OPEN ACCESS

University of BRISTOL

Peer reviewed version

Link to published version (if available):
10.1037/ebs0000067

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Cross-race Misaggregation: Its Detection, a Mathematical Decomposition, and Simpson's

Paradox

Bryan L. Koenig[a], Florian van Leeuwen[b], Justin H. Park[c]

[a]University College, Washington University in St. Louis, St. Louis, MO, 63130, USA.

bryanleekoenig@gmail.com

[b]Dynamique du Langage, University of Lyon, Lyon, France. f.van.leeuwen.84@gmail.com

[c]School of Experimental Psychology, University of Bristol, Bristol, UK. j.h.park@bristol.ac.uk

Author Note

Correspondence concerning this article should be addressed to Bryan L. Koenig,

University College, Washington University in St. Louis, One Brookings Dr, Campus Box 1085,

St. Louis, MO, 63130. 1-406-531-0470.

Email: bryanleekoenig@gmail.com

Abstract

Researchers sometimes aggregate data, such as combining resident data into state-level means. Doing so can sometimes cause valid individual-level data to be invalid at the group level. We focus on cross-race misaggregation, which can occur when individual-level data are confounded with race. We discuss such misaggregation in the context of Simpson's Paradox and identify four diagnostic indicators: aggregated rates that correlate strongly with the relative size of one or more subgroup(s), unequal sample sizes across subgroups, unequal rates or mean values across subgroups, and aggregated rates that do not correlate with subgroup rates. To illustrate these diagnostic indicators, we decomposed data on the prevalence of sexually transmitted diseases (STDs) to confirm cross-race misaggregation in *Parasite Stress USA*, an ostensible index of parasite prevalence known to be confounded with the proportion of African American residents per state.

Keywords: Simpson's Paradox, ecological fallacy, parasite-stress theory, sexually transmitted diseases, population demographics

Cross-race Misaggregation: Its Detection, a Mathematical Decomposition, and Simpson's

Paradox

Researchers sometimes test hypotheses by analyzing data aggregated at the level of US states or countries. Such analyses face several obstacles to validity: Data points are non-independent, measures might mean different things in different countries, and individual-level or subgroup relationships cannot be reliably inferred from group-level aggregated data (Pollet, Tybur, Frankenhuis, & Rickard, 2014). We elaborate on one cause of the latter obstacle: a validity threat we refer to as *misaggregation*. By this we mean valid data combined together such that the aggregate represents neither what the researcher intended it to represent, nor what it represented at the individual level. Instead, the aggregate is undermined by a confounder variable such that a true effect at the individual or subgroup level is obscured at the aggregate level because of a confounder. Confounder variables can be hard to identify, making misaggregation easy to overlook. We focus on race, a relatively easy-to-identify confounder variable, although many variables are potential confounders across states or countries, such as poverty levels, ethnic groups, and rural-vs-urban residency rates. To help researchers who use aggregated data avoid using invalid aggregated variables, we connect misaggregation to Simpson's Paradox, identify four red flags that can help researchers detect misaggregation, and show how cross-race misaggregation can occur. Our primary illustration uses Parasite Stress USA, a variable that was intended to be an index of pathogen prevalence for the fifty states of the USA (Fincher & Thornhill, 2012) that has been found to be invalid due to confounding with race (Hackman & Hruschka, 2013; Hruschka & Hackman, 2014).

Simpson's Paradox and Cross-race Misaggregation

A key problem with aggregated data is when relationships observed at the group level are not the same as the relationships that occur at the subgroup or individual level. This is Simpson's Paradox[1] (Simpson, 1951). It can be seen in the following example, in which an effect of treatment is present in subgroups of men and women but disappears when their data are aggregated into a single group. Survivorship can be higher for treated men (61%, or 8/13) compared to untreated men (57%, or 4/7) as well as for treated women (44%, or 12/27) compared to untreated women (40%, or 2/5). When aggregated across the sexes, however, the efficacy of the treatment disappears: treated people (50%, or 20/40) survive no better than untreated people (50%, or 6/12; Simpson, 1951). The possibility of divergence in relationships across levels implies that researchers should not infer that relationships observed at the group level hold for subgroups or individuals (i.e., the ecological fallacy; Robinson, 1950). Simpson's Paradox can occur when the relationship differs across subgroups due to a third, confounder variable.

