A critical reflection on computing the sampling variance of the partial correlation coefficient

van Aert, R.C. M.; Goos, C.

Published in:
Research Synthesis Methods

DOI:
10.1002/jrsm.1632

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Received: 27 October 2022 | Accepted: 19 February 2023

DOI: 10.1002/jrsm.1632

RESEARCH IN BRIEF

Research Synthesis Methods WILEY

A critical reflection on computing the sampling variance of the partial correlation coefficient

Robbie C. M. van Aert | **Cas Goos**

Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

Correspondence
Robbie C. M. van Aert, Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.
Email: r.c.m.vanaert@tilburguniversity.edu

Funding information
Veni grant financed by the Dutch Research Council (NWO), Grant/Award Number: VI.Veni.211G.012

Abstract

The partial correlation coefficient quantifies the relationship between two variables while taking into account the effect of one or multiple control variables. Researchers often want to synthesize partial correlation coefficients in a meta-analysis since these can be readily computed based on the reported results of a linear regression analysis. The default inverse variance weights in standard meta-analysis models require researchers to compute not only the partial correlation coefficients of each study but also its corresponding sampling variance. The existing literature is diffuse on how to estimate this sampling variance, because two estimators exist that are both widely used. We critically reflect on both estimators, study their statistical properties, and provide recommendations for applied researchers. We also compute the sampling variances of studies using both estimators in a meta-analysis on the partial correlation between self-confidence and sports performance.

KEYWORDS

meta-analysis, partial correlation coefficient, sampling variance, standard error

Highlights

What is already known

- The partial correlation coefficient quantifies the relationship between two variables while taking into account the effect of one or multiple control variables.
- Meta-analyses on partial correlation coefficients are regularly conducted in different research fields.

What is new

- The literature on estimators of the sampling variance of the partial correlation coefficient is diffuse. Two different estimators of the sampling variance are frequently used in practice.
- Our critical reflection and assessment of the statistical properties of the estimators show that the estimator derived in Olkin and Siotani (1976) and Anderson (1984) is preferred for meta-analyzing partial correlation coefficients.

Potential impact for RSM readers outside the authors' field

- The quality of meta-analyses on partial correlation coefficients will be improved if researchers are using the recommended estimator of the sampling variance of the partial correlation coefficient.

1 | INTRODUCTION

Meta-analysts often want to synthesize the results of a linear regression model where the effect of interest is the relation between an independent variable and a dependent variable when controlling for other variables in the model. Meta-analyses on these so-called partial effects are becoming more common[1] and are regularly conducted in, for example, psychology[2–6] and economics.[7–11] The partial correlation coefficient (PCC) can be used as effect size measure in these meta-analyses. The PCC quantifies the relationship between the independent and dependent variables where there is controlled for the effect of the other variables in both the independent and dependent variable.[12,13] Each study's PCC needs to be accompanied by its sampling variance, because effect sizes in a meta-analysis are generally weighted by the inverse of the sampling variance. The existing literature is diffuse on how to compute the sampling variance of the PCC. Two different estimators are available in the literature and both are used in practice. The goal of this paper is to compare both estimators and provide recommendations for researchers on how to compute the sampling variance of a PCC.

2 | STATISTICAL MODEL

Let N be the number of independent observations with Y_i denoting the observed score on the dependent variable of participant i. We write the population linear regression model as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_M X_{iM} + \epsilon_i$$

where β_0 is the intercept, $\beta_1 X_{i1}$ is the regression coefficient of the independent variable X_1, $\beta_M X_{iM}$ refers to the $m = 1, \cdots, M^{\text{th}}$ regression coefficient of the independent variable X_m, and ϵ_i is the sampling error that is assumed to follow a normal distribution with mean zero and variance σ^2.

