Still no support for the win-win effect when controlling for non-independence between countries of the same region

Toon Kuppens, University of Groningen, t.kuppens@rug.nl
Thomas V. Pollet, VU University Amsterdam

There has been a debate about whether or not a country's level of gender equality is related to the number of Olympic medals for female *and* male athletes (the "win-win" effect). There was the original Berdahl, Uhlmann, and Bai (2015) article, our commentary (Kuppens and Pollet, 2015), and a rejoinder by Bai, Uhlmann, and Berdahl (2015). This has been an interesting and open exchange of viewpoints and we hope that others will find it useful, too. Here we discuss a few remaining issues, respond to the criticism that our division of countries into regions is unscientific, and present new, advanced multilevel analyses that address some points of criticism and find no support for the hypothesis that athletes in more gender equal countries win more Olympic medals.

Should GDP per capita and population size be log-transformed?

Bai, Uhlmann, and Berdahl (2015) are correct in pointing out that it would be more consistent to log-transform both GDP per capita and population size, as both are very skewed. We (Kuppens & Pollet, 2015) had only transformed GDP per capita. Log-transforming population size as well shows stronger support for the win-win effect (but even then overall gender equality does not show a relation with male medals). However, when region is added to the model (either 10 regions as in Kuppens & Pollet, 2015, or the 22 UN regions as an objective division of countries into regions, see below) the only model that shows a significant positive relation between gender equality and Olympic medals is the one where the educational gender gap predicts female medals when controlling for 10 regions. This is not the case for the overall gender gap, the economic gender gap, the political gender gap, or the health gender gap, not the case for any of the models for male medals, and not the case when controlling for 22 rather than 10 regions. In sum, models using the log of the population size and controlling for region show little support for the win-win effect.

Is the effect of gender equality underestimated in models including GDP per capita as a control variable?

Bai et al. (2015) claim that the effect of gender equality is underestimated when controlling for GDP per capita because there also is an indirect effect of gender equality running through GDP per capita. This is technically correct but they do not mention that the bidirectional effect between gender equality and GDP per capita (see Dollar & Gatti, 1999) implies that the effect of GDP per capita is *also* underestimated in these models. Therefore, the effect of GDP per capita remains much larger than the effect of gender equality, and a large part of the total effect of gender equality actually occurs as an indirect effect through national wealth.

Additional analyses via machine learning support the view that GDP per capita is a substantially more important predictor of winning medals than any of the gender equality variables for both men and women's medals. These additional analyses can be found at https://dl.dropboxusercontent.com/u/14442317/Additional%20analyses%20BUB.zip .

Furthermore, from a theoretical point of view an indirect effect of gender equality through national wealth is incompatible with the proposed explanation of the gender equality effect in terms of gender stereotypes that impede both women's and men's performance. An indirect effect through wealth on medals is presumably simply due to more resources rather than stereotypes *directly* affecting athletic performance.

Is our division of countries into regions subjective?

Bai et al. (2015) call the world regions K&P (2015) use "arbitrarily-defined," "based on folk stereotypes," and "based on naïve stereotypes." These are strong words and we think that Bai et al. mean that our division into regions is *subjective*, that is, that others could make different regions and that we do not show that our division is better than others or based on objective criteria. This is true. In fact, we (Kuppens & Pollet, 2015) explicitly acknowledged this in the supplemental material to our paper:

> "The division in world regions is one we made ourselves, based on our assessment of geographical, cultural, and historical differences and similarities between countries. We did this because we could not find any official or widely used classification of countries into world regions. Without any doubt, other researchers would make different divisions in world regions, and we do not claim that our classification is better than other classifications. It is therefore important to show that our results do not depend on the particular division into world regions that one uses, as long as the regions do not become so large that they are too heterogeneous (like lumping together all of America or all of Europe)."

This was then followed by analyses using two alternative divisions into regions (one being the one we had used previously, see Kuppens & Pollet, 2014), and showing very similar results to the analysis reported in the main text. So, in our reply, we have already acknowledged this issue and showed that the exact division of countries into regions does not matter much. Also, the fact that the division in 10 regions explains half the between-country variance in gender equality means that we must have been doing something right. We can only note that Bai et al. (2015) did not make any suggestions either about what a scientific division in regions should look like. In short, the criticism of Bai et al. on our choice for a certain division into regions is an issue we had explicitly acknowledged and one that does not affect the results, and they do not make suggestions for how to improve it.

