The Robustness of the Win-Win Effect

Feng Bai

University of British Columbia

Eric Luis Uhlmann

INSEAD

Jennifer L. Berdahl

University of British Columbia

WORD COUNT: 2,500 (Main Text), 149 (Abstract)

REFERENCES: 20

CORRESPONDING AUTHOR:

Jennifer L. Berdahl
Sauder School of Business
University of British Columbia
2053 Main Mall
Vancouver, British Columbia
Canada V6T 1Z2
Tel: 1-604-827-2759
E-mail: jennifer.berdahl@sauder.ubc.ca

Abstract

We demonstrate that positive relationships between measures of national gender equality and Olympic medal wins are robust across a variety of appropriate statistical approaches to analyzing cross-national data. First demonstrated by Berdahl, Uhlmann, and Bai (2015), who controlled for GDP, population, latitude, and income inequality, we show that relationships between gender equality and medal wins remain positive when controlling for GDP per capita, consistently log-transforming positively skewed variables, and fully analyzing all four gender gap subindexes. The Win-Win effect is most robust for gender equality in education and earnings. Controlling for arbitrarily-defined world regions ("Anglo-Saxon countries" vs. "Africa") is inappropriate, as such groupings are based on folk stereotypes, not objective scientific criteria, and risks masking meaningful differences between countries. There is, however, often more than one right way to analyze a dataset; we discuss how this can be addressed by crowdsourcing the analysis of complex datasets prior to publication.

Kuppens and Pollet (2015; henceforth K&P) argue that the positive relationship between national gender equality and Olympic medal wins reported by Berdahl, Uhlmann, and Bai (2015) is nonsignificant and even reverses when controlling for GDP per capita and world regions. K&P inconsistently log transform variables, however, in a manner that artificially reduces the relationship between gender equality scores and medal wins. Specifically, they log transform GDP per capita to correct for positive skew, but not national population, which is also positively skewed. This is an easy mistake to make, but has a large effect on the degree of empirical support obtained for the Win-Win effect. As we demonstrate below, when GDP per capita and population are both log-transformed, significant positive relationships between measures of gender equality and medal wins remain.

In addition, K&P (2015) analyze the overall gender gap score from the World Economic Forum but only one of its four subindexes: educational gender equality. We demonstrate that when all four subindexes (educational, economic, health, and political gender equality) are fully analyzed, both educational and economic equality emerge as important predictors of medal wins.

Controlling for arbitrarily defined world regions, as K&P (2014, 2015) advocate, is not done in research on cross-national comparisons, and for good reason. Such groupings are based on folk stereotypes rather than objective scientific criteria, and risk obscuring real national differences.

Although we disagree with K&P's (2015) conclusion that gender equality is either unrelated or negatively related to Olympic medal wins, we acknowledge that there is often more than one right way to analyze a dataset. The Win-Win effect, like any other finding based on a complex set of interrelated variables, cannot remain significant at the $p < .05$ level no matter which statistical approach and control variables are used (Anderson & Anderson, 1996). As the field moves toward a new era of open data and scientific transparency, we can collectively address the issue of analysis-contingent results by crowdsourcing the analysis of complex datasets prior to publication (Silberzahn et al., 2015).

Measures of Gender Equality Positively Predict Medal Wins After Controlling for GDP Per Capita

As emphasized by Berdahl et al. (2015), any analysis controlling for national wealth when predicting an outcome from gender equality is inherently conservative. Gender equality predicts future economic growth: societies in which men and women are given more equal opportunities exploit their human capital more efficiently and therefore enjoy greater prosperity (Barsh & Yee, 2011; Chaaban & Cunningham, 2011; Inglehart & Norris, 2003; Lagerlöf, 2003; Löfström, 2009; Morrison, Raju, & Sinha, 2007; World Economic Forum, 2014). To the extent that gender equality contributes to national wealth *and* athletic success, controlling for national wealth underestimates the effects of gender equality on medal wins (S1 elaborates on this point in greater detail).

That said, it remains important to examine the relationship between national-level gender equality and Olympic medal wins after controlling for national wealth. Table 1 reports the zero-order correlations between all of the variables in the present analysis. Tables 2 to 5 summarize the results repeating K&P’s (2015) analyses when GDP per capita and national population are both log-transformed and all four gender gap subindexes are analyzed.