3 | PARTIAL CORRELATION COEFFICIENT

The PCC between variables Y and X_1 controlled for variable X_2 can be estimated using[12–15]

$$r_p = \frac{r_{Y1} - r_{Y2} r_{12}}{\sqrt{1 - r_{Y2}^2}\sqrt{1 - r_{12}^2}} \tag{1}$$

where the Pearson correlation coefficients between Y and X_1, Y and X_2, and X_1 and X_2 are denoted by r_{Y1}, r_{Y2}, and r_{12}, respectively. Estimating Equation (1) can only be used when controlling for a single variable. An alternative estimating equation for the PCC between Y and X_1 that allows for controlling for one or more variables is[16–18]

$$\sqrt{\frac{b_1^2}{b_1^2 + \text{Var}[b_1] df}} = \sqrt{\frac{t_1^2}{t_1^2 + df}} \tag{2}$$

where b_1 is the estimate of β_1, df refers to the degrees of freedom that are equal to $N - M - 1$, and t_1 is the t-statistic of the test $H_0 : \beta_1 = 0$. Estimating Equation (2) is especially useful in the context of meta-analysis, because Pearson correlation coefficients between all variables included in the model are usually not reported in the primary studies and it allows for more than one control variable.

Fisher[19] noted that the probability density function (PDF) of the PCC is the same as of a Pearson correlation coefficient except for the degrees of freedom. Hence, the exact PDF of the PCC is (equation (25) of Hotelling[20])

$$f(r_p | \rho; df) = \frac{df}{\sqrt{2\pi}} \frac{\Gamma(df+1)}{\Gamma(df+1.5)} (1-\rho^2)^{0.5(df+1)} \left(1 - r_p^2\right)^{0.5(df-2)} (1 - \rho r_p)^{-df-0.5} {}_2\text{F}_1\left(0.5, 0.5, df+1.5, \frac{1+\rho r_p}{2}\right) \tag{3}$$

where ρ is the PCC in the population, the degrees of freedom (i.e., df) are equal to $N-M-1$, Γ is the gamma function, and ${}_2F_1$ is the Gaussian hypergeometric function. The exact PDF if $\rho=0$ simplifies to a Student's t-distribution and is commonly used for testing $H_0: \rho=0$,[21]

$$\frac{r_p}{\sqrt{\left(1-r_p^2\right)/df}} \sim t_{df}. \tag{4}$$

4 | SAMPLING VARIANCE OF THE PARTIAL CORRELATION COEFFICIENT

We continue with describing two popular estimators for estimating the sampling variance of the PCC before we reflect on their statistical properties.

4.1 | Estimators

The first estimator of the sampling variance of the PCC that we describe has been derived in Olkin and Siotani[14] and Chapter 4 of Anderson.[22] Its estimating equation is

$$s_1^2 = \frac{\left(1-r_p^2\right)^2}{df}. \tag{5}$$

Note that the numerator in estimating Equation (5) is equal to the numerator of the sampling variance of the Pearson correlation coefficient (e.g., see equation (11.35) in Borenstein and Hedges[23]). Only the denominators differ, because the number of regression coefficients (i.e., M) are included in the degrees of freedom that are in the denominator of estimating Equation (5). This estimator is the large sample approximation of the variance of the PCC if r_p is replaced by ρ in Equation (5).

The second estimator is reported on page 25 of the popular book on meta-analysis by Stanley and Doucouliagos.[17] Its estimating equation is

$$s_2^2 = \frac{1-r_p^2}{df}. \tag{6}$$

The square root of this second estimator is actually equal to the standard error used in Equation (4) for computing the t-statistic. This implies that this estimator is derived conditional on $\rho=0$.

Preliminary observations when comparing both estimators are that the difference between the estimators s_1^2 and s_2^2 is in the numerators where the term $\left(1-r_p^2\right)$ is squared in s_1^2 but not in s_2^2. This implies that both estimators yield the same estimates if $r_p=0$. Furthermore, it also implies that $s_1^2 < s_2^2$ for $r_p \neq 0$, because the numerator of s_1^2 is always smaller than the numerator of s_2^2.