Nonetheless, there is one suggestion for improvement that we would like to make ourselves. We used 10 regions because reviewers of similar previous analyses (Kuppens & Pollet, 2014) found the use of a large number of regions too much. We felt that 10 was as low as we could

go without making the regions too heterogeneous. One alternative would be to use the 22 "UN geographical regions for statistical use" (see http://unstats.un.org/unsd/methods/m49/m49regin.htm). One advantage is that these are less arbitrary because they were defined by the UN rather than by individual researchers. A second advantage is that a larger number of regions also allows an analysis at the regional level, instead of or in addition to the national level. How does this play out for the equality-medals relation? There is no relation between gender equality and Olympic medals when controlling for GDP per capita and the 22 UN regions (both population and GDP per capita were logged for this analysis). See below for details. In sum, we think the findings we present are fairly robust to changes in the division of countries in regions.

Should researchers account for non-independence of countries in cross-national analyses?

Yes, some attempt should be made to account for non-independence, and controlling for region could be a first step. Bai et al.'s (2015) reasoning that dependence between students in classes should be solved differently than dependence between countries in regions, is unconvincing. It does not matter how exactly the nesting is defined: the proof of the pudding is in the eating. If countries close to one another tend to be more similar, they are not independent data points (for whatever reason) and this violates a key assumption of statistical models in which such dependence is not taken into account (Fox, 2008; Gelman & Hill, 2007; Snijders & Bosker, 2012). This, in turn, leads to more false positive results, an increase in the type I error rate. Any introductory multilevel modelling textbook will explain this in detail.

Adding region to the model changes the results in important ways. Here we briefly report models that include the log of GDP per capita, the log of population, and region (either 10 regions as in Kuppens & Pollet, 2015, or the 22 UN regions as an objective division of countries into regions, see above). The only model that shows a significant positive relation between gender equality and Olympic medals is the one where the educational gender gap predicts female medals when controlling for 10 regions. This is not the case for the overall gender gap, the economic gender gap, the political gender gap, or the health gender gap, not the case for male medals, and not the case when controlling for 22 rather than 10 regions. In sum, models using the log of the population size and controlling for region show little support for the win-win effect.

Is using dummy variables that code for region the ultimate solution?

Using dummy variables to code for world regions is the same as a fixed effect approach or analysis of covariance model (Snijders & Bosker, 2012). Importantly, this is not the only and not the best way to control for the dependence of countries within regions. In the discussion we wrote: "Many solutions to this problem exist […] but a simple and first robustness check can consist of controlling for region (e.g., Kuppens & Pollet, 2014), or analyzing the data at a higher level (e.g., region). If such an analysis does not uphold the statistical patterns, then geographical or cultural clustering could drive the found effect and it is less likely that it reflects a real phenomenon." In other words, controlling for region is just a robustness check

that is easily carried out. We reiterate that other researchers have accounted for non-independence via other methods, for example via language phylogenies (Mace & Pagel, 1994), but one thing that is clear is that non-independence of observations exists in these data.

Is a more advanced analysis of the Olympic medals data possible?

Yes. Given the obvious relevance of regions, one thing we could do is fit a multilevel model of countries nested within regions. We performed such an analysis and used the 22 UN regions (see above) as an objective rather than subjective division in regions. We used grand mean centering for the gender gap scores and GDP per capita, and added region-level gender gap scores and GDP per capita as predictors at the regional level. In the calculation of the region-level predictors we took into account the population size of each country. We estimated multilevel Poisson models including a parameter for overdispersion, and estimated these models with Bayesian Markov Chain Monte Carlo (MCMC) estimation in MLwiN 2.29. This type of estimation is known to work well for small samples (e.g., Stegmueller, 2013). We also included the log of the population size, as well as the other control variables present in the original Berdahl et al. (2015) article. We conducted these analyses for the overall gender gap, the economic gender gap, and the education gender gap separately, as these are the most promising in terms of finding support for the win-win effect. The coefficients of the gender gap variables are reported in Table 1.

Table 1: Multilevel regression coefficients and standard errors for gender gap scores

		Country level		Regional level	
	Gender gap	B	SE	B	SE
Female medals	Overall	-0.362	0.418	**24.280**	10.213
	Economic	0.192	0.428	**8.770**	4.178
	Educational	-0.103	0.992	12.610	12.603
Male medals	Overall	-0.067	0.343	4.615	5.746
	Economic	-0.065	0.398	4.358	3.424
	Educational	0.296	0.824	9.765	9.443

Note: Coefficients have 95% confidence intervals not including zero when they are nearly twice as large as their standard error. These coefficients appear in boldface in the table.

At the country level, none of the relations between gender gap scores and medals are significant, and four out of six coefficients have a negative sign (see Table 1). We thus find no support for the hypothesis that athletes from more gender equal nations win more Olympic medals. This is important because in this model we addressed the criticism from Bai et al. (2015): we used an objective division in regions, log-transformed both GDP per capita and population, and we did not use dummy variables to code for region (we used a random effect in a multilevel model instead). At the same time, these models also do not provide support for a negative relation between gender equality and medals for male athletes. We did report such a negative relation previously (Kuppens & Pollet, 2015), but we did not interpret it as being meaningful (we wrote: "We do not, however, want to conclude that gender equality impedes men's performance at the Olympic games"). In hindsight, this was a wise decision.