Known methods for detecting Simpson's Paradox are pertinent to detecting misaggregation. Kievit, Frankenhuis, Waldorp, and Borsboom (2013) provided four methods to detect Simpson's Paradox. First, if the data are bivariate and continuous, one can look at a scatterplot for any obvious subgroups with different patterns of results. Second, for contingency tables with an observed relationship at the aggregate level, one can compute a chi-square test of

---

[1] Tu, Gunnell, & Gilthorpe (2008) argue that many labels, such as Simpson's Paradox (Simpson, 1951), Lord's Paradox (Lord, 1967), or suppression, refer to the same underlying phenomenon, the *reversal paradox*. We use the best known term, Simpson's Paradox, to refer to reversals regardless of other features of the situation.

independence to see if the frequency distributions differ across subgroups. If they differ, then the subgroups should be analyzed separately. Third, researchers using regression can check the residuals for systematic (subgroup-based) differences in homoscedasticity, which could reflect the different slopes of the subgroups. The fourth diagnosis technique is to use latent cluster analysis to detect subgroups whose patterns of results diverge. Such clusters are based on their position on a bivariate scatterplot, although the technique can also be applied to multiple regression. (Kievit and colleagues made available an analysis tool that detects diverging clusters and statistically evaluates whether the observed relationship of interest differs statistically across those clusters.)

However, for Parasite Stress USA, a measure of infectious-disease prevalence aggregated at the level of US states—which is known to be confounded with race—the diagnostic methods of Kievit and colleagues (2013) do not reliably indicate Simpson's Paradox (see Supplemental Material S1 available online). This is because the checks assume that each case has only one level on the confounder variable. For example, half of the participants are male and the other half female, and a positive correlation is observed when analyzing all participants together, but a negative correlation is observed within each sex. In such cases the confounder variable might be considered a *between-cases confound* because each data point is associated with only one level of the confounder variable. Parasite Stress USA (and likely other variables suffering from cross-race misaggregation) differ systematically in that the confounder can be thought of as a *within-cases confound*: each data point has within it all levels of the confounder variable. For such data, the methods suggested by Kievit and colleagues (2013) would be unlikely to reveal misaggregation.

Red Flags for Cross-race Misaggregation

Given the frequency of research that uses data aggregated across demographic characteristics such as race, researchers could benefit from evaluating their data for the following four red flags, which are suggestive of misaggregation. The first red flag is when aggregated rates correlate strongly with the relative size of one or more subgroup. The second is when sample sizes are unequal across subgroups. The third is when rates or mean values differ across subgroups. The fourth is when aggregated rates correlate weakly or not at all with subgroup rates. Checks for the red flags require access to information about subgroups, which might be difficult to obtain. In many cases, relevant information is available for demographic subgroups, such as the population sizes of different racial groups. Thus the check for the first red flag is relatively easy to do for cross-race misaggregation. The first three checks are suggestive; the fourth is diagnostic (i.e., if total rates correlate with all subgroups, misaggregation has not occurred across those subgroups). We illustrate these red flags with Parasite Stress USA.

Parasite Stress USA

We build upon research that has identified problems with an aggregated index of pathogens—Parasite Stress USA—that has been used in tests of parasite stress theory. Parasite stress theory proposes that norms and practices that reduce the likelihood of pathogenic infection will be heightened among cultural groups situated in regions with more instances of infectious disease (Thornhill & Fincher, 2011). Evidence presented in support of parasite stress theory includes correlations of sociality variables with Parasite Stress USA or similar pathogen indexes (Fincher & Thornhill, 2012; Shrira, Wisman, & Webster, 2013; Thornhill & Fincher, 2011; Varnum, 2012, 2014). However, these findings may be invalid as Parasite Stress USA is confounded with the percentage of state populations that was African American, *%Black*

(Hruschka & Hackman, 2014). The confounding of Parasite Stress USA with %Black is argued

to have resulted from African Americans having higher STD rates than non-Hispanic Whites and

%Black varying substantially across states. Hruschka and Hackman (2014) provide suggestions

for researchers who desire to use aggregated data but avoid the pitfalls, such as replicating with

new data and at multiple levels, considering historical and social context, and testing alternative

hypotheses.