4.2 | Comparing the estimators

We compared s_1^2 and s_2^2 with the variance obtained using the exact PDF in (3) for a PCC between two variables while controlling for one variable (i.e., $M=2$). That is, we computed s_1^2, s_2^2, and the variance using the exact PDF for 1000 equidistant values of ρ that range from -0.99 to 0.99. We assumed that the PCC is an unbiased estimate of ρ* and computed s_1^2, s_2^2 by replacing r_p in Equations (5) and (6) by ρ. The selected levels of the sample size were $N=8;16;32;64$. The variance based on the exact PDF was obtained by computing the second moment of the PDF,

$$\int_{-1}^{1} \left(r_p-\rho\right)^2 f\left(r_p|\rho;df\right) dr_p.$$

The software R[27] (Version 4.2.2) was used for the analyses. The Gaussian hypergeometric function implemented in the R package "gsl"[28] (Version 2.1.7.1) was used for evaluating the exact PDF. R code of the analyses is available at https://osf.io/pqvyx.

Figure 1 shows the variance of the PCC for the different estimators where each subfigure contains the results of a particular sample size N. The properties of the different estimators were most manifest for the top-left subfigure with $N=8$. The variances of the PCC using s_1^2 (green dashed line) and s_2^2 (orange dashed line) were equal to each other for $\rho=0$ but larger than the variance based on the exact PDF (black solid line). If $\rho \neq 0$, s_2^2 approached the variance based on the exact distribution whereas s_1^2 was always larger than the variance based on the exact PDF and s_2^2. These patterns are also apparent in the other subfigures of Figure 1 with larger N. The differences between both s_1^2 and s_2^2 and the variance based on the exact PDF decreased as a function of N where s_1^2 approached the variance based on the exact PDF more rapidly than s_2^2.

5 | EXAMPLE

We computed s_1^2 and s_2^2 for the meta-analysis by Craft et al.[29] on the PCC between self-confidence and sports performance when controlling for cognitive and somatic anxiety. These data were obtained from table 16.2 of

