Psychophysiology, 37 (2000), 123–125. Cambridge University Press. Printed in the USA.

BRIEF REPORT

EOG correction: Which regression should we use?

RODNEY J. CROFT AND ROBERT J. BARRY
Department of Psychology, University of Wollongong, Australia

Abstract

Electrooculogram (EOG) correction is used to remove eye-movement-related contamination from electroencephalograms (EEG). Correction is reliant on the regression procedure, although when multiple EOG channels are used in the correction, the appropriate type of regression to use is not known. In the present study, we aimed to resolve this matter. Computer simulations were used to compare the simultaneous, multiple-stage, and single-channel regression methods of correction. EOG propagation was modeled on prior findings, under conditions of varying vertical and horizontal EOG (VEOG/HEOG) correlation. The dependent variable was the correlation between the uncontaminated and the corrected EEG. The simultaneous regression procedure gave the best correction, with its advantage increasing as a function of VEOG/HEOG correlation. It is recommended that the simultaneous regression procedure be used for EOG correction of the EEG.

Descriptors: EEG, EOG correction, Regression method, EOG artifact

The removal of eye-movement-related artifact from the electroencephalogram (EEG), termed *electrooculogram (EOG) correction*, is used in EEG research for a number of reasons. As is described more fully elsewhere (Croft & Barry, 1998), these reasons include a reduction in data loss, greater flexibility in task type, the secondary task of "not moving the eyes or blinking" becomes unnecessary, and the bias created by subjects' differing abilities to perform the secondary task becomes irrelevant.

EOG correction is reliant on the regression procedure. In the simplest case, this procedure estimates and removes the proportion of EOG that is present in the EEG using the "least squares" criterion. Formally, a coefficient ($\boldsymbol{B}$) is calculated using Eq. (1), where X_i and Y_i are the measured EOG and EEG voltages, respectively, at time point i.

$$\boldsymbol{B} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) / \sum (X_i - \bar{X})^2 \tag{1}$$

$$\mathrm{CEEG}_i = \mathrm{MEEG}_i - \boldsymbol{B} * \mathrm{EOG}_i - \mathrm{C} \tag{2}$$

The resultant $\boldsymbol{B}$ is then used to correct the EEG according to Eq. (2), where CEEG_i and MEEG_i are the corrected and measured EEG voltages, respectively, at time i, and C is the constant from the least-squares formula ($\mathrm{C} = \bar{Y} - [\boldsymbol{B} * \bar{X}]$).

It has been demonstrated that the use of two EOG channels in the correction procedure provides a better correction than one (Elbert, Lutzenberger, Rockstroh, & Birbaumer, 1985), and so *multiple* regression is generally used for correction. There are a number of ways of calculating $\boldsymbol{B}$s in multiple regression, each producing different corrections, and the manner in which EOG correction is affected by the type of multiple regression used has not been demonstrated.

The question of which regression procedure to use arises primarily from the problem of multicollinearity (for a discussion of which, see Cohen & Cohen, 1975). Put simply, in the simultaneous model, the $\mathbf{r}^2$ value for an EOG channel represents the effect of that EOG channel on the EEG *after* the effect of the other EOG channels has been removed. This means that shared variance between the EOG channels will not be included in the $\mathbf{r}^2$ values for any of the EOG channels. Therefore, when the EOG channels are correlated, the total $\mathbf{r}^2$ value is generally smaller than the sum of the first-order $\mathbf{r}^2$ values.

Woestenburg, Verbaten, and Slangen (1983) and van Driel, Woestenburg and van Blokland-Vogelesang (1989) attempted to overcome this problem by using a multiple-stage regression,[1] in which the first stage is to calculate a $\boldsymbol{B}$ from Eq. (1) using MEEG and one channel of EOG, and then to correct MEEG using Eq. (2). The corrected data are then used as the dependent variable in Eq. (1) with the second EOG channel, and then a second correction is applied using Eq. (2). Corrected data from Eq. (2) are reentered iteratively into Eq. (1) until there are no further EOG channels to

Address reprint requests to: Rodney J. Croft, Department of Psychology, University of Wollongong, Wollongong NSW 2522, Australia.