Table 1. Correlations Between Study Variables

Variable	1	2	3	4	5	6	7	8	9	10
1. Women's medals										
2. Men's medals	.82**									
3. Overall Gender Equality	.22*	.24**								
4. Educational Gender Equality	.19*	.23*	.58**							
5. Economic Gender Equality	.22*	.18*	.74**	.19*						
6. Health Gender Equality	.04	.07	.19*	.19*	.06					
7. Political Gender Equality	.06	.12	.74**	.16	.30**	.08				
8. Gini Index[a]	-.07	-.20*	-.10	.01	-.06	.07	-.15			
9. Population[b]	.41**	.36**	-.14	-.19*	-.22*	-.12	.09	-.05		
10. GDP per capita[b]	.37**	.47**	.47**	.61**	.20*	.22*	.23*	-.21*	-.12	
11. Latitude	.19*	.26**	.16	.21*	.06	-.01	.10	-.67**	.02	.40**

[a] Higher scores indicate more economic inequality.
[b] GDP per capita and population (in thousands) are log transformed.
Notes: *$p < .05$. **$p < .01$.

As seen in Tables 2 and 3, in quasipoisson regressions, overall gender equality significantly predicts Olympic medal wins for women (*b* = .33, *SE* = .13, *p* = .013) but not for men (*b* = .05, *SE* = .11, *p* = .643). Tables 4 and 5 show this pattern of results is similar using negative binomial regressions: Overall gender equality scores are significantly and positively related to medal wins for women (*b* = .83, *SE* = .24, $p < .001$), a relationship that is in the same direction but not significant for men (*b* = .27, *SE* = .18, *p* = .128). Thus, controlling for GDP per capita reduces the relationship between overall gender equality and medal wins to nonsignificance for men, but it does not reverse the sign of the effect. A higher overall level of gender equality in a society benefits female athletes, without hurting the success of male athletes.

This does not mean, however, that important forms of gender equality in a society never benefit male athletes. Gender equality is multi-dimensional, and for this reason Berdahl et al. (2015) fully analyzed the World Economic Forum's gender gap subindexes for educational, economic, health, and political gender equality. As seen in Table 1, educational equality and economic equality exhibit positive zero-order correlations with medal wins for both men and women, whereas equality in health and political representation do not.

As noted by Berdahl et al. (2015), the equality-medals effect is most strongly supported across different analytic approaches for the educational equality subindex. In quasipoisson regressions, educational equality predicts medal wins for both women (*b* = 1.42, *SE* = .49, *p* = .004) and men (*b* = 1.01, *SE* = .39, *p* = .010). In the negative binomial regressions, educational equality is likewise a significant predictor of medal wins for both women (*b* = .75, *SE* = .35, *p* = .031) and men (*b* = .72, *SE* = .33, *p* = .031).

In the analyses reported across Tables 2-5, economic gender equality also emerges as a positive predictor of athletic performance. As seen in Tables 2 and 3, in quasipoisson regressions the economic equality subindex predicts women's medals (b = .56, SE = .13, p < .001). This relationship is in the same direction but marginally significant for men (b = .20, SE = .11, p = .082). As seen in Tables 4 and 5, in negative binomial regressions, economic gender equality significantly predicts medal wins for both women (b = .97, SE = .22, p < .001) and men (b = .41, SE = .15, p = .005).

The final columns of Tables 2-5 report analyses with all four gender gap subindexes entered into the regression model simultaneously. These represent some of the most conservative tests of the "Win-Win" hypothesis, given that different types of gender equality correlate with each another. For educational and economic equality to predict medal wins, they must do so above-and-beyond each other as well as above-and-beyond gender equality in health and political representation. Despite this, in all regressions, educational and/or economic gender equality significantly and positively predict medal wins for male and female athletes.

It is clear from the analyses in Tables 1-5 that higher levels of gender equality in health outcomes and political representation do not facilitate the success of either male or female athletes at the Olympic games. When entered with the other three subindexes, the political equality subindex negatively relates to men's medal wins in the negative binomial model (Table 5). However, the zero-order correlation between political equality and medal wins for male athletes is positive in sign (Table 1) and political equality does not predict men's medal wins when entered by itself in the model (see Tables 3 and 5). This overall pattern of results suggests a null rather than a negative relationship between political equality and medal wins.

The heterogeneous results across gender gap subindexes underscore the multi-dimensional nature of gender equality. Some types of gender equality (education and economic) are a “Win-Win” for male and female athletes, whereas others (health and political representation) do not affect their success or failure.

In sum, when positively skewed variables are consistently log transformed and gender gap measures are fully analyzed, robust empirical support emerges for positive relationships between measures of gender equality and Olympic medal wins when controlling for GDP per capita (see S2 for the full set of analyses represented by the decision to log-transform GDP per capita and/or population). Female athletes from more gender equal countries overall win more medals. Gender equality in education and earnings predict more success at the Olympic games for *both* male and female athletes. Five of the six zero-order correlations between measures of gender equality and medal wins reported in Table 1 remain significant after controlling for GDP per capita, and all six remain positive in sign. Taken together, these results provide strong evidence that greater equality between men and women benefits their athletic performance.

