_a$ and $\bar{Z}_1$ (instead of the globally standardized means) will be used to determine the standardized ethnic group difference after the prior selection on X_1.

The example considers a total of 20 situations. These situations correspond to the cells of a design in which the following three factors are crossed: selection rate, s_t, with five levels (i.e., s_t equal to .1, .3, .5, .7 and .9); intercorrelation of the predictors, $\rho_{X_1X_2}$ with two levels (i.e., $\rho_{X_1X_2}$ equal to .19 and .38); and effect size of the first

Table 3: Effect of prior selection on standardized ethnic group differences

	Predictor correlation equal to .19 and and $d_{X_1} equalto.52$				Predictor correlation equal to .38 and and $d_{X_1} equalto.52$			
Selection	$d - X_1 = .72$		$d - X_1 = .98$		$d - X_1 = .72$		$d - X_1 = .98$	
Rate	ME	AE	ME	AE	ME	AE	ME	AE
.1	.30	.40	.55	.36	.37	.29	.14	.20
.3	.40	.42	.45	.38	.34	.31	.27	.23
.5	.45	.43	.43	.39	.38	.33	.32	.26
.7	.47	.44	.45	.41	.42	.37	.38	.31
.9	.50	.47	.46	.46	.46	.43	.46	.39

Notes:
ME: Monte Carlo simulation estimate (cf. Roth *et al.* 2001)
AE: Accurate, analytically computed estimate

predictor, d_{X_1}, also with two levels (i.e., d_{X_1} equal to .72 and .98). Also, all situations assume a value of .52 for the population effect size of the second predictor, d_{X_2}.

Table 3 summarizes the results of the example application. For each of the above described situations, the table has two entries that both provide an estimate of the effect size of the second predictor after selection on the basis of X_1. The first entry, labeled ME, is the value obtained by Roth *et al.* (cf. their Table 3 on p. 602) whereas the second entry (AE) corresponds to the accurate value computed by the present method.

Inspection Table 3 shows that the two estimates are generally different and that some of these differences are quite substantial. As an example, consider the estimate associated with an overall selection rate of .10 in the situation where $d_{X_1} = .98, d_{X_2} = .52$ and the predictor intercorrelation equals .19. In that case, Roth *et al.* (2001) obtain a value of .55 for the effect size (after the selection) of the second predictor, whereas the present method indicates a much lower value of .36. Observe that the Roth *et al.* (2001) estimate is logically impossible because the restricted effect size (i.e., the effect size after the selection) cannot be larger than the initial population effect size of .52 when the predictors have a positive correlation.

In general, the differences between the corresponding estimates of Table 3 stem from two origins. First, as Roth *et al.* (2001) invoke Monte Carlo simulation methods, their results can be only approximate. Second, and most importantly, their actually used simulation procedure is somewhat deficient because the data are not sampled from the apparently intended bivariate normal distribution (Switzer, personal communication) of the predictors, but instead are characterized in terms of only the marginal univariate distributions. Despite the errors in the results, it should be recognized that the main conclusion of the Roth *et al.* study remains valid. Some of the other observations made by these authors must be amended, however. In particular, their finding that the downward bias with respect to the effect size of the second predictor is not always inversely related to the overall selection rate should be corrected.

Discussion

The previous examples illustrate the potential of the present method. The method and, in particular, the related computer program enable the selection researcher and practitioner to explore the probable effects of an anticipated selection decision in terms of both the expected adverse impact and the expected quality of the selected work force. Unlike previous procedures, the approach is comprehensive and it provides analytically computed, precise estimates of the major selection outcomes, provided that the required input values are accurate and, in particular, that the correlations between the elementary predictor and criterion dimensions are expressed in a common metric (see above). However, as acknowledged from the outset, the method focuses on single-stage selection decisions. In addition, top-down selection, without refusals from selected candidates, has been assumed throughout.

With respect to the first limitation, it was observed that the extension of the method to the case of general multistage selection from a heterogeneous applicant group has not been attempted yet. Although such an extension seems plausible, at least in a formal sense, related work by De Corte (1998) indicates that its practical implementation will pose considerable computational problems. Similar problems are expected when addressing other than top-down selection scenarios. Furthermore, incorporating the effects of, for example, job refusal, requires fairly detailed information about the underlying process and its eventual relation with performance on the predictor variables. Such information may not be available or be situation specific and therefore difficult to model within a generally applicable framework.

Apart from the above limitations, the method has one other shortcoming that it shares with all the thus far developed approaches to estimate the effects of a selection decision before it is actually implemented. As outlined by De Corte (1999b), these procedures, as well as the present one, assume implicitly that the applicant group consists of an unlimited number of candidates. As a consequence, the resulting estimates refer to population values that must be distinguished from the corresponding sample statistics that characterize selections from a finite applicant pool. Also, using a result proven by Gillett (1991), De Corte (1999b) shows that these population values overestimate the correct sample values and that the magnitude of the overestimation is inversely related to the actual size of the total applicant group.

