Industrial and Organizational Psychology (2023), **16**, 322–327
doi:10.1017/iop.2023.38

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

To correct or not to correct for range restriction, that is the question: Looking back and ahead to move forward

In-Sue Oh[1], Jorge Mendoza[2], and Huy Le[3]

[1]Department of Management, Fox School of Business, Temple University, Philadelphia, PA, USA, [2]Department of Psychology, Dodge Family College of Arts and Sciences, University of Oklahoma, Norman, OK, USA and [3]Department of Management, Alvarez College of Business, University of Texas at San Antonio, San Antonio, TX, USA
Corresponding author: In-Sue Oh; Email: insue.oh@temple.edu

(Received 13 March 2023; revised 7 April 2023; accepted 12 April 2023)

Sackett et al. (2023) start their focal article by stating that they identified "previously unnoticed flaws" in range restriction (RR) corrections in most validity generalization (VG) meta-analyses of selection procedures reviewed in their 2022 article. Following this provocative opening statement, they discuss how both researchers and practitioners have handled (and should handle) RR corrections in estimating the operational validity of a selection procedure in both VG meta-analyses (whose input studies are predominantly concurrent studies) and individual validation studies (which serve as input to VG meta-analyses). The purpose of this commentary is twofold. We first provide an essential review of Sackett et al.'s (2022) three propositions serving as the major rationales for their recommendations regarding RR corrections (e.g., no corrections for RR in concurrent validation studies). We then provide our critical analyses of their rationales and recommendations regarding RR corrections to put them in perspective, along with some additional thoughts.

Essential review of Sackett et al.'s three propositions regarding RR corrections

Sackett et al. (2022) advance three propositions regarding RR and RR corrections in concurrent validation studies: (a) "any range restriction (on the predictor of interest, X) will only be indirect"; (b) "with rare exceptions Z is not highly correlated with X" where Z is the direct basis for selection; and (c) "indirect range restriction restricts variance on the selection predictor of interest to only a very modest degree under virtually all realistic circumstances" (pp. 2041–2042). Testing these propositions using simulated data, Sackett et al. (2022, Table 1; 2023) state that when RR is indirect (e.g., in concurrent validation studies), ux (= restricted, incumbent-based SDx/unrestricted, applicant-based SDx) tend to be .90 or higher, and, thus, the RR correction effect will be relatively small (see Oh et al., 2023, Table 2 for counterevidence). This serves as a basis for their recommendation for not correcting for RR in concurrent studies and most VG meta-analyses reviewed in their 2022 article, whose input is predominantly (approximately 80%) concurrent studies. Relatedly, Sackett et al. (2023) also state that the degree of RR "can be quite substantial" in predictive studies, whereas, as noted above, RR "is likely to be small anyway" in concurrent studies. This differential RR by validation design (predictive vs. concurrent) serves as another basis for their recommendation against applying RR data (i.e., the ux distribution) from applicant-based predictive studies to RR corrections in concurrent studies or VG meta-analyses whose input is mostly concurrent studies. They argue that this (problematic in their view) "uniform RR correction" approach[1] has been implemented in most previous VG meta-analyses, which is why they claim that RR corrections applied in such meta-analyses have "flaws."

In summary, Sackett et al.'s (2022, 2023) solution to the "previously unnoticed flaws" in RR corrections in most VG meta-analyses reviewed in their 2022 article is to *question* and, in most cases, *undo* RR corrections with the premise that most input studies in those VG meta-analyses are concurrent studies. That is, most of their newly suggested meta-analytic mean operational validity estimates for various selection procedures are, basically, those *uncorrected* for RR. Along this line, Sackett et al. (2022) propose a sensational recommendation for individual validation (and VG) studies: "Absent credible U ($= 1/u$) ratios for concurrent studies, and in light of the above demonstration that U ratios are likely to be close to 1.0 in concurrent studies, we *recommend no range restriction correction for concurrent studies*" (p. 2044).

Having reviewed the major underpinnings of Sackett et al.'s recommendations regarding RR corrections, we share our critical analyses of Sackett et al.'s (2022, 2023) recommendations as follows, focusing on their potential pitfalls and challenges.