Table 2: Results for quasipoisson regressions on women's medals won

	Overall Gender Equality			Educational Gender Equality			Economic Gender Equality			Health Gender Equality			Political Gender Equality			All Gender Equality Subindexes		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Intercept	.12	.20	.538	-.28	.28	.316	.23	.18	.196	.14	.20	.503	.16	.20	.436	.05	.22	.812
Latitude	-.09	.11	.408	-.10	.10	.323	-.09	.09	.338	-.06	.11	.586	-.06	.11	.620	-.09	.09	.285
GDP per capita	1.18	.15	<.001	1.03	.16	<.001	.99	.13	<.001	1.41	.16	<.001	1.37	.13	<.001	.89	.16	<.001
Population	1.27	.11	<.001	1.31	.12	<.001	1.25	.10	<.001	1.20	.11	<.001	1.22	.11	<.001	1.21	.11	<.001
Gini Index	-.13	.13	.291	-.31	.14	.031	-.25	.12	.039	-.12	.13	.367	-.15	.14	.278	-.35	.13	.006
Overall Gender Equality	.33	.13	.013															
Educational Gender Equality				1.42	.49	.004										.76	.39	.050
Economic Gender Equality							.56	.13	<.001							.55	.13	<.001
Health Gender Equality										-.06	.10	.580				-.08	.09	.386
Political Gender Equality													-.05	.10	.653	-.19	.11	.077

Notes: GDP per capita and population (in thousands) are log transformed.

Table 3. Results for quasipoisson regressions on men's medals won

	Overall Gender Equality			Educational Gender Equality			Economic Gender Equality			Health Gender Equality			Political Gender Equality			All Gender Equality Subindexes		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Intercept	.49	.17	.005	.22	.22	.331	.52	.17	.002	.47	.18	.008	.48	.17	.006	.28	.22	.215
Latitude	-.05	.10	.591	-.08	.09	.396	-.05	.09	.571	-.05	.10	.624	-.04	.10	.716	-.06	.09	.511
GDP per capita	1.22	.14	<.001	.98	.14	<.001	1.12	.14	<.001	1.29	.13	<.001	1.27	.12	<.001	1.03	.15	<.001
Population	.99	.10	<.001	1.05	.10	<.001	1.00	.10	<.001	.97	.10	<.001	.97	.10	<.001	1.00	.10	<.001
Gini Index	-.35	.12	.005	-.47	.13	<.001	-.38	.12	.002	-.35	.12	.005	-.40	.13	.002	-.51	.13	<.001
Overall Gender Equality	.05	.11	.643															
Educational Gender Equality				1.01	.39	.010										.82	.40	.041
Economic Gender Equality							.20	.11	.082							.17	.12	.169
Health Gender Equality										-.06	.10	.514				-.11	.09	.220
Political Gender Equality													-.12	.09	.195	-.18	.10	.066

Notes: GDP per capita and population (in thousands) are log transformed.

Table 4. Results for negative binomial models on women's medals won

	Overall Gender Equality			Educational Gender Equality			Economic Gender Equality			Health Gender Equality			Political Gender Equality			All Gender Equality Subindexes		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Intercept	.10	.17	.554	.16	.18	.368	.03	.16	.847	.26	.17	.129	.25	.17	.135	-.03	.17	.840
Latitude	.05	.17	.761	-.17	.18	.364	.02	.17	.904	-.05	.17	.796	-.03	.18	.864	-.09	.19	.632
GDP per capita	.75	.19	<.001	.77	.23	<.001	1.00	.16	<.001	1.15	.18	<.001	1.11	.19	<.001	.74	.24	.002
Population	1.29	.18	<.001	1.30	.19	<.001	1.37	.19	<.001	1.20	.19	<.001	1.20	.19	<.001	1.44	.19	<.001
Gini Index	-.40	.18	.022	-.47	.19	.015	-.23	.17	.165	-.31	.18	.079	-.32	.18	.075	-.36	.19	.056
Overall Gender Equality	.83	.24	<.001															
Educational Gender Equality				.75	.35	.031										.59	.34	.087
Economic Gender Equality							.97	.22	<.001							.93	.22	<.001
Health Gender Equality										-.05	.18	.790				-.02	.16	.915
Political Gender Equality													.06	.19	.735	-.05	.16	.758

Notes: GDP per capita and population (in thousands) are log transformed.