To further demonstrate and elucidate the confounding of Parasite Stress USA with

%Black, we removed components of the data from the numerator and denominator of the

aggregate to evaluate the contributions of the components. This decomposition confirms that

unequal STD rates across racial subgroups played a key role. It also highlights the importance of

unequal sample sizes (i.e., the proportion of state populations composed of non-Hispanic Whites

compared with African Americans). Our decomposition did not address the contribution of

variation in %Black; however, it was important because if it were constant across states, then the

numerator would have been driven by the relatively high STD rates of African Americans and

the aggregate would have been strongly correlated with STD rates of African Americans rather

than with %Black. In short, cross-race misaggregation can result when a minority subgroup has

disproportionate influence on the numerator but not the denominator of an aggregate index.

Method

We obtained Parasite Stress USA from supplementary materials of Fincher and Thornhill

(2012). They derived Parasite Stress USA from the *Summary of Notifiable Diseases, United

States* for years 1993 to 2007, part of the annual Morbidity and Mortality Weekly Report of the

Centers for Disease Control and Prevention (CDC). It is a standardized measure of the total

incidence of all notifiable diseases reported by all states for a year divided by the population, calculated separately for each state.

Hackman and Hruschka (2013) demonstrated that Parasite Stress USA mainly represents STDs, because STD cases dwarf cases of other notifiable diseases. Using data from the CDC WONDER spanning 1998-2009, Hackman and Hruschka developed STD indexes using the two most common STDs, chlamydia and gonorrhea (CG), for the total population and separately for non-Hispanic Whites and African Americans. For the total population (i.e., collapsed over racial subgroups), the CG index was strongly correlated with Parasite Stress USA; $r = .95$, $N = 50$, $p < .001$. They reasoned that if parasite stress theory is valid, the pattern observed when collapsed over racial subgroups should also occur when analyzing the White and African American subsamples separately. We also used CDC WONDER data to recreate CG rates for our analyses. We developed three indexes of CG rates per 100,000 residents: one for the whole population, *CG rates$_{total}$*; another for non-Hispanic Whites, *CG rates$_{white}$*; and one for non-Hispanic African Americans, *CG rates$_{black}$*. For consistency we also used this data source for calculating values %Black (note this source for %Black differs from that used by Hackman & Hruschka, 2013, and that their African American population included Hispanics).

Results

Parasite Stress USA Failed All Red Flag Checks

Reproducing Hackman and Hruschka (2013), CG rates$_{total}$ strongly correlated with Parasite Stress USA, $r = .96$, $N = 50$, $p < .001$. Previous reports have noted the high correlation between Parasite Stress USA and %Black (Eppig, Fincher, & Thornhill, 2011; Hackman & Hruschka, 2013). Our data showed the same pattern: Parasite Stress USA correlated strongly

with %Black, $r = .90$, $N = 50$, $p < .001$. Thus, Parasite Stress USA failed the check for the first red flag.

Parasite Stress USA failed the checks for the second and third red flags because across states non-Hispanic Whites and African Americans diverged substantially in sample size and rates of STDs, respectively. African Americans, on average, comprised only 7.50% of state populations whereas non-Hispanic Whites comprised 60.68% of state populations. In addition, across US states, the African American CG rates per 100,000 ($Mdn = 1,810.79$, $M = 1,954.96$, $SD = 747.72$) were an order of magnitude higher than those of non-Hispanic Whites ($Mdn = 157.78$, $M = 162.95$, $SD = 54.83$), $d = 2.45$, Wilcoxon signed-rank test, $Z = 6.15$, $p < .001$.

