Ho et al. reminds us that correlation analysis suffers from limitations in assessing agreement between methods. Such agreement should be assessed using Bland-Altman plots. This method is largely known as the original 1986 paper of Bland and Altman has been the most frequently cited article ever to appear in the Lancet and is one of the ten most frequently cited statistical articles ever (Bland and Altman, 2012).

Nevertheless, we decided to use Spearman's rank correlation, since our original intent was to compare methods and biological fluids, which is not exactly the same as assessing the degree of agreement. Ultimately, several conditions are required to evaluate agreement with Bland-Altman graphics. Differences of measurements between methods need to be normally distributed and the variability of the paired differences has to be uniform along the range of measurements (homoscedasticity). Heteroscedastic data should be transformed logarithmically or investigated with an analysis based on ranks (Atkinson and Nevill, 1998). Furthermore, the Bland-Altman method is based on a qualitative appreciation of concordance. Deciding whether the agreement between methods or samples is sufficient depends on the context in which the measurements are used (Bartlett and Frost, 2008). In psychiatry, the frequent absence of a gold standard flow chart for measuring of soluble proteins and the lack of tools providing the right absolute amount of the assayed protein renders any decision on appropriate limits of agreement for evaluating two relative quantification assays highly questionable.

Ho et al. also recommend the use of Bland-Altman plots to evaluate the stability of cytokine measurements in healthy individuals over time. They also state that the chance that cytokine levels remain constant over a 210-day period is unlikely. However, a biomarker that would discriminate psychiatric patients from the general population has to demonstrate both a higher inter-individual variability and a lower intra-individual variability in healthy subjects. To explore stability within subjects, we chose to calculate the Intra-Class Correlation coefficient (ICC) in our analyses, which is a well-recommended method (Liu et al., 2016). With respect to the aforementioned remark on the required stability of a biomarker over a long time period in healthy individuals, we do not see any advantage of segregating the calculation of ICC over different retest intervals.

Too many studies describing biomarkers only compared differences between groups without clear validation process of measurement for reproducibility and stability in healthy subjects. Correlations could be a first step to seek for disagreement, while other more specific analyses should be used when available technical tools reach consistency and provide validated measurements compared to absolute values.

References

Atkinson, G., Nevill, A.M., 1998. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 26, 217–238.

Bartlett, J.W., Frost, C., 2008. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet. Gynecol. 31, 466–475.

Bland, J.M., Altman, D.G., 2012. Agreed statistics: measurement method comparison. Anesthesiology 116, 182–185.

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., Tu, X.M., 2016. Correlation and agreement: overview and clarification of competing concepts and measures. Shanghai Arch. Psychiatry 28, 115–120.

Raoul Belzeaux [a,b,c,d]
[a] *Aix-Marseille Univ, CNRS, CRN2M-UMR7286, Marseille, France*
[b] *Pôle de Psychiatrie, Assistance Publique-Hôpitaux de Marseille, Marseille, France*
[c] *Fondation FondaMental, Fondation de Recherche et de Soins en Santé Mentale, Créteil, France*
[d] *McGill Group for Suicide Studies, Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Quebec, Canada*

Julia-Lou Consoloni [a,b,c]
[a] *Aix-Marseille Univ, CNRS, CRN2M-UMR7286, Marseille, France*
[b] *Pôle de Psychiatrie, Assistance Publique-Hôpitaux de Marseille, Marseille, France*
[c] *Fondation FondaMental, Fondation de Recherche et de Soins en Santé Mentale, Créteil, France*

El Chérif Ibrahim [a,b,*]
[a] *Aix-Marseille Univ, CNRS, CRN2M-UMR7286, Marseille, France*
[b] *Fondation FondaMental, Fondation de Recherche et de Soins en Santé Mentale, Créteil, France*

* Corresponding author at: Institut de Neurosciences de la Timone (INT), UMR 7289, 27 Bd Jean Moulin, 13385 Marseille Cedex 5, France.
E-mail address: el-cherif.ibrahim@univ-amu.fr (E.C. Ibrahim)