Table 5. Results for negative binomial models on men's medals won

	Overall Gender Equality			Educational Gender Equality			Economic Gender Equality			Health Gender Equality			Political Gender Equality			All Gender Equality Subindexes		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Intercept	.53	.14	<.001	.44	.15	.005	.50	.14	<.001	.53	.14	<.001	.53	.14	<.001	.39	.15	.009
Latitude	.11	.14	.422	-.01	.13	.940	.09	.13	.485	.07	.13	.571	.01	.14	.921	-.06	.14	.656
GDP per capita	1.12	.18	<.001	.94	.20	<.001	1.16	.15	<.001	1.29	.16	<.001	1.34	.17	<.001	1.02	.21	<.001
Population	.81	.13	<.001	.84	.13	<.001	.88	.14	<.001	.75	.13	<.001	.78	.13	<.001	.96	.14	<.001
Gini Index	-.28	.16	.077	-.40	.17	.017	-.28	.16	.075	-.27	.16	.091	-.30	.16	.066	-.43	.17	.011
Overall Gender Equality	.27	.18	.128															
Educational Gender Equality				.72	.33	.031										.58	.33	.073
Economic Gender Equality							.41	.15	.005							.47	.16	.003
Health Gender Equality										-.10	.14	.482				-.08	.13	.524
Political Gender Equality													-.17	.15	.267	-.30	.15	.041

Notes: GDP per capita and population (in thousands) are log transformed.

Arbitrarily-Defined World Regions Should Not Be Used to Analyze Country-Level Data

An interesting question is precisely which variables should be controlled for when probing the relationship between gender equality and medal wins. Population size? Wealth? Distance from the equator? Any control variable requires a strong theoretical foundation motivating its inclusion in the analyses.

K&P (2014, 2015) argue that research on cross-national comparisons should include controls for world regions. They use the analogy of research on education and families, in which students are nested within classrooms and individuals are nested within family. Their analogy of regions with classrooms and biological relatives breaks down rapidly. School classes and families are objectively defined groups, like nations; world regions are not.

The world regions K&P (2015) control for include widely geographically dispersed “Anglo-Saxon countries” and arbitrary cut-offs within continents, such as “Northern and Western Europe” versus “Central and Eastern Europe,” and “Central and East Asia” versus “South and Southeast Asia.” They combine 24 countries into one “Africa” region and group Iran, Israel, Morocco, and Turkey as part of “Middle East and North Africa.” In their supplement, they report analyses with alternative regional groupings, such as “Asia” (e.g., Iran, China, Japan, India, and Kazakhstan) and “Insular Pacific” (e.g., Philippines, Indonesia, Australia, and Fiji).

The national groupings proposed by K&P (2015) are arbitrary and subjective and should not be used when analyzing country-level data. In some cases, there are arguably greater national similarities *across* than *within* these regions. These groupings do not follow a defensible logic based on objective and consistent empirical criteria. Rather, they appear to represent subjective judgments based on folk stereotypes. Highlighting this arbitrariness, K&P employ inconsistent sets of regional groupings in their different critiques (2014, 2015).

For this reason, as well as empirical ones (see S3), psychological research on cross-national comparisons does not make it a practice to use regional dummies of any kind. A far superior strategy is to assess how countries with objectively defined legal borders differ on a dimension of interest and to examine whether those national differences predict relevant outcomes. This is precisely what was done in Berdahl et al. (2015) and in other research on cross-national comparisons (e.g., Bernard & Busse, 2004; Gelfand et al., 2011; Guiso et al., 2008).

Crowdsourcing Data Analysis

As seen above with the Win-Win effect, there are often multiple ways to analyze the same dataset—some more defensible than others—that can lead to variability in results. Two of the present authors have, together with a large group of colleagues, launched a program of research on crowdsourcing data analysis that both highlights and seeks to address this issue (Robinson et al., 2015; Silberzahn et al., 2015). In these crowdsourced projects, different teams of researchers are provided with the same dataset to test the same hypothesis. The teams at first work independently, and then are informed of each other's methods (but not results) so that they can provide detailed feedback and engage in methodological debates.

In our first project, 29 teams used the same dataset to test the hypothesis that soccer referees are more likely to give red cards to dark skin toned players than to light skin toned players (Silberzahn et al., 2015). Analytic approaches varied widely across teams, as did estimated effect sizes, which ranged from 0.89 to 2.93 in odds ratio units. The range of estimated effect sizes among analysts with a very high level of statistical expertise was similar to that for comparatively less expert analysts. Despite this striking variability, there was also some convergence in findings and conclusions: Over two-thirds of the teams (69%) reported a statistically significant effect in the expected direction, and team leaders indicated overall agreement that the data supported the hypothesis that dark skin toned players received more red cards.