If Parasite Stress USA were simply a race-independent index of STD rates, it might correlate with race-stratified CG rates. The strongest correlation might be observed with the CG rates of the largest racial subgroup, non-Hispanic Whites. However, Parasite Stress USA did not correlate with CG rates$_{white}$, $r = -.03$, $N = 50$, $p = .858$ (nor was CG rates$_{total}$ correlated with CG rates$_{white}$, $r = .13$, $N = 50$, $p = .354$). Alternatively, given that African Americans accounted for a large number of CG cases in absolute numbers, Parasite Stress USA might correlate with CG rates$_{black}$. This also did not occur, $r = .11$, $N = 50$, $p = .464$ (nor was CG rates$_{total}$ correlated with CG rates$_{black}$, $r = .23$, $N = 50$, $p = .112$. Parasite Stress USA therefore failed the check for the fourth red flag.

Decomposition of Parasite Stress USA: Why Aggregate Rates Correlate With %Black

Parasite Stress USA is uncorrelated with race-stratified CG rates, but is strongly correlated with %Black. This is because the numerator of Parasite Stress USA was mostly STDs, and a large proportion of CG cases were in the African American subcomponent of state populations. This results in African American CG cases strongly influencing the numerator of

CG rates$_{total}$ across US states (and by implication Parasite Stress USA), but not its denominator, which is largely determined by members of other races.

If African American CG cases exert such a strong influence on Parasite Stress USA, then removing them from the numerator of CG rates$_{total}$ should reduce its correlation with Parasite Stress USA. On the other hand, including only African American CG cases in CG rates$_{total}$ should not reduce the correlation much. These predictions were confirmed (see Table 1). Removing African American CG cases eliminated the significant correlations of CG rates$_{total}$ with both Parasite Stress USA and %Black. Conversely, using *only* African American CG cases hardly changed these correlations. For comparison, we excluded non-Hispanic White CG cases (see Table 1). Doing so had little effect on the relationships of CG rates$_{total}$ with either Parasite Stress USA or %Black. By contrast, including only non-Hispanic White CG cases dramatically altered the relationships. These results indicate that African American CG cases were critical for the strong relationships of CG rates$_{total}$ and Parasite Stress USA with %Black.

To see whether African American or non-Hispanic White population sizes had a strong influence on the denominator, we excluded each in turn. Removing non-Hispanic Whites from state populations substantially altered correlations of CG rates$_{total}$ with Parasite Stress USA and %Black, but removing African Americans from the population size did not (Supplemental Material S2 available online). Therefore, African Americans had little influence over the denominator of CG rates$_{total}$, and by implication Parasite Stress USA. In sum, African American CG cases can account for the observed strong relationships of Parasite Stress USA and CG rates$_{total}$ with %Black.

Table 1. Decomposition of Parasite Stress USA: Correlations among Parasite Stress USA, population percentage African American (%Black), and CG rates$_{total}$. Modified CG rates$_{total}$ were produced by excluding African American or non-Hispanic White CG cases from the total number of cases, or including only African American or non-Hispanic White CG cases. The formulas for various calculations of CG rates are shown.

	CG rates$_{total}$	No CG$_{black}$	No CG$_{white}$	Only CG$_{black}$	Only CG$_{white}$
	CG$_{total}$ / pop$_{total}$	(CG$_{total}$ − CG$_{black}$) / pop$_{total}$	(CG$_{total}$ − CG$_{white}$) / pop$_{total}$	CG$_{black}$ / pop$_{total}$	CG$_{white}$ / pop$_{total}$
Parasite Stress USA	.96***	.26$^+$	.97***	.93***	-.45***
%Black	.80***	-.10	.84***	.89***	-.34*
CG rates$_{total}$	-	.42**	.97***	.84***	-.20

$^+ p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Other Examples of Checks for Cross-Race Misaggregation