23 January 2017
http://dx.doi.org/10.1016/j.psyneuen.2017.02.020

Burnout and the hypothalamic-pituitary-thyroid axis: A methodological comment☆

Guo et al. (2017) examined the link between burnout and the activity of the hypothalamic-pituitary-thyroid (HPT) axis in a small sample ($N < 100$) of Chinese female nurses. The authors used the Maslach Burnout Inventory-General Survey (MBI-GS; Maslach et al., 1996) to assess burnout (score range: 0–6). The MBI-GS consists of three subscales, namely, exhaustion, cynicism, and (reduced) professional efficacy. For the purpose of categorical analyses, Guo et al. (2017) relied on cut-off scores derived from research conducted on burnout in Finland (Kalimo et al., 2003). The authors found no clear association between burnout and an altered functioning of the HPT axis. In our estimation, these null findings may be a by-product of an inadequate sampling of the study participants.

In analytical studies such as Guo et al.'s (2017) one, a key criterion must be met in order for the study to generate valid results: the inclusion of participants with various degrees of exposure—in the present case, of burnout symptoms (Kristensen, 1995, p. 21). That this criterion was met in Guo et al.'s (2017) study is unlikely. Indeed, while the authors have claimed that their participants showed "high levels of burnout symptoms" (p. 49), the mean scores of exhaustion, cynicism, and reduced professional efficacy in the study sample were 1.91, 1.67, and 1.75, respectively. Such scores correspond to burnout symptoms experienced on average *less than once a month* (Maslach et al., 1996). With standard deviations around 1, these scores suggest that the study sample mostly comprised healthy workers and under-included workers scoring at the high end of the burnout continuum. Consistent with this view, only

☆ The present correspondence paper concerns the following article: Guo, Y., Lam, L., Luo, Y., Plummer, V., Cross, W., Li, H., Yin, Y., Zhang, J., 2017. Female nurses' burnout symptoms: no association with the hypothalamic-pituitary-thyroid (HPT) axis. Psychoneuroendocrinology 77, 47–50.

two participants fell in the "severe burnout" category defined by the authors. In passing, we note that Guo et al. (2017) did not use nation-specific cut-off scores to categorize their participants with "severe burnout," a research practice that can be problematic (Schaufeli and Enzmann, 1998, p. 58).

Because (a) burnout overlaps with depression (Bianchi et al., in press) and (b) depression has been associated with an altered functioning of the HPT axis (e.g., Sullivan et al., 1997), changes in the activity of the HPT axis can be expected in burnout. Most probably, Guo et al.'s (2017) results are flawed by a severe form of the "healthy worker effect" (Schaufeli and Enzmann, 1998, p. 74).

References

Guo, Y., Lam, L., Luo, Y., Plummer, V., Cross, W., Li, H., Yin, Y., Zhang, J., 2017. Female nurses' burnout symptoms: no association with the hypothalamic-pituitary-thyroid (HPT) axis. Psychoneuroendocrinology 77, 47–50.
Kalimo, R., Pahkin, K., Mutanen, P., Toppinen-Tanner, S., 2003. Staying well or burning out at work: work characteristics and personal resources as long-term predictors. Work Stress 17 (2), 109–122.
Kristensen, T.S., 1995. The demand-control-support model: methodological challenges for future research. Stress Med. 11 (1), 17–26.
Maslach, C., Jackson, S.E., Leiter, M.P., 1996. Maslach Burnout Inventory Manual, third ed. Consulting Psychologists Press, Palo Alto.
Schaufeli, W.B., Enzmann, D., 1998. The Burnout Companion to Study and Practice: a Critical Analysis. Taylor & Francis, London.
Sullivan, P.F., Wilson, D.A., Mulder, R.T., Joyce, P.R., 1997. The hypothalamic-pituitary-thyroid axis in major depression. Acta Psychiatr. Scand. 95 (5), 370–378.