The practical consequence of the above finding is that the present approach produces somewhat optimistic estimates of the expected criterion performance of the selected individuals, especially when the number of initial applicants is less than 20. Also, although De Corte (1999b) presented a procedure to obtain accurate estimates for small sample selection decisions, it is not yet clear how this approach can be extended to the situation in which the applicant group is a mixture of majority and minority group members. Alternatively, it is equally important to emphasize that the overestimation problem relates essentially to the absolute value of the expected selection results. Comparisons between alternative selection scenarios, that all pertain to the same (finite) applicant pool but differ in the weights assigned to the different predictors, are much less affected by the problem. As the present method and the related program are first and foremost intended as a tool to assist the selection practitioner in deciding between such alternative selection scenarios, it is expected that its results, when used comparatively, are fairly accurate even for small sample decisions.

Appendix

This Appendix provides details on the method used in the CAIQS program to compute the expected adverse impact and the expected criterion score of the minority, the majority and the total group selected applicants. Given the notation introduced in the article and using the vector representation $\mathbf{x} = (X_1, \ldots, X_P)'$ and $\mathbf{y} = = (Y_1, \ldots, Y_C)'$ for the predictor and criterion variables, it follows from the assumptions that in the majority applicant group the joint distribution of **x** and **y** is multivariate normal with zero mean vector and variance/covariance matrix equal to the correlation matrix, $\boldsymbol{R}$, of the predictor and criterion dimensions: $(x, p) \sim N_{P+C}(\mathbf{0},\boldsymbol{R})$. Likewise, $(\mathbf{x},\ \mathbf{y}) \sim N_{P+C}((\boldsymbol{d}_{\mathrm{X}},\ \boldsymbol{d}_{\mathrm{Y}})$', $\boldsymbol{R})$ in the minority applicant group, where $\boldsymbol{d}_{\mathrm{X}} = d_x, \ldots, d_{X_p}$ and $\boldsymbol{d}_{\mathrm{Y}} = d_{Y_1} \ldots d_{Y_c})'$. The assumptions also imply that the composite predictor $U = \boldsymbol{a}'x$, with $\boldsymbol{a}$ the vector of predictor weigts, and the composite criterion $Z = \boldsymbol{b}'y$, with $\boldsymbol{b}$ the vector of criterion weights, have a standard bivariate normal distribution with correlation parameter, ρ_{UZ}, equal to $(\boldsymbol{a},\ \mathbf{0}_c)'\boldsymbol{R}(\mathbf{0}_p,\ \boldsymbol{b})$ in the majority group, where $\mathbf{0}_c$ and $\mathbf{0}_p$ represent zero vectors with c and p elements respectively. Alternatively,

$$(U, Z) \sim \mathrm{N}_2\left(d, \begin{pmatrix} 1 & \rho_{uz} \\ \rho_{uz} & 1 \end{pmatrix}\right)$$

in the minority group, where $\boldsymbol{d} = (d_U,\ d_Z)'$, $d_U = \boldsymbol{a}'\boldsymbol{d}_{\mathrm{X}}/\rho_U$ and $d_Z = \boldsymbol{b}'\boldsymbol{d}_{\mathrm{Y}}\ /\sigma_Z$. Also, the variance of the predictor composite, σ_Z^2, and the criterion composite, σ_U^2, can for both applicant groups be determined as $\sigma_Z^2 = \Sigma_{i=1}^{p}\Sigma_{j=1}^{p} a_i a_j$ and, $\sigma^2 = \Sigma_{i=1}^{c}\Sigma_{j=1}^{c} b_i b_j$ respectively.

The above results entail that the overall selection rate, s_t, that corresponds to a particular composite predictor cutoff value, u_c, can be determined as

$$s_t = p_a[1 - \Phi(u_c)] + p_1[1 = \Phi(u_c - d_U)]$$

where p_a and p_i express the proportion of majority and minority applicants in the total applicant population and $\Phi(\cdot)$ denotes the standard normal distribution function. As a consequence, the predictor cutoff value that corresponds to a given overall selection rate can be found by solving a non-linear equation in terms of u_c. Several excellent methods exist to obtain such a solution and the thus determined value can then be used to calculate s_a as $s_a = 1 - \Phi(u_c), s_j$ as $s_i = 1 - \Phi(u_c - d_U)$ and i as s_i/s_a. Next, using $\phi(\cdot)$ to denote the standard normal density function, the expected composite criterion score of the minority, the majority and the total group selected applicants can be estimated as

$$\bar{Z}_i = d_z + \rho_{UZ} \frac{\phi(u_c - d_U)}{1 - \Phi(U_c - d_u)}$$

$$\bar{Z} = \rho_{UZ} \frac{\phi(u_c)}{1 - \Phi(u_c)}$$

and

$$\bar{Z} = \frac{p_i s_i \bar{Z}_i + p_a s_a \bar{Z}_a}{s_t}$$

respectively.