It is worth noting that the coefficients in this model do not represent within-region relations only, as is the case when using region as a fixed effect (i.e. using dummy variables). Instead, the country-level coefficients in Table 1 represent relations between countries all over the world (just as the original Berdahl et al. analysis did), while at the same time controlling (imperfectly) for the statistical non-independence of countries through the use of multilevel modelling.

At the regional level, the coefficients for gender gap are consistently positive. They are statistically significant for female medals for the overall gender gap and the economic gender gap scores. Finding a relation between gender equality and female medals at the regional level is one thing, interpreting it is another. We cannot rely on the explanation in terms of stereotypes that was proposed by Berdahl et al. (2015) because there is no reason why this mechanism would work at the regional level and *not* at the national level (the results imply that the relation is stronger at the regional compared to the national level). In other words, the explanation should be one that applies to differences between regions more than it applies to differences between countries. Our hunch is that regional gender equality is related to other broad cultural/political/economic differences that affect the number of medals for female athletes. Suggesting a specific mechanism through which this occurs based on region-level correlational data is difficult.

Finally, it is also important to note that other analytical approaches exist. We do not claim that this is the only or optimal approach. Some researchers might for example rely on geographical coordinates, others might have relied on a language or cultural phylogeny (see Mace & Pagel, 1994), still others might have attempted modeling the non-independence via Generalized Estimating Equations (Barthes, Crochet, & Raymond, 2015). We did attempt one alternative approach, machine learning, which does not rely on the assumption of non-independence, and this approach does not provide support for gender equality variables as viable predictors of medals won in men and women, when considering geographical and economic variables (see https://dl.dropboxusercontent.com/u/14442317/Additional%20analyses%20BUB.zip).

Concluding remarks

Controlling for GDP per capita reduces the win-win effect but there still is a significant relation between the economic dimension of gender equality and Olympic medals won by female and male athletes (see Bai et al., 2015). However, controlling for region reduces all effects of gender equality (including all subdimensions) to non-significance and reverses the sign of most effects. The multilevel analyses that we present here are better than the ones presented previously. They also do not represent purely within-region relations but relations between countries from all over the world. These analyses are still slightly liberal in terms of the win-win effect because they only control imperfectly for the dependence of countries that are situated close to one another. Yet even then, not a single effect of gender equality could be found at the national level. If there are relations between gender equality and Olympic medals, our results suggest that they are limited to female medals and exist only at the

regional level, but not at the *national* level, which makes interpretations in terms of a win-win effect problematic. We welcome further analyses, but for now we conclude that, in our view, the evidence for a win-win effect at the national level in these data is very weak.

Acknowledgements

We would like to think Eric Uhlmann for his comments on a previous version. Thomas Pollet is supported by a grant from the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS).

References

Barthes, J., Crochet, P.-A., & Raymond, M. (2015). Male homosexual preference: Where, when, why? *PLoS One*, 10(8): e0134817. doi: 10.1371/journal.pone.0134817

Berdahl, J. L., Uhlmann, E. L., & Bai, F. (2015). Win–win: Female and male athletes from more gender equal nations perform better in international sports competitions. *Journal of Experimental Social Psychology*, 56, 1–3. doi:10.1016/j.jesp.2014.08.003

Bai, F., Uhlmann, E. L., & Berdahl, J. L. (2015). The robustness of the win-win effect. *Journal of Experimental Social Psychology*

Dollar, D., & Gatti, R. (1999). Gender inequality, income, and growth: Are good times good for women? Policy research report on gender and development working paper series ; no. 1. Washington, DC: The World Bank. Retrieved from http://documents.worldbank.org/curated/en/1999/05/443556/gender-inequality-income-growth-good-times-good-women

Fox, J. (2008). *Applied regression analysis and generalized linear models* (Second edition). Los Angeles, CA: Sage.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, NY: Cambridge University Press

Kuppens, T., & Pollet, T. V. (2014). Mind the level: problems with two recent nation-level analyses in psychology. *Frontiers in Psychology*, 5, 1110. doi:10.3389/fpsyg.2014.01110

Kuppens, T., & Pollet, T. V. (2015). Gender equality probably does not affect performance at the Olympic games: A comment on Berdahl, Uhlmann, and Bai (2015). *Journal of Experimental Social Psychology*.

Mace, R., & Pagel, M. (1994). The comparative method in anthropology. *Current Anthropology*, *35*, 549–564.

Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel analysis* (Second edition). Los Angeles: Sage.

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and bayesian approaches. *American Journal of Political Science*, 57(3), 748–761. doi:10.1111/ajps.12001