Critical analyses of Sackett et al.'s recommendations regarding RR corrections

Dichotomous thinking and sweeping recommendations

We are surprised, if not shocked, by Sackett et al.'s (2022, 2023) "dichotomous thinking" and sweeping recommendation for "no" RR corrections in any concurrent studies, regardless of which predictor is being considered. If their recommendation were taken at face value, the implication would be that all concurrent studies are alike in terms of the degree of RR (i.e., no RR) and, thus, the resulting ux values are the same (ux = 1) across all such studies and across all predictors. Sackett et al. do not provide sufficient evidence in support of this. In fact, research evidence shows that the degree of RR varies by predictor. In fact, ux values tends to be smaller for cognitively loaded selection procedures than for noncognitive selection procedures. For example, average ux values for personality traits are in the .90 range (e.g., Schmidt et al., 2008).[2] Conversely, average ux values for cognitive ability tests tend to be in the .70 range (e.g., Alexander et al., 1989).

In addition, Sackett et al.'s recommendation for no corrections for RR in concurrent studies, unlike their disclaimer, does not appear to be limited to input studies included in most VG meta-analyses reviewed in Sackett et al. (2022, Table 2). It does, however, appear to be generalizable to future individual validation studies (see the last section about "correcting individual studies" in Sackett et al., 2023). One issue here is their assumption about the magnitude and nature of the correlation between X and Z ("rzx")[3]; as "rzx" increases, ux decreases. Sackett et al.'s assumption

[1]Sackett et al. object to using ux values computed from predictive studies and applying them to an entire VG study (a "uniform RR correction" approach) for the following two reasons: (a) applicant-based SDx is only available from applicant-based predictive studies (and unavailable from concurrent studies) and (b) such predictive studies *may not be* random draws (representative samples) from an entire VG meta-analysis database whose input is predominantly concurrent studies. Morris (2023) points out that there is no clear evidence for (b).

[2]To be clear, this is not intended to provide support for Sackett et al.'s recommendation for no corrections for RR in concurrent validation studies, as even ux values of .90 or higher should not be ignored as they can lead to a substantial amount of RR correction (see Oh et al., 2023, Table 2).

[3]On one hand, Sackett et al.'s (2022) assumption that "rzx" is the unrestricted true-score correlation, thus having no measurement error, can be questioned in many situations. But let's assume that "rzx" is the unrestricted true-score correlation between the selection variable Z and the concurrent predictor X. Then, the next question is the magnitude of this correlation. If the "rzx" correlation is high, then selection on Z will create a significant indirect RR on X, substantially attenuating the correlation between X and Y, the criterion. However, according to Sacket et al. (2023), this is not the case in most concurrent studies although their argument is only based on their speculation that high "rzx" correlations over .50 or .70 are "implausible" in those VG meta-analyses reviewed in their original article. Then, the next question is whether Z and X measure the same construct (e.g., two different predictors, such as structure interviews and work sample tests, developed to measure the same set of job-relevant knowledge, skills and abilities). If they measure the same construct without measurement error, one expects the

https://doi.org/10.1017/iop.2023.38 Published online by Cambridge University Press

that "rzx" always falls in a certain range (e.g., $\leq .50$) regardless of the nature of Z and X and their specific application in selection settings certainly lacks sufficient evidence and needs caution (see Oh et al., 2023). For example, as noted in their focal article, we are likely to see increasingly more validation studies on "a comparison between the validity of legacy predictors and gamified assessments, asynchronous video interviews, natural language processing-based scoring of essays, and the like" or a linear composite of multiple predictors, such as a cognitive ability test and a job-relevant personality (e.g., conscientiousness) measure as an effort to balance validity (performance) and adverse impact (diversity). We believe this would produce much higher "rzx" correlations and thus result in much lower ux values than suggested in Sackett et al.'s articles, thus warranting RR corrections.[4]

In summary, our major issue with Sackett et al.'s (2022, 2023) articles is their sweeping recommendations (i.e., no RR correction in concurrent validation studies regardless of which predictor is being considered) without compelling evidence. As noted in Roberts (2022), "scientific arguments should not be presented without evidence, as doing so shifts the burden of proof to the reader and pretends that the claims are established facts (burden of proof fallacy). Scientific arguments should also not be framed within dichotomous "either–or" scenarios, as doing so is unsound (false dichotomy fallacy)" (p. 22).