[1]Footnote 1 in that paper described the procedure on p. 132 as a stepdown regression function. We find this confusing, as a stepdown regression generally does not give extra weighting to any independent variable once it has been accepted into the equation (Cohen & Cohen, 1975). If the footnote is taken as a stipulation, then it contradicts the formulae given, and when both VEOG and HEOG channels are included in the equation, it becomes merely the simultaneous regression typically used in EOG correction.

enter. This method assigns any joint variance between EOG channels to the EOG channel entered first into the correction, and thus none is omitted.

However, it can be shown that where there is multicollinearity, the first stage of this correction will introduce EOG artifact, and the second stage cannot remove this artifact (Tuan, 1989). The overcorrection becomes more serious as both the degree of multicollinearity and the EOG magnitude increase. In contrast, it is only the *interpretation of* $\boldsymbol{r}^2$ *values* that is problematic with the simultaneous regression; even when the EOG channels are correlated, ***B*** should still be a good predictor of the EOG channels' effect. Correspondingly, many researchers use a form of simultaneous multiple regression (e.g., Gratton & Coles, 1989).[2]

It is important to determine empirically which method is more accurate, and to accomplish this task, computer simulations were used. These simulations compared the two methods under conditions of varying multicollinearity. It was predicted that the simultaneous would correct better than the multiple-stage regression. Because the difference between methods is based on their handling of multicollinearity, it was predicted also that the advantage of the simultaneous method would vary with multicollinearity.

Method

Simulations

To simulate EEG, a 50-data-point series was generated randomly (TEEG). To simulate HEOG, a 50-data-point series similar to a left–right eye movement was designed. To simulate VEOG, a 50-data-point series similar to a blink was designed. To vary the VEOG/HEOG correlation, the VEOG series was adjusted so that the maximum amplitude occurred randomly in the epoch. $\mathrm{VEOG_n}$ series were then grouped according to $\mathrm{HEOG/VEOG_n}$ correlation: MC1 contained those with correlations 0.00–0.22, MC2 = 0.23–0.44, MC3 = 0.45–0.66, and MC4 = 0.67–0.88. $\mathrm{MEEG_n}$ series were created by adding a proportion of the HEOG and $\mathrm{VEOG_n}$ series to the EEG. The proportions added were based on propagation factors presented by the Tilburg group at the Tilburg symposium (Brunia, Mocks, & van den Berg-Lenssen, 1989), and were 0.12 and 0.14 for VEOG, and 0.04 and 0.05 for HEOG, at C3 and C4, respectively.

Data Manipulation

Five types of regression were used to calculate the correction ***B***s: Simple regression with $\mathrm{VEOG_n}$ (VE) or $\mathrm{HEOG_n}$ (HE) as independent variable, simultaneous multiple regression (SIM), multiple-stage regression using $\mathrm{VEOG_n}$ and then $\mathrm{HEOG_n}$ (VE-HE), or $\mathrm{HEOG_n}$ and then $\mathrm{VEOG_n}$ (HE-VE) as independent variables. VE, HE, and SIM used $\mathrm{MEEG_n}$ as the dependent variable. VE and HE were performed according to Eq. (1), and SIM according to Eq. (3). Corrections were then performed according to Eq. (2) for VE and HE and according to Eq. (4) for SIM ***B***s, creating $\mathrm{C1EEG_n}$ to $\mathrm{C3EEG_n}$. For the multiple-stage regressions, ***B***s were calculated using $\mathrm{C1EEG_n}$ as the dependent variable and $\mathrm{HEOG_n}$ as the independent variable, and using $\mathrm{C2EEG_n}$ as the dependent variable and $\mathrm{VEOG_n}$ as the independent variable. These were then corrected using Eq. (2), creating $\mathrm{C4EEG_n}$ and $\mathrm{C5EEG_n}$, respectively.