Crowdsourcing data analysis highlights both convergence in conclusions and the contingency of results on choices of analytic strategy. Such analysis-contingent results are likely a basic property of most complex datasets and empirical findings, the present dataset on Olympic medal wins not excluded. One implication is that even true findings will often not be statistically significant at the $p < .05$ level with every possible operationalization of key variables, combination of covariates, and choice of statistical tests. In the case of the Win-Win effect, the relationship between gender equality and Olympic medal wins finds empirical support with a variety of controls and analytic strategies, but of course not all of them.

That there are many defensible ways to analyze the same dataset also holds important implications for both the analysis and re-analysis of data from scientific papers. When data analysis is crowdsourced, numerous analysts test the same hypothesis in distinct ways, with limited individual incentive to produce anything other than the most error-free and accurate results. But in a standard scientific investigation, there is a strong individual incentive to produce a publishable result. It is possible for a single analyst to test the same hypothesis numerous ways with the same dataset and choose the one specification that produces statistically significant support for his or her theory (i.e., to engage in *p-hacking*; Simmons, Nelson, & Simonsohn, 2011).

The issue of perverse incentives and analytic flexibility on the part of original authors has been the topic of considerable discussion in the field of late. Less discussed is the fact that parallel problems emerge with respect to the re-analysis of data. To publish a reply to an original paper, the re-analyst has a strong incentive to attempt many specifications and chose an approach that produces the *least* possible support for the original paper's hypothesis (i.e., to engage in *reverse p-hacking*; Schnall, 2014). This is highly problematic. Analysis-contingent results make it almost inevitable that any true relationship will first become non-significantly positive and then eventually reach zero or even reverse as more and more controls are added or different analytic approaches are employed (Anderson & Anderson, 1996; Silberzahn et al., 2015). One of the many advantages of crowdsourcing data analysis prior to publication is to avoid non-scientific incentives in both the original analysis and re-analysis of data.

Conclusion

Countries that are more gender equal in important ways enjoy greater athletic success at the Olympic games, an effect that holds not only for female but also for male athletes. This Win-Win effect finds empirical support with a variety of valid statistical controls and analytic approaches, including controls for GDP per capita. It is most robustly supported with regard to gender equality in education and earnings. Further controlling for arbitrarily-defined world regions is neither appropriate in this case nor for research on cross-national comparisons more generally. Such groupings are based on naïve stereotypes rather than scientific criteria, and controls for region risk obscuring the true effects of the characteristics of the world's many diverse nations.

To address the challenge of scientific results that are contingent, in part, on the analytic choices of researchers, we as a community can crowdsource the analysis of complex datasets prior to publication.

Acknowledgments

The authors thank Adam Galinsky, Toon Kuppens, Jim Sherman, and an anonymous reviewer for their helpful feedback on drafts of this paper, Markus Brauer for analytic insights and suggestions, and Dennis Ma for triple-checking our statistical calculations. Our dataset on gender equality and Olympic medal wins was publicly posted on the Open Science Framework before the publication of the original Berdahl et al. (2015) Win-Win paper and can be downloaded here: https://osf.io/d4m95/files/

References

Anderson, C. A., & Anderson, K. B. (1996). Violent crime rate studies in philosophical context: A destructive testing approach to heat and southern culture of violence effects. *Journal of Personality and Social Psychology, 70*, 740–756.

Barsh, J., & Yee, L. (2011, April). Unlocking the full potential of women in the U.S. Economy. *McKinsey Report*. Retrieved from http://www.mckinsey.com/client_service/organization/latest_thinking/unlocking_the_full_potential

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

Bernard, A.B., & Busse, M.R. (2004). Who wins the Olympic games: Economic resources and medal totals. *The Review of Economics and Statistics, 86(1),* 413–417.

Chaaban, J., & Cunningham, W. (2011). Measuring the economic gain of investing in girls: The girl effect dividend. *Policy Research Working Paper*, 5753.

Gelfand, M.J., Raver, J.L., Nishii, L.M., Leslie, L., Lun, J., Lim, B.C., et al. (2011). Differences between tight and loose cultures: A 33-Nation Study. *Science, 332*, 1100-1104.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science, 320*, 1164-1165.

Inglehart, R., & Norris, P. (2003). *Sacred and secular: Religion and politics worldwide.* New York, 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*.

Lagerlöf, N. P. (2003). Gender equality and long-run growth. *Journal of Economic Growth, 8,* 403-426.