Differential RR by validation (study) design

As briefly noted above, Sackett et al. (2022, 2023) advance statements that appear to invoke differential RR by validation design (concurrent vs. predictive) without compelling evidence. In particular, Sackett et al. (2023) state that:

"[T]he degree of RR can be quite substantial in predictive studies; however, Sackett et al. (2022) demonstrated that restriction will commonly (though not always) be quite small in concurrent studies. Applying a large correction factor derived from predictive studies to concurrent studies thus results in an overcorrection, which is often quite large."

To put the statement above in perspective, it is worth noting other esteemed personnel selection scholars' opinions and findings concerning Sackett et al.'s assertion. In particular, Morris (2023) raises a legitimate apprehension: "While it is important to consider the representativeness of the data used for artifact corrections, it is not clear that predictive and concurrent designs will

"rzx" correlation to be high. On the other hand, if Z and X measure different but job-relevant constructs (i.e., job knowledge and social skill, respectively), then the correlation of interest in the concurrent study represents the correlation between scores on a social skill measure and the criterion (say, r for a social skill test) after restriction on scores on a job knowledge test. However, using the job knowledge test alone in a follow-up future validation study represents the correlation between social skill and the criterion ignoring job knowledge (say, r * for a social skill measure). In other words, if one selects based on a job knowledge test in the initial validation study, then a future validation study examining the validity of a social skill measure would be based on individuals with the knowledge needed to perform the job (a sample that is already restricted in job knowledge test scores due to prior selection on that test). Whether r is the same as r * is not the point. Fundamentally, these are two very different correlations. The point is that the r * does not represent the typical validity of the social skill measure by itself in a future validation study. Furthermore, in cases in which "rzx" can be unbiasedly estimated, one should do so and not make assumptions about the nature of this correlation. As noted in the conclusion section, future research along this line should pay more attention to multiple-hurdle selection systems (see Hunter et al., 2006, pp. 606–607; Mendoza et al., 2004). Finally, we should point out that although sample-based correlational estimates are always affected by RR, regression coefficients are not if RR is direct or the predictor of interest is measured without measurement error. If raw data are available, we may obtain unbiased regression coefficients and substantiate them with a utility measure.

[4]Such comparisons would be inevitably based on the measurement of the same construct (e.g., cognitive ability) using different methods (e.g., a legacy method such as written exams vs. an emerging method such as gamifications; Arthur & Villado, 2008). In addition, in the simplest form, a correlation between a unit-weighted two-predictor composite (Z; e.g., a structured interview and a cognitive ability test) and one of the predictors (X; e.g., a structured interview) in the composite is a part–whole correlation. This correlation (representing Sackett et al.'s "rzx") is lowest at .71 when they are unrelated. However, it increases up to .87 when the two predictors are correlated at .50, suggesting that "rzx" can be much higher than suggested in Sackett et al., when using a predictor composite (selection battery) as an actual selection basis (i.e., X is part of Z).

https://doi.org/10.1017/iop.2023.38 Published online by Cambridge University Press

necessarily produce substantially different levels of range restriction" (p. 238). Morris further elaborate that although it is true that RR can be more substantial in predictive rather than concurrent studies (because RR can be direct), most, if not all, predictive studies are also subject to indirect RR as is the case for concurrent studies; that is, few, if any, predictive studies are subject to direct RR (Schmidt et al., 2006). We agree with Morris that the field has yet to see compelling evidence for differential RR by validation design.

In fact, existing evidence does not appear to support Sackett et al.'s (2022) statements suggesting different RR by validation design. For example, in their influential meta-analysis, Schmitt et al. (1984) report that "concurrent validation designs produce (observed) validity coefficients roughly equivalent to those obtained in predictive validation designs," thus illustrating that predictive and concurrent designs are unlikely to result in very different degrees of RR (p. 37). The same conclusion was found in the largest-scale meta-analysis currently available (Pearlman et al., 1980, p. 380). Based on all these findings, Schmidt et al. (1985) state, "contrary to general belief, predictive and concurrent studies suffer from range restriction to about the same degree" (p. 750). In summary, although there might be debate over how to correct for RR in concurrent studies, there is no compelling evidence for differential RR by validation design for each selection procedure reviewed in Sackett et al.'s (2022) article.