[2]Miller, Gratton, and Yee (1988) also published an iterative procedure, but as they decorrelated the EOG channels at each step, their results will be similar to those from the simultaneous multiple regression, and the criticisms of the multiple-stage approach are not applicable to it.

$$\boldsymbol{B}_{\mathrm{yx.z}} = sd_{\mathrm{y}}/sd_{\mathrm{x}} * (r_{\mathrm{yx}} - r_{\mathrm{yz}} * r_{\mathrm{zx}}/1 - r_{\mathrm{xz}}^2) \tag{3}$$

$$\mathrm{CEEG_i} = \mathrm{MEEG_i} - (\boldsymbol{B}_{\mathrm{yx.z}} * \mathrm{EOG_{xi}}) - (\boldsymbol{B}_{\mathrm{yz.x}} * \mathrm{EOG_{zi}}) - \mathrm{C} \tag{4}$$

where y represents the EEG, x the EOG channel whose ***B*** we wish to calculate, and z an EOG channel that we wish to partial out of the analysis. A similar formula calculates $\boldsymbol{B}_{\mathrm{zy.x}}$ by substituting x for z and z for x. Terms for Eq. (4) are the same as for Eq. (2).

Statistical Analysis

Correlations were performed on each of the $\mathrm{TEEG_n}$-$\mathrm{C1EEG_n}$, $\mathrm{TEEG_n}$-$\mathrm{C2EEG_n}$, $\mathrm{TEEG_n}$-$\mathrm{C3EEG_n}$, $\mathrm{TEEG_n}$-$\mathrm{C4EEG_n}$, and $\mathrm{TEEG_n}$-$\mathrm{C5EEG_n}$ pairs. This calculation produced 20 correlations for each site, for each regression type. Due to the noninterval nature of correlation data, scores were normalized using Fisher's transformation. Matched planned contrasts compared the one-channel with the multiple-stage regression correlations, and the simultaneous to the multiple-stage regression correlations, as a function of the between-subjects factors laterality (C3 vs. C4) and multicollinearity (MC).

Results

As can be seen in Figure 1, simultaneous regression was superior to multiple-stage regression, $F(1,32) = 169.62$, $p < .001$, with the advantage interacting with multicollinearity, $F(3,32) = 4.90$, $p = .007$. The multiple-stage was superior to single-channel regression, $F(1,32) = 60.46$, $p < .001$, with the advantage again interacting with multicollinearity, $F(3,32) = 17.68$, $p < .001$. These effects did not interact with laterality, $F(3,32) < 0.01$, $p > .999$.

A follow-up analysis determined the levels of multicollinearity where the differences between simultaneous and multiple-stage, and multiple-stage and single-channel regressions occurred. This involved four within-subjects contrasts each, and so α was ad-

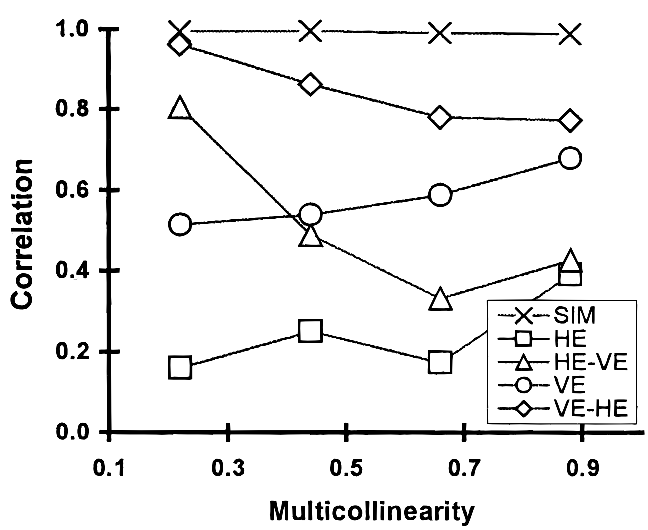

Figure 1. Mean correlations (vertical axis) between original and corrected electroencephalograms (EEG) are shown for each of the regression types, for each vertical and horizontal electrooculogram (VEOG/HEOG) correlation, for C3 and C4 combined. Error bars are omitted because they are too small to be easily seen. The simultaneous regression method is labeled "SIM," the multiple-stage methods "VE-HE" and "HE-VE," and the single-channel methods "VE" and "HE."

justed to 0.0125 using the Bonferroni adjustment. The simultaneous was superior to the multiple-channel regression at all levels of multicollinearity, $F(1,4) > 30.22$, $p < .005$, and the multiple-stage was superior to the single-channel regression at all levels of multicollinearity, $F(1,6) > 19.24$, $p < .005$ except 0.67–0.88, where it approached significance, $F(1,4) = 16.68$; $p = .015$.