Löfström, Å. (2009). Gender equality, economic growth and employment. Report financed by the Swedish Ministry of Integration and Gender Equality. Retrieved from http://www.forschungsnetzwerk.at/downloadpub/2009_12_Gender_Equality_study.pdf

Morrison, A., Raju, D., & Sinha, N. (2007). Gender equality, poverty and economic growth. The World Bank Policy Research Working Paper. Retrieved from https://books.google.ca/books?hl=en&lr=&id=2ch8Sh63xggC&oi=fnd&pg=PA1&dq=gender+equality+gdp&ots=8A61wTx0Qy&sig=0Q3AU_m86I3g7M6PGIwnpbdICRE#v=onepage&q=gender%20equality%20gdp&f=false

Robinson, E.S., Snellman, K., Silberzahn, R., Sommer, A., Otner, S., Robinson, D., Adie, P., Alstott, J., Armstrong, W.S., Bahnik, S., Balling, L.W., Banker, S., Banks, G.C., Barr, D. Bauer, M., Berry, S., Boelen, L., Cambo, S., Canela, M., Chan, C.S.R. Clark, M., Clayden, J., Cohn, B., Colclough, G., Coppock, A., Cugueró-Escofet, N.,deRuijter, S., Duarte de Souza, C., Duggal, S., Dube, B. T., Dunn, B., Bednall, T., Gong, Y., Gordon-McKeon, S., Gnambs, T., Greenawalt, B., Guess, A., Hanowell, B., Hemphill, L., Henninger, F., Huntsinger, J., Jiang, L., Jong, J., Kappler, G., Kane, J. Kaszubowski, E., Kavanagh, C., Kelchtermans, S., Kent, H., Khabsa, M., Kim, J., Klamer, B., Kouros, J., Lindsay, T., Liverani, S., Low, D., Madan, C., Magee, J., Majumdar, B., Manley, H., Martin, C., Martinez, A., Matskewich, E., Mertens, U., Miller, D., Mohr, A. H., Moore, B., Nelson, A., O'Brien, C. E., O'Neill, P., O'Shea, K., Olita, H., Ong, C.W., Otterbacher, J., Palsetia, D., Ray, C., Reyt, J., Rubineau, B., Russo, R., Scheepers, C., Shotwell, G., Simecek, P., Simecokova, M., Sotak, K., Spain, S., Suchow, J., Taylor, L., Tomkins, S.Tran, M., Vranka, M., Vul, E., Wadsworth, D., Wee, S., Whelpley, C., Won, A., Wu, L., Yang, W., Yarkoni, T., & Uhlmann, E. L. (2015). *Crowdsourcing data analysis: Gender, status, and science.* Data analysis phase in progress.

Schnall, S. (2014, November 18). Further thoughts on replications, ceiling effects and bullying. Retrieved from http://www.psychol.cam.ac.uk/cece/blog

Silberzahn, R., Uhlmann, E. L. Martin, D., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., Fong, N., Gamez-Djokic, M., Glenz, A., Gordon-McKeon, S., Heaton, T. J., Hederos Eriksson, K., Heene, M., Hofelich Mohr, A. J., Högden, F., Hui, K., Johannesson, M., Kalodimos, J., Kaszubowski, E., Kennedy, D., Lei, R., Lindsay, T. A., Liverani, S., Madan, C. R., Molden, D., Molleman, E., Morey, R. D., Mulder, L. B., Nijstad, B. A., Pope, N. G., Pope, B., Prenoveau, J. M., Rink, F., Robusto, E., Roderique, H., Sandberg, A., Schlüter, E., Schönbrodt, F. D., Sherman, M. F., Sommer, S., Sotak, K., Spain, S., Spörlein, C., Stafford, T., Stefanutti, L., Tauber, S., Ullrich, J., Vianello, M., Wagenmakers, E., Witkowiak, M., Yoon, S., & Nosek, B.A. (2015). *Many analysts, one dataset: Making transparent how variations in analytical choices affect results.* Manuscript under review.

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: *Undisclosed* flexibility in data collection and analysis allow presenting anything as significant. *Psychological Science, 22(11),* 1359-1366.

World Economic Forum (2014). The global gender gap index: The case for gender equality. Retrieved from http://reports.weforum.org/global-gender-gap-report-2014/part-1/the-case-for-gender-equality/

Online Supplements for Bai, Uhlmann, & Berdahl,

"The Robustness of the Win-Win Effect"

Supplement 1: National wealth is a conservative control when predicting Olympic medal wins from gender equality (but should still be controlled for)

It is worth elaborating on why controlling for national wealth represents a conservative test of the "Win-Win" hypothesis when estimating the relationship between gender equality and medal wins. Let us make the plausible assumption that the causal relationship between gender equality and economic growth is bidirectional. If the effects are reciprocal, then there likely exists (1) a direct effect of gender equality on medal wins, (2) an indirect effect of gender equality on medal wins mediated by national wealth, and (3) a spurious component of the correlation between gender equality and medal wins that is really due to the third variable of national wealth. If so, then controlling for national wealth removes not only (3) but also (2), underestimating the true relationship between gender equality and medal wins, which consists of both direct (unmediated) and indirect (mediated) effects. Figure S1 below illustrates this visually.