(Un)Availability of credible and sound information about RR

Sackett et al. (2022, 2023) repeatedly state that without "credible" and "sound" information about RR, we'd better not attempt to correct for RR. This raises a critical question: What do Sackett et al. indeed mean by the "credible" and "sound" information about RR with which we correct for RR in concurrent studies? The answer is clear: *applicant*-based SDx. This, in turn, raises another critical question: Is such credible information on RR available from concurrent (and most validation) studies? The answer is clear: NO! Such information as *applicant*-based SDx is unavailable from concurrent studies because concurrent studies, by definition, are based on *incumbents*, not applicants. As such, we are left wondering how feasible it is to obtain credible or sound information about RR in concurrent studies because such information is unavailable. Importantly, a lack of credible information about RR does not necessarily mean that RR does not exist in concurrent studies. This naturally leads to the question of whether there are good alternatives to applicant SDx absent in concurrent studies.

Alternatives to applicant-based SDx

In fairness to Sackett et al. (2022, 2023), they discuss a possible solution to the lack of applicant-based SDx in concurrent (and incumbent-based predictive[5]) studies. Specifically, Sackett's own study based on a cognitive ability test (Sackett & Ostgaard, 1994) shows that a large number of *job-specific applicant*-based SDx values collected across various jobs are, on average, 10% smaller than their *applicant-based national norm* SDx. Subsequent studies have been conducted based on this approach (e.g., Hoffman, 1995 for various ability and mechanical tests; Ones & Viswesvaran, 2003 for self-reported personality measures). With this approach, we will know how much downward adjustment should be made to the applicant-based national workforce norm SDx to derive a reasonable estimate of the expected applicant-based SDx value that can be used to correct for RR in concurrent and other types of validation studies without such unrestricted SDx values. However, as in the following paragraph from Sackett et al. (2022), their tone (note "an argument for skepticism about this approach" below) is not encouraging but rather skeptical as they now do not view it as credible information regarding RR.

[5]Although Sacket et al. did not make it clear, not all predictive studies are based on applicants and not all applicant-based predictive studies report applicant-based SDx values (Van Iddekinge & Ployhart, 2008).

"Sackett and Ostgaard (1994) obtained applicant pool SDx values for a large number of jobs and then pooled the data across jobs as an estimate of workforce SDx; they reported that the *applicant* pool SDx values were on average 10% smaller than the workforce SDx estimate. So, based on Sackett and Ostgaard's finding, it is *at least hypothetically possible* that it could be reasonable to pool *incumbent* data across jobs to estimate the unrestricted SDx, and then reduce that SDx by 10%. However, we offer *an argument for skepticism about this approach*, at least in terms of the U ratio estimate it produced in Hunter (1983)." (p. 2045)

What is odd to us is that we know that Hunter's (1983) approach was based on only *incumbent* data without any downward adjustments, whereas Sackett and Ostgaard's (1994) approach mentioned above is rightfully based on only *applicant* data with proper downward adjustments. Thus, we are surprised that Sackett et al. (2022) discuss these two approaches side by side without clarifying the stark difference in input data (also see Roth et al., 2017). To be clear, we advocate the use of *applicant*-based national workforce norm SDx values with proper downward adjustments.

In the focal article, Sackett et al. (2023) note that it may be acceptable to use job/occupation-specific (vs. national workforce) norm-based applicant SDx. In theory, we do not disagree with this suggestion because it is analogous to using available local applicant-based SDx values (e.g., prior applicant data kept in an organization) for correcting for RR in present concurrent studies (see Hoffman, 1995, Table 4). However, we are concerned about the (un)availability of job/occupation-specific norm-based applicant SDx values from many test manuals and large-scale validation studies. Sackett et al. appear to be quite restrictive and skeptical (not encouraging) about this applicant-based national workforce norm approach, which is further elaborated in their caution. However, unlike Sackett et al. (2022, 2023), we believe that applicant-based national workforce norm SDx values can be used as good surrogates for applicant-based SDx with proper downward adjustments; this is better than assuming no RR in concurrent studies. That is, we agree with Sackett and Ostgaard (1994) and other subsequent studies that applicant-based national workforce norm SDx should not be used liberally without proper adjustments (Roth et al., 2017).