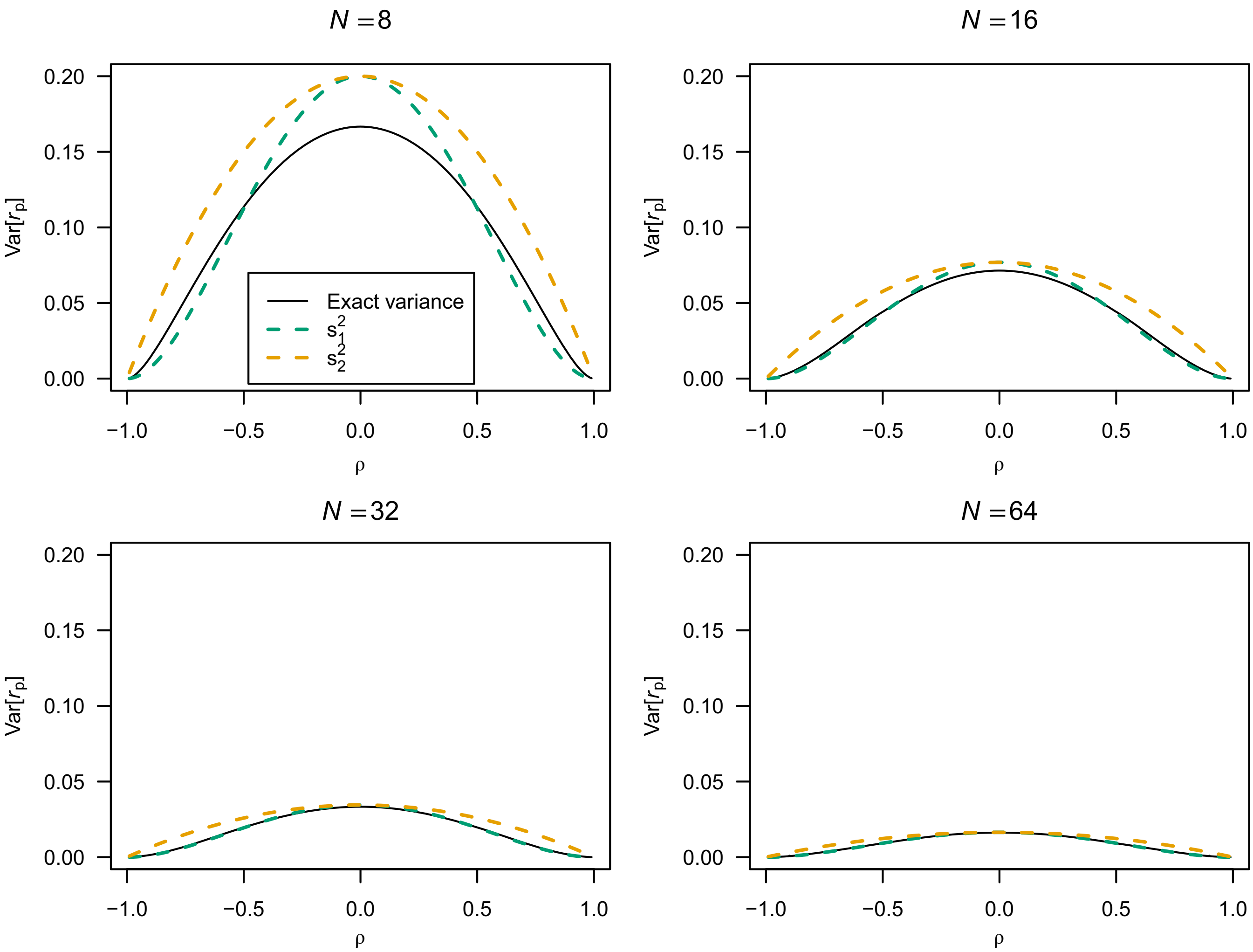

FIGURE 1 Results of analyses examining the statistical properties of estimating the variance of the PCC based on the exact PDF (solid black line, Equation 3) and estimators s_1^2 (green dashed line, Equation 5) and s_2^2 (orange dashed line, Equation 6). Different subfigures refer to different total sample sizes (i.e., N). [Colour figure can be viewed at wileyonlinelibrary.com]

Becker and Aloe.[1] Table 1 shows these data together with s_1^2 and s_2^2 and the corresponding standard errors (i.e., $\sqrt{s_1^2}$ and $\sqrt{s_2^2}$). R code of this analysis is available at https://osf.io/gtcux.

The results confirm the earlier observations of section 4.2. First, $s_1^2 < s_2^2$ for all studies. Second, the difference between s_1^2 and s_2^2 were small if the PCC was close to zero (e.g., IDs 10 and 38). However, the difference between s_1^2 and s_2^2 was larger if the PCC was not close to zero. For example, $r_p = 0.654$ for the study with ID = 22, and s_2^2 was almost twice as large as s_1^2 (0.0060 vs. 0.0034).

6 | DISCUSSION

PCCs are frequently used as effect size measure in a meta-analysis. We have reflected on two estimators of

TABLE 1 Data of eight studies of the meta-analysis by Craft et al.[29]

ID	N	r_p	s_1^2	s_2^2	$\sqrt{s_1^2}$	$\sqrt{s_2^2}$
1	142	0.536	0.0037	0.0052	0.0607	0.0719
3	37	0.332	0.0240	0.0270	0.1549	0.1642
10	14	−0.070	0.0990	0.0995	0.3147	0.3155
22	100	0.654	0.0034	0.0060	0.0584	0.0772
26	51	0.044	0.0212	0.0212	0.1456	0.1457
28	128	0.247	0.0071	0.0076	0.0843	0.0870
36	70	0.434	0.0100	0.0123	0.0999	0.1109
38	30	−0.024	0.0384	0.0384	0.1960	0.1961

Note: ID = study identifier, N = total sample size; r_p = estimated partial correlation coefficient; s_1^2 = estimated variance with Equation (5); s_2^2 = estimated; variance with Equation (6); $\sqrt{s_1^2}$ = square root of s_1^2; $\sqrt{s_2^2}$ = square root of s_2^2.

the sampling variance of the PCC and examined their properties. This revealed that the estimator proposed in Stanley and Doucouliagos[17] was derived under the assumption of a zero PCC in the population. This estimator is especially biased when the PCC in the population is different from zero and the sample size is small. Hence, we recommend researchers to use the estimator that was derived in Olkin and Siotani[14] and Anderson,[22] because this estimator can be used for a zero and non-zero PCC in the population.