Renzo Bianchi*
Institute of Work and Organizational Psychology, University of Neuchâtel, Neuchâtel, NE, Switzerland

Irvin Sam Schonfeld
The City College of the City University of New York, New York, NY, USA

Eric Laurent
Bourgogne Franche-Comté University, Besançon, France

* Corresponding author.
E-mail addresses: renzo.bianchi@unine.ch (R. Bianchi), ischonfeld@ccny.cuny.edu (I.S. Schonfeld), eric.laurent@univ-fcomte.fr (E. Laurent).

16 January 2017

http://dx.doi.org/10.1016/j.psyneuen.2017.02.007

Reply to Letter to the Editor: Below is our response to the concerns raised by Dr. Bianchi et al. regarding our manuscript (Guo, Y., Lam, L., Luo, Y., Plummer, V., Cross, W., Li, H., Yin, Y., Zhang, J., 2017. Female nurses' burnout symptoms: no association with the hypothalamic-pituitary-thyroid (HPT) axis. Psychoneuroendocrinology 77,47–50.)

1. The null fingings may be a by-product of an inadequate sampling of the study participants

Response: this study is part of a large study on nurse burnout and one intervention's effect. The sample size is calculated basing on a large survey and the suggestion from some statisticians. For the large survey, the sample size calculation formula "$n = (u_{\alpha/2}\sigma/\delta)^2$ "has been used, the α was 0.05, $u_{\alpha/2}$ was 1.96. According to Xiao et al. (2012), the σ of emotional exhaustion, cynicism and reduced professional efficacy were 11.27, 5.80 and 9.60, respectively. And the control error δ of these three metrics were 0.95, 0.63, and 1.54, respectively. The largest sample size was accepted and it was 541. Then we consulted the statisticians, they suggested that ten to twenty percent of this sample size was enough to investigate the association between nurse burnout and HPT axis. We adopted 15% to calculate the size and it was 82. While, the sample size in this paper was 94.

2. The inclusion of participants with various degrees of exposure of burnout symptoms

Response: literature review shows that the most of stuff has mild-moderate burnout, and severe ones usually turnover or ask for a break. In this study, only two nurses experienced severe burnout. As the sample size was too small, we divided the participants into the non-burnout group and burnout group. And in the limitation part, we also mentioned that the relationship between severe burnout and thyroid hormones could not be demonstrated in this study.

3. Did not use the nation-specific cut-off scores to categorize the sample

Response: we accepted the explanation from Kalimo et al. (2003). "By the definition, burnout is a multifaceted three-dimensional syndrome and none of the metrics alone could indicate burnout. Therefore, each of the dimensions is, as such, unspecific to burnout". And we used the calculation method, developed by Kalimo et al. (2003), to compute the total score of the MBI-GS.

4. Results are flawed by a severe form of the "healthy worker effect"

Response: According to Shah (2009), time related factors (age at hire, age at risk), duration of employment, socioeconomic status and gender are factors affecting healthy worker effect (HWE). And the ways to minimize HWE are the follows: first, avoid using general population as a reference group and use active workers from another industry. Second, compare the health outcome between high exposure group and low or no exposure group. Third, conduct a study including participants in different seniority. Last, include the experience of every person who ever worked in a particular faculty. In this study, we recruited nurses from the same hospital and compared nurse burnout between burnout group and non-burnout group. Moreover, nurses in this study have different length of service, which could minimize the effect of age at risk.

References

Kalimo, R., Pahkin, K., Mutanen, P., Toppinen-Tanner, S., 2003. Staying well or burning out at work: work characteristics and personal resources as long-term predictors. Work Stress 17 (2), 109–122.
Shah, D., 2009. Healthy worker effect phenomenon. Indian J. Occup. Environ. Med. 13 (2), 77–79.
Xiao, B.J., Liu, K., You, L.M., et al., 2012. A survey about job burnout among Chinese nurses. Med. Philos. 33 (11B), 69–71.

Yufang Guo
Nursing Psychology Research Center of Xiangya Nursing School, Central South University, Changsha, 410013, Hunan, China