It is important to observe that the above derived expected composite criterion scores are expressed relative to the distribution of the composite criterion scores in the majority applicant population. To obtain the values expressed relative to the standardized distribution of the composite criterion score in the total applicant population (cf. the globally standardized values $\bar{Z}_i^{(g)}$, $\bar{Z}_a^{(g)}$ and $\bar{Z}_t^{(g)}$ in the article), one final transformation is required:

$$\bar{Z}_i^{(g)} = \frac{\bar{Z}_i - p_i d_z}{\sqrt{1 + p_1 P_a d_z^2}}$$

$$\bar{Z}_i^{(g)} = \frac{\bar{Z}_a - p_i d_z}{\sqrt{1 + p_1 P_a d_z^2}}$$

$$\bar{Z}_i^{(g)} = \frac{\bar{Z}t - p_i d_z}{\sqrt{1 + p_1 P_a d_z^2}}$$

where $p_i d_Z$ and $1 + p_i p_a$ correspond to the mean and the variance of the composite predictor scores in the total applicant population.

Given the weights that are used to combine the predictors to the composite predictor, the CAIQS program also determines the overall selection ratio that corresponds to a value of 0.80 for the adverse impact ratio (cf. the selection ratio $s^{(0)}$ in the article). To obtain the value of $s^{(0)}$, the following non-linear equation must be solved in terms of the composite predictor cutoff u_c:

$$s^{(0)} - p_l s_l - p_a s_a = 0.$$

To solve this equation, the dfzero routine from the slatec library (see http://www.netlib.org/slatec/) is invoked.

References

Aguinis, H., Cortina, J.M. and Goldberg, E. (1998) A new procedure for computing equivalence bands in personnel selection. *Human Performance*, **11**, 351–365.

Bartram, D. (1995) Predicting adverse impact in selection testing. *International Journal of Assessment and Selection*, **3**, 52–61.

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

Borman, W.C. and Motowidlo, S.J. (1997) Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, **10**, 99–110.

Borman, W.C., Penner, L.A., Allen, T.D. and Motowidlo, S.J. (2001) Personality predictors of citizenship performance. *International Journal of Assessment and Selection*, **9**, 52–69.

Campion, M.A., Outtz, J.L., Zedeck, S., Schmidt, F.L., Kehoe, J.F., Murphy, K.R. and Guion, R.M. (2001) The controversy over score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology*, **54**, 149–185.

Cascio, W.F., Outtz, J., Zedeck, S. and Goldstein, I.L. (1991) Statistical implications of six methods of test score use in personnel selection. *Human Performance*, **4**, 233–264.

De Corte, W. (1998) Estimating and maximizing the utility of sequential selection decisions with a probationary period. *British Journal of Mathematical and Statistical Psychology*, **51**, 101–121.

De Corte, W. (1999a) Weighing job performance predictors to both maximize the quality of the selected workforce and control the level of adverse impact. *Journal of Applied Psychology*, **84**, 695–702.

De Corte, W. (1999b) A note on the success ratio and the utility of fixed hiring rate personnel selection decisions. *Journal of Applied Psychology*, **84**, 952–958.

Gillett, R. (1991) Population comparison based on an extreme fraction of the sample: A caveat. *British Journal of Mathematical and Statistical Psychology*, **44**, 289–297.

Hattrup, K., Rock, J. and Scalia, C. (1997) The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predictor performance. *Journal of Applied Psychology*, **82**, 656–664.

Hough, L.M., Oswald, F.L. and 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.

McManus, M.A. and Kelly, M.L. (1999) Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry. *Personnel Psychology*, **52**, 137–148.

Motowidlo, S.J., Borman, W.C. and Schmit, M.J. (1997) A theory of individual differences in task and contextual performance. *Human Performance*, **10**, 71–83.

Murphy, K.R. and Shiarella, A.H. (1997) Implications of the multidimensional nature of job performance for the validity of selection tests: multivariate framework for studying test validity. *Personnel Psychology*, **50**, 823–854.

Roth, P.L., Bobko, P., Switzer, F.S., III and Dean, M.A. (2001) Prior selection causes biased estimates of standardized ethnic group differences: simulation and analysis. *Personnel Psychology*, **54**, 591–617.

Sackett, P.R. and 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. and Roth, L. (1996) Multistage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology*, **49**, 549–572.

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

Sackett, P.R. and Wilk, S.L. (1994) Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, **49**, 929–954.

Schmidt, F.L. and Hunter, J.E. (1998) The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, **124**, 262–274.

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

te Nijenhuis, J.T. and Van der Flier, H. (1997) Comparability of GATB scores for immigrants and majority group members: Some Dutch findings. *Journal of Applied Psychology*, **82**, 675–687.

Viswesvaran, C. and Ones, D.S. (2000) Perspectives on models of job performance. *International Journal of Selection and Assessment*, **8**, 216–226.