Using a suboptimal estimator of the sampling variance biases the results of a meta-analysis, because the inverse of the variances are typically used as weights. Another reason why accurate estimation of the sampling variance is of importance is that the sampling variance or the square root of the sampling variance (i.e., standard error) are often used for assessing small-study effects. Small-study effects refer to the tendency of small studies to go along with larger effect sizes, and one of the possible causes of small-study effects is publication bias.[30,31] Commonly used methods to test and correct for small-study effects are Egger's regression test[30] and PET-PEESE[32] that require researchers to include the sampling variance or its square root as moderator in a meta-regression model.

We have focused on the PCC as effect size measure. An alternative option is to not meta-analyze PCCs directly, but first apply Fisher's z transformation and use the transformed PCCs as effect size measure in the meta-analysis.[18,22] This is analogous to how Pearson correlation coefficients are frequently meta-analyzed. A desirable property of this Fisher's z transformation is that the sampling distribution of a study's transformed effect size is approximately normally distributed. It is especially beneficial to apply the Fisher's z transformation for PCCs that are not close to zero, because the normality assumption is then more likely violated when meta-analyzing untransformed PCCs. Another desirable property of the Fisher's z transformation is that the sampling variance is independent of the PCC. This sampling variance of Fisher's transformed PCCs can be estimated with $1/(N-3-M-1)$ where $M-1$ are the number of control variables.[19,22]

To summarize, the frequently used estimating equation of the sampling variance of the PCC by Stanley and Doucouliagos[17] should only be used in a meta-analysis if the PCC in the population is zero. Hence, we recommend to abandon this estimator and use the estimator derived in Olkin and Siotani[14] and Anderson[22] instead.

ACKNOWLEDGMENTS

We would like to thank Prof. Dr. Tom Stanley for providing feedback on an earlier version of this paper.

FUNDING INFORMATION

Robbie C.M. van Aert is supported by a Veni grant financed by the Dutch Research Council (NWO). Grant Number: VI.Veni.211G.012.

DATA AVAILABILITY STATEMENT

Data of the example in section 5 were obtained from Table 16.2 of Becker and Aloe. R codes of the analyses in section 4.2 and section 5 are available at https://osf.io/pqvyx and https://osf.io/gtcux, respectively.

ORCID

Robbie C. M. van Aert https://orcid.org/0000-0001-6187-0665

ENDNOTE

* This assumption is violated in practice, because the estimator of the Pearson correlation coefficient is known to have a small negative bias,[24–26] and Pearson's correlation coefficients are used for estimating the PCC (see Equation (1)).

REFERENCES

1. Becker BJ, Aloe AM. *Model-Based Meta-Analysis and Related Approaches*. 3rd ed. Russell Sage Foundation; 2019: 339-363.
2. Polanin JR, Espelage DL, Grotpeter JK, et al. A meta-analysis of longitudinal partial correlations between school violence and mental health, school performance, and criminal or delinquent acts. *Psychol Bull*. 2021;147(2):115-133. doi:10.1037/bul0000314
3. Peng P, Lin X, Ünal ZE, et al. Examining the mutual relations between language and mathematics: a meta-analysis. *Psychol Bull*. 2020;146(7):595-634. doi:10.1037/bul0000231
4. Chiang JJ, Lam PH, Chen E, Miller GE. Psychological stress during childhood and adolescence and its association with inflammation across the lifespan: a critical review and meta-analysis. *Psychol Bull*. 2022;148(1–2):27-66. doi:10.1037/bul0000351
5. Curran T, Hill AP. Young people's perceptions of their parents' expectations and criticism are increasing over time: implications for perfectionism. *Psychol Bull*. 2022;148(1–2):107-128. doi:10.1037/bul0000347
6. Righetti F, Sakaluk JK, Faure R, Impett EA. The link between sacrifice and relational and personal well-being: a metaanalysis. *Psychol Bull*. 2020;146(10):900-921. doi:10.1037/bul0000297
7. Anwar AI, Mang CF. Do remittances cause Dutch disease? A meta-analytic review. *Appl Econom*. 2022;54(36):4131-4153. doi:10.1080/00036846.2021.2022091
8. Sun Z, Zhu D. Investigating environmental regulation effects on technological innovation: a meta-regression analysis. *Energy Environ*. 2021:1-30. doi:10.1177/0958305X211069654
9. Filomena M, Picchio M. Retirement and health outcomes in a meta-analytical framework. *J Econom Surv*. 2022;joes.12527. doi:10.1111/joes.12527
10. Floridi A, Demena BA, Wagner N. Shedding light on the shadows of informality: a meta-analysis of formalization