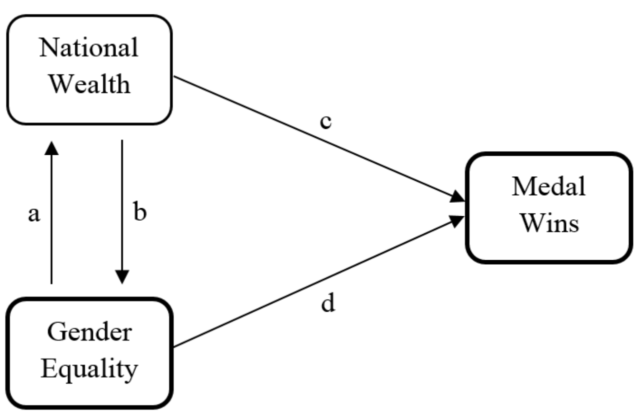

Figure S1. Theoretical path model of the relationships between gender equality, national wealth, and Olympic medal wins.

Let us assume a path model with two variables that cause each other (national wealth and gender equality, paths a and b) and a third (outcome) variable that is caused by these two variables (medals, paths c and d). The total effect of gender equality on medals is (1) the direct effect (path d) and (2) the indirect effect through national wealth (paths a*c). By statistically controlling for national wealth, we remove the influence of national wealth and thus examine only the direct effect (i.e., the part of the total effect that is not genuinely mediated by national wealth *and* that is not a spurious relationship due to the third variable of national wealth). As a consequence, the regression coefficients controlling for national wealth in Tables 2-5 of the main text underestimate the total effect of gender equality on medal wins.

At the same time, it is also the case that estimating the relationship between gender equality and medal wins without controlling for economic wealth would represent an overly liberal test of the hypothesis. This would leave not only (1) the direct effect of gender equality on medal wins and the (2) indirect effect of gender equality on medal wins that is mediated by economic wealth, but also (3) the spurious component of the correlation between gender equality and Olympic medal wins that is due to the third variable of national wealth.

Supplement 2: Effects of different analytic approaches on support for the Win-Win hypothesis

GDP per capita and national population are both positively skewed variables. Kuppens and Pollet (2015) inconsistently log transform GDP per capita to correct for this skew and use raw scores for national population. This is a simple and understandable mistake to make, but turns out to have major implications for the degree of support obtained for the Win-Win hypothesis.

The potential analytic approaches can be more fully represented as a 2 (log transform vs. use raw scores) x 2 (GDP per capita vs. national population) matrix. How do these different specifications affect the results?

There exist six significant zero-order correlations between measures of gender equality and Olympic medal wins (Table 1): specifically, between overall gender equality, educational gender equality, and economic gender equality and medal wins for both male and female athletes. Table S2 below summarizes whether these relationships remain statistically significant in regressions using each analytic approach.

			GDP per capita															
			Log transformed								Raw							
			Women's medals				Men's medals				Women's medals				Men's medals			
			Quasipoisson		Negative binominal		Quasipoisson		Negative binominal		Quasipoisson		Negative binominal		Quasipoisson		Negative binominal	
			b	*p*	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
Population	Log transformed	Overall gender equality	.33	.013	.83	<.001	.05	.643	.27	.128	.40	.009	1.09	<.001	.23	.097	.54	.006
		Educational gender equality	1.42	.004	.75	.031	1.01	.010	.72	.031	2.09	<.001	1.24	<.001	1.83	<.001	1.42	<.001
		Economic gender equality	.56	<.001	.97	<.001	.20	.082	.41	.005	.64	<.001	.97	<.001	.34	.013	.41	.020
	Raw	Overall gender equality	**.06**	**.702**	*.71*	*.010*	**-.15**	**.259**	*.11*	*.556*	.29	.097	.87	.002	.15	.327	.29	.171
		Educational gender equality	**.72**	**.194**	*.27*	*.427*	**.52**	**.215**	*.32*	*.308*	1.69	.005	.78	.011	1.48	.001	1.06	<.001
		Economic gender equality	.43	.025	.76	.001	-.001	.993	.20	.180	.60	.004	.67	.007	.23	.167	.12	.485

Table S2. Regression results under each possible specification. Results based on inconsistently log transformed variables are presented with a grey background. The analyses reported in Kuppens and Pollet's (2015) main text are in **bold**. Analyses reported in their supplement are presented in *italics*. Analyses that were not part of their commentary are in regular font.