Misaggregation may have invalidated other aggregated variables used in research on ecological (environmental) effects. For illustration purposes, we identified three for which African Americans and non-Hispanic Whites were likely to differ: incarceration rates (Kruger & De Loney, 2009), life expectancy (Thornhill & Fincher, 2011; Eppig et al., 2011), and homicide rates (Hackman & Hruschka, 2013; Shrira et al., 2013; Thornhill & Fincher, 2011). Table 2 illustrates with these three variables how the identified red flags indicate the presence or absence of cross-race misaggregation. For incarceration rates, the fourth red flag for total, state-level rates

strongly suggests they may be confounded with race. Removing African American cases resulted

in the correlation between total incarceration rates and %Black becoming negative ($r = -.35$, $p =$

.016), and including only African American cases increased it ($r = .96$, $p < .001$). Thus the

decomposition of state-level incarceration rates showed a pattern similar to that observed for CG

rates$_{total}$ and Parasite Stress USA (see Supplemental Materials S3). This suggests that total

incarceration rates indeed suffers from cross-race misaggregation. Despite three suggestive red

flags for both life expectancy and homicide rates, total state-level values strongly correlated with

both race-stratified rates; therefore, the fourth diagnostic checks indicates no misaggregation for

life expectancy or homicide rates. The absence of misaggregation might be due to positive

correlations for African Americans with non-Hispanic Whites (life expectancy, $r(36) = .65$, $p <$

.001; homicide rates $r(45) = .38$, $p = .009$).

Table 2. Illustrative red flag checks for three state-level aggregate variables: incarceration, life

expectancy, and homicide. These examples pertain to US states and so all exhibit the second red

flag of unequal racial subgroup sizes.

variable (total state-level rate)	r with %Black	first red flag?	mean African American value	mean non-Hispanic White value	third red flag?	r with African American rates	fourth red flag?	r with non-Hispanic White rates	fourth red flag?	cross-race mis-aggre-gation?
incar-ceration[1]	.65*	yes	2,572.63/ 100,000	425.23/ 100,000	yes[2]	.03	yes	.77*	no	yes
life expectancy[3]	-.57*	yes	74.46 years	78.62 years	yes[2]	.73*	no	.95*	no	no
homicide[4]	.70*[4]	yes	22.99/ 100,000	2.97/ 100,000	yes[2]	.62*	no	.58*	no	no

*$p < .001$
[1]data from Harrison and Beck, 2006; excludes New Mexico and Wyoming
[2]rates differ across races significantly, $p < .001$, using Wilcoxon signed-rank test
[3]life expectancy at birth; data from www.measureofamerica.org, 2010-2011 dataset; correlation sample sizes are $N = 45$
 for %Black, $N = 38$ for African American mean, and $N = 50$ for non-Hispanic White mean

[4]data from the Uniform Crime Report (Federal Bureau of Investigation 2003, 2005, 2006, 2007, & 2009), stratified by offender race; no data for Florida; following Hackman and Hruschka (2013), we excluded data from New Mexico and Nevada because they had large Hispanic populations but did no distinguish Hispanic from non-Hispanic Whites; other states with large Hispanic populations distinguished between Hispanic and non-Hispanic Whites; $N = 49$ for total and African American, $N = 47$ for non-Hispanic White.

Discussion

Researchers are sometimes interested in testing whether ecological variables motivate particular kinds of behavior, using this as evidence of context-specific adaptations. If an ecological variable is confounded with demographic variables, then we are mistaking cross-group differences, which might have any number of contextual or historical bases, with the ecological effect that we are specifically hypothesizing. Our analyses showed how data of infectious disease rates are confounded with race, and they illustrated four red flags suggestive of misaggregation. Although identifying confounding variables can be hard (and showing that no confounder is present may be impossible), researchers compiling and using an aggregated variable may check for these red flags. The red flags can be checked for demographic variables other than race (e.g., age, income). When the red flags indicate misaggregation, statistically controlling for the confounder variable (e.g., by including %Black as a predictor) might not solve the problem (e.g., due to multicollinearity). Stratified analyses may provide more valid inferences (Hruschka & Hackman, 2014). If lower-level data (e.g., stratified by race) are unavailable, these can sometimes be estimated from aggregated data (i.e., ecological inference) using methods developed in political science (King, 1997; Rosen, Jiang, King, & Tanner, 2001). Confounders can be hard to identify, so even in the absence of any of the identified red flags we advise researchers to be cautious when using aggregated data.