interventions targeted at informal firms. *Labour Econom*. 2020; 67:101925. doi:10.1016/j.labeco.2020.101925
11. Geyer-Klingeberg J, Hang M, Rathgeber AW. What drives financial hedging? A meta-regression analysis of corporate hedging determinants. *Int Rev Financ Anal*. 2019;61:203-221. doi:10.1016/j.irfa.2018.11.006
12. Cohen J, Cohen P, West SG, Aiken LS. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd ed. Lawrence Erlbaum Associates; 2003.
13. Pedhazur EJ. *Multiple Regression in Behavioral Research: Explanation and Prediction*. 3rd ed. Harcourt Brace College Publishers; 1997.
14. Olkin I, Siotani M. Asymptotic distribution of functions of a correlation matrix. In: Ikeda S, ed. *Essays in Probability and Statistics*. Shinko Tsusho; 1976:235-251.
15. Kutner MH, Nachtsheim C, Neter J, Li W. *Applied Linear Statistical Models*. fifth ed. McGraw-Hill Irwin; 2005.
16. Gustafson RL. Partial correlations in regression computations. *J Am Stat Assoc*. 1961;56(294):363-367. doi:10.1080/01621459.1961.10482120
17. Stanley TD, Doucouliagos H. *Meta-Regression Analysis in Economics and Business*. Routledge; 2012.
18. Aloe AM, Thompson CG. The synthesis of partial effect sizes. *J Soc Soc Work Res*. 2013;4(4):390-405. doi:10.5243/jsswr.2013.24
19. Fisher RA. The distribution of the partial correlation coefficient. *Metron*. 1924;3:329-332.
20. Hotelling H. New light on the correlation coefficient and its transforms. *J R Stat Soc Ser B*. 1953;15(2):193-232. doi:10.1111/j.2517-6161.1953.tb00135.x
21. Levy KJ, Narula SC. Testing hypotheses concerning partial correlations: some methods and discussion. *Int Stat Rev/Revue Internationale de Statistique*. 1978;46(2):215. doi:10.2307/1402814
22. Anderson TW. *An Introduction to Multivariate Statistical Analysis*. second ed. Wiley; 1984.
23. Borenstein M, Hedges LV. Effect sizes for meta-analysis. In: Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis*. 3rd ed. Rusell Sage Foundation; 2019:207-244.
24. Olkin I, Pratt JW. Unbiased estimation of certain correlation coefficients. *Ann Math Stat*. 1958;29(1):201-211.
25. Johnson NL, Kotz S, Balakrishnan N. *Continuous univariate distributions*. second ed. Wiley; 1995.
26. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*. 1915;10(4):507-521. doi:10.2307/2331838
27. R Core Team. R: A Language and Environment for Statistical Computing. 2022.
28. Hankin RKS. Special functions in R: introducing the gsl package. *R News*. 2006;6(4):1-7.
29. Craft LL, Magyar TM, Becker BJ, Feltz DL. The relationship between the competitive state anxiety inventory-2 and sport performance: a meta-analysis. *J Sport Exerc Psychol*. 2003;25(1):44-65. doi:10.1123/jsep.25.1.44
30. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J*. 1997; 315(7109):629-634. doi:10.1136/bmj.315.7109.629
31. Sterne JAC, Becker BJ, Egger M. The funnel plot. In: Rothstein HR, Sutton AJ, Borenstein M, eds. *Publication Bias in Meta-Analysis: Prevention, Assessment and adjustments Chichester*. Wiley; 2005:73-98.
32. Stanley TD, Doucouliagos H. Meta-regression approximations to reduce publication selection bias. *Res Synth Methods*. 2014; 5(1):60-78. doi:10.1002/jrsm.1095

How to cite this article: van Aert RCM, Goos C. A critical reflection on computing the sampling variance of the partial correlation coefficient. *Res Syn Meth*. 2023;14(3):520-525. doi:10.1002/jrsm.1632