Concluding thoughts

There is no disagreement among scholars that "the influence of selection (RR) upon the resulting validity coefficients becomes a very substantial matter where a high standard of selectivity exists" (Thorndike, 1949, p. 170). Looking back, all thoughts and evidence presented above suggest that there is no compelling evidence for Sackett et al.'s (2022, 2023) sweeping (e.g., without considering which predictor is being considered and in which situation/application) recommendation that the effect of RR correction is so small that it is better to not attempt RR corrections in concurrent validation studies. Looking ahead, we agree with Morris (2023) that "additional work is needed to fully understand the representativeness of range restriction estimates and optimal correction procedures under typical conditions" (p. 238). For example, considering (a) that many firms choose to use a multiple-hurdle model of personnel selection increasingly often, to save hiring-related time and costs, and (b) that traditional RR correction methods based on a compensatory model of selection discussed in this article do not directly apply to the multiple hurdle model (see Mendoza et al., 2004 for exceptions), it will be fruitful for future studies to examine a long overdue issue of how to correct for RR in such selection situations, along with developing more innovative and less restrictive RR correction methods. We also recommend the creation of publicly available large applicant-based data reservoirs that can be used collectively in personnel selection.

References

Alexander, R. A., Carson, K. P., Alliger, G. M., & Cronshaw, S. F. (1989). Empirical distributions of range restricted *SD*x in validity studies. *Journal of Applied Psychology*, **74**, 253–258.

Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, **93**, 435–442.

Hoffman, C. C. (1995). Applying range restriction corrections using publisher norms: Three case studies. *Personnel Psychology*, **48**, 913–923.

Hunter, J. E. (1983). Test validation for 12,000 jobs: An application of synthetic validity and validity generalization to the GATB. US Employment Service, US Department of Labor.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, **91**, 594–612.

Mendoza, J. L., Bard, D. E., Mumford, M. D., & Ang, S. C. (2004). Criterion-related validity in multiple-hurdle designs: Estimation and bias. *Organizational Research Methods*, **7**, 418–441.

Morris, S. B. (2023). Meta-analysis in organizational research: A guide to methodological options. *Annual Review of Organizational Psychology and Organizational Behavior*, **10**, 225–259.

Oh, I.-S., Le, H., & Roth, P. L (2023). Sackett et al.'s (2022) recommendation against correcting for range restriction in concurrent validation studies. *Journal of Applied Psychology*. http://dx.doi.org/10.2139/ssrn.4308528.

Ones, D. S., & Viswesvaran, C. (2003). Job-specific applicant pools and national norms for personality scales: Implications for range-restriction corrections in validation research. *Journal of Applied Psychology*, **88**, 570–577.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, **65**, 373–406.

Roberts, S. O. (2022, December 2). Dealing with diversity in psychology: Science and ideology. https://doi.org/10.31234/osf.io/xk4yu.

Roth, P. L., Le, H., Oh, I. S., Van Iddekinge, C. H., & Robbins, S. B. (2017). Who r u?: On the (in) accuracy of incumbent-based estimates of range restriction in criterion-related and differential validity research. *Journal of Applied Psychology*, **102**, 802–828.

Sackett, P. R., & Ostgaard, D. J. (1994). Job-specific applicant pools and national norms for cognitive ability tests: Implications of range restriction corrections in validation research. *Journal of Applied Psychology*, **79**, 680–684.

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, **107**, 2040–2068.

Sackett, P. R., Zhang, C., Lievens, F., & Berry, C. M. (2023). Revisiting the design of selection systems in light of new findings regarding the validity of widely used predictors. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, **16**(3), 283–300.

Schmidt, F. L., Oh, I.-S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. *Personnel Psychology*, **59**, 281–305.

Schmidt, F. L., Pearlman, K., Hunter, J. E., Hirch, H. R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, **38**, 697–798.

Schmidt, F. L., Shaffer, J. A., & Oh, I. S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, **61**, 827–868.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, **37**, 407–422.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. Wiley.

Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology*, **61**, 871–925.

Cite this article: Oh, I-S., Mendoza, J., & Le, H. (2023). To correct or not to correct for range restriction, that is the question: Looking back and ahead to move forward. *Industrial and Organizational Psychology* **16**, 322–327. https://doi.org/10.1017/iop.2023.38