References

Eppig, C., Fincher, C. L., & Thornhill, R. (2011). Parasite prevalence and the distribution of intelligence among the states of the USA. *Intelligence, 39*, 155–160.

Federal Bureau of Investigation. (2003, 2005, 2006, 2007, 2009). *Uniform Crime Reporting Program.* Downloaded from https://www2.fbi.gov/ucr on 5 September 2014.

Fincher, C. L., & Thornhill, R. (2012). Parasite-stress promotes in-group assortative sociality: The cases of strong family ties and heightened religiosity. *Behavioral and Brain Sciences, 35,* 61–79.

Hackman, J., & Hruschka, D. (2013). Fast life histories, not pathogens, account for state-level variation in homicide, child maltreatment, and family ties in the US. *Evolution and Human Behavior, 34, 118–124.*

Harrison, P. M., & Beck, A. J. (2006). Prison and jail inmates at midyear 2005. *NCJ*, 213133.

Hruschka, D. J., & Hackman, J. (2014). When are cross-group differences a product of a human behavioral immune system? *Evolutionary Behavioral Sciences, 8,* 265-273.

Kievit, R. A., & Epskamp, S. (2012). *Simpsons: Detecting Simpson's Paradox. R package version 0.1.0.* Available online at: http://CRAN.R-project.org/package=Simpsons

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology, 4*, Article 513, 1-14.

King, G. (1997). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data.* Princeton: Princeton University Press.

Kruger, D. J., & De Loney, E. H. (2009). The association of incarceration with community health and racial health disparities. *Progress in Community Health Partnerships: Research, Education, and Action, 3.2,* 113–121.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68,* 304-305.

Pollet, T. V., Tybur, J. M., Frankenhuis, W. E. F., & Rickard, I. J. (2014). What can cross-cultural correlations teach us about human nature? *Human Nature*, *25*, 410–429.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Rosen, O., Jiang, W., King, G., & Tanner, M. A. (2001). Bayesian and frequentist inference for ecological inference: The R× C case. *Statistica Neerlandica, 55*, 134–156.

Shrira, I., Wisman, A., & Webster, G. D. (2013). Guns, germs, and stealing: Exploring the link between infectious disease and crime. *Evolutionary Psychology, 11,* 270–287.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological), 13*, 238–241.

Thornhill, R., & Fincher, C. L. (2011). Parasite stress promotes homicide and child maltreatment. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*, 3466–3477.

Tu, Y. K., Gunnell, D. & Gilthorpe, M. S. (2008). Simpson's Paradox, Lord's Paradox, and suppression effects are the same phenomenon – the reversal paradox. *Emerging Themes in Epidemiology, 5*:2, 1-9.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for HIV, STD and TB Prevention (NCHSTP), Division of STD/HIV Prevention, Sexually Transmitted Disease Morbidity for selected STDs by age, race/ethnicity and gender 1996-2009, CDC WONDER On-line Database, June 2011. Accessed at http://wonder.cdc.gov/std-v2009-race-age.html on 16 June 2013

Vandello, J. A. & Cohen, D. (1999). Patterns of individualism and collectivism across the United

States. *Journal of Personality and Social Psychology, 77*, 279–292.

Varnum, M. E. W. (2012). Frontiers, germs, and nonconformist voting. *Journal of Cross-

Cultural Psychology, 44,* 832–837.

Varnum, M. E. W. (2014). Sources of regional variation in social capital in the United States:

Frontiers and pathogens. *Evolutionary Behavioral Sciences, 8*, 77–85.

S1

   The methods by Kievit et al. (2013) need not always indicate that a bi-variate association is confounded. (Note that Kievit et al. did not claim that their methods would always reveal a Simpon's Paradox.) Below we show for the bi-variate association of Parasite Stress USA that the diagnostic checks suggested by Kievit et al. need not indicate that associations involving Parasite Stress USA (a variable known to be confounded with race) are confounded with a third variable.