It is illustrative to examine the number of statistically significant effects in each quadrant of Table S2. As seen in the top left quadrant, when both GDP per capita and national population are log transformed (as in our analyses from the main text), 9 of 12 tests of the relationship between gender equality and medal wins are positive and statistically significant, one is positive and marginally significant, and two are positive but nonsignificant.

The bottom right quadrant of Table S2 displays the results if raw scores are used for both GDP per capita and national population despite their positively skewed distributions. In this specification, seven statistical tests for the relationship between gender equality and medal wins are positive and significant, one is positive and marginally significant, and four are positive but nonsignificant.

The top right quadrant of Table S2 displays the results if raw scores are used for GDP per capita and national population is log transformed. In this specification, 11 statistical tests for the relationship between gender equality and medal wins are significantly positive, one is positive and marginally significant, and none are nonsignificant.

The specification that produces the least support for the Win-Win hypothesis is log transforming GDP per capita but not national population, as Kuppens and Pollet (2015) did. As shown in the bottom left quadrant of Table S2, this analytic approach results in nine nonsignificant relationships between gender equality and medal wins: eight positive and one negative. The other three relationships are significant and positive. One of these, a significant positive relationship between overall gender equality and medal wins for female athletes in the negative binomial regression, is reported in their supplement.

Two of these three significant positive effects are not included in Kuppens and Pollet's (2015) commentary (either the main text or supplement). The reason is that Kuppens and Pollet analyzed overall World Economic Forum gender gap scores and only one of the four World Economic Forum gender gap subindexes, specifically educational equality between men and women. This was done because educational gender equality was the best predictor out of the four subindexes in Berdahl et al. (2015). Unfortunately, however, this approach led Kuppens and Pollet to overlook significant effects of *economic* gender equality.

In sum, log transforming GDP per capita and using raw scores for population yields less support for the Win-Win effect than the other potential analytic approaches (see Table S2). Kuppens and Pollet also overlook positive relationships between measures of gender equality and medal wins that emerge when the available measures are more fully analyzed.

Supplement 3: Further empirical issues with the use of regions in cross-national analyses

Researchers should strive to avoid not only Type 1 but also Type 2 errors. There are a limited number of nations with reliable data for many variables of interest, and further including regions in the analysis risks increasing the false negative rate too high. The loss of degrees of freedom when regional controls are included is a comparatively minor issue when the sample includes 121 countries, as in the Olympic medals and gender equality dataset. However, it is a significant problem for studies that only include twenty or thirty countries (e.g., Gelfand et al., 2011; Glick et al., 2000). Cross-national investigations based on new original data collected by the primary investigators and colleagues at other universities (e.g., Glick et al., 2000) are potentially crippled by reduced power.

At the same time, meaningful variability in both predictors and outcomes is reduced, in that prediction must occur within regions (Brauer, 2015). For instance, for national collectivism to predict an outcome of interest, it must do so within each area of the world (e.g., "Central and Eastern Europe," "Southern Europe"), rather than across all countries of the world. The field of cross-national comparisons becomes the study of variability within regions. Typically, however, the hypothesis is that *across the world*, nations with high scores on variable A (e.g., gender equality) further exhibit characteristic B (e.g., Olympic medal wins).

These empirical issues add to the more fundamental problem that regional groupings of nations are inherently subjective and arbitrary, regardless of whether the regional distinctions are made by Kuppens and Pollet (2014, 2015) or by other investigators.

References

Brauer, M. (2015). *Different ways to deal with non-independent data in scientific research: The example of world regions in cross-cultural research.* Manuscript under review.

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

Gelfand, M.J., Raver, J.L., Nishii, L.M., Leslie, L., Lun, J., Lim, B.C., et al. (2011). Differences between tight and loose cultures: A 33-Nation Study. *Science, 332*, 1100-1104.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J, Abrams, D., Masser, B., Adetoun, B., Osagie, J., Akande, A., Alao, A., Brunner, A., Willemsen, T. M., Chipeta, K., Dardenne, B., Dijksterhuis, A., Wigboldus, D., Eckes, T., Six-Materna, I., Expósito, F., Moya, M., Foddy, M., Kim, H-J., Lameiras, M., Sotelo, M. J., Mucchi-Faina, A., Romani, M., Sakalli, N., Udegbe, B., Yamamoto, M., Ui, M., Ferreira, M. C., & López, W. L. (2000). Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology, 79*, 763-775.

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.*