The results show that the simultaneous is better than the multiple-stage regression at all levels of multicollinearity, and that the multiple-stage is better than the single-channel regression at all but the 0.67–0.88 multicollinearity range.

Discussion

The simultaneous multiple regression was the only method that gave accurate estimates of propagation and that result was not affected by multicollinearity. As was predicted, the multiple-stage methods performed worse with increasing multicollinearity, and significantly worse than the simultaneous method at all levels of multicollinearity. Further, consistent with previous findings (e.g., Elbert et al., 1985), it was demonstrated that where vertical and horizontal eye movement affect the EEG, both VEOG and HEOG channels should be used for correction.

It can be seen in Figure 1 that as multicollinearity approaches 0, the multiple-stage method gives good correction. It follows that this method is appropriate where there is no multicollinearity, such as may occur at central sites. However, as there was no advantage of this method here, its utility is questionable.

The regression method of correction has limitations. Of particular importance is that even if an accurate correction coefficient is obtained, an overcorrection may occur due to the removal of neural potentials from the EEG that have been recorded at the EOG (the problem of forward propagation). Methods have been proposed to overcome this problem, such as those based on source localization procedures (e.g., Berg & Scherg, 1991; Lins, Picton, Berg, & Scherg, 1993). These estimate the source of ocular artifact, which by definition is not contaminated by neural potentials, and remove the ocular artifact from the EEG analysis. The relative success of these methods will not be addressed here.

It is concluded that where the regression approach to EOG correction is used and multicollinearity occurs between the EEG and EOG channels, correction coefficients are best estimated with simultaneous multiple regression rather than single-channel or the multiple-stage regression of Woestenburg et al. (1983).

REFERENCES

Berg, P., & Scherg, M. (1991). Dipole models of eye-movements and blinks. *Electroencephalography and Clinical Neurophysiology*, *79*, 36–44.

Brunia, C. H. M., Mocks, J., & van den Berg-Lenssen, M. M. C. (1989). Correcting ocular artifacts in the EEG: A comparison of several methods. *Journal of Psychophysiology*, *3*, 1–50.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Croft, R. J., & Barry, R. J. (1998). EOG correction: A new perspective. *Electroencephalography and Clinical Neurophysiology*, *107*, 387–394.

Elbert, T., Lutzenberger, W., Rockstroh, B., & Birbaumer, N. (1985). Removal of ocular artifacts from the EEG—A biophysical approach to the EOG. *Electroencephalography and Clinical Neurophysiology*, *60*, 455–463.

Gratton G., & Coles, M. G. H. (1989). Generalization and evaluation of eye-movement correction procedures. *Journal of Psychophysiology*, *3*, 14–16.

Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in EEG and event related potentials: II. Source dipoles and source components. *Brain Topography*, *6*, 65–78.

Miller, G. A., Gratton, G., & Yee, C. M. (1988). Generalized implementation of an eye-movement correction procedure. *Psychophysiology*, *25*, 241–243.

Tuan, P. D. (1989). A comment from the viewpoint of time series analysis. *Journal of Psychophysiology*, *3*, 46–48.

van Driel, G., Woestenburg, J. C., & van Blokland-Vogelesang, A. W. (1989). Frequency domain methods: A solution for the problems of EOG-EEG contamination in ERPs. *Journal of Psychophysiology*, *3*, 29–34.

Woestenburg, J. C., Verbaten, M. N., & Slangen, J. L. (1983). The removal of the eye movement related artifact from the EEG by regression analysis in the frequency domain. *Biological Psychology*, *16*, 127–147.

(Received September 9, 1998; Accepted March 16, 1999)