   For the association of Parasite Stress USA with collectivism (Fincher & Thornhill, 2012; collectivism data from Vandello & Cohen, 1999), a scatterplot of both variables does not suggest different clusters (Figure S1.1). A Breusch–Pagan test did not reject the assumption of the residuals being homoscedastic, $BP = 0.20$, $df = 1$, $p = .652$. A cluster analysis with the R package "Simpsons" by Kievit and Epskamp (2012) detected two clusters (Figure S1.2) but indicated no evidence for Simpson's Paradox (the first cluster differed from the regression over the entire sample, the second cluster did not differ from the regression over the entire sample).

   As the three diagnostic checks are performed on a bi-variate association, their ability to detect confounds depends on the variables included. For example, a cluster analysis of the association of Parasite Stress USA and Honor killings (Thornhill & Fincher, 2011) revealed one cluster (Figure S1.3). A cluster analysis of the association of Parasite Stress USA and IQ (Eppig, Fincher, & Thornhill, 2011) revealed two clusters with slopes that differed from the slope of the total sample (Figure S1.4). However, in all three examples Parasite Stress USA was confounded with race to the same extent.
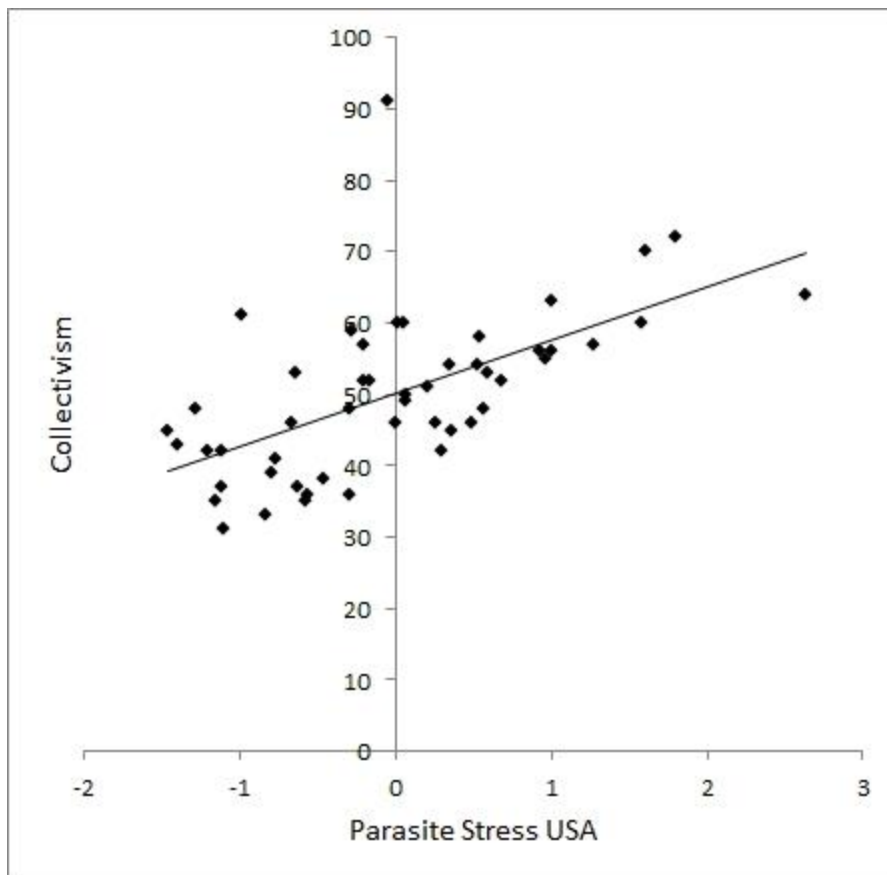

Figure S1.1. Scatterplot and regression line for values of Parasite Stress USA and Collectivism for the 50 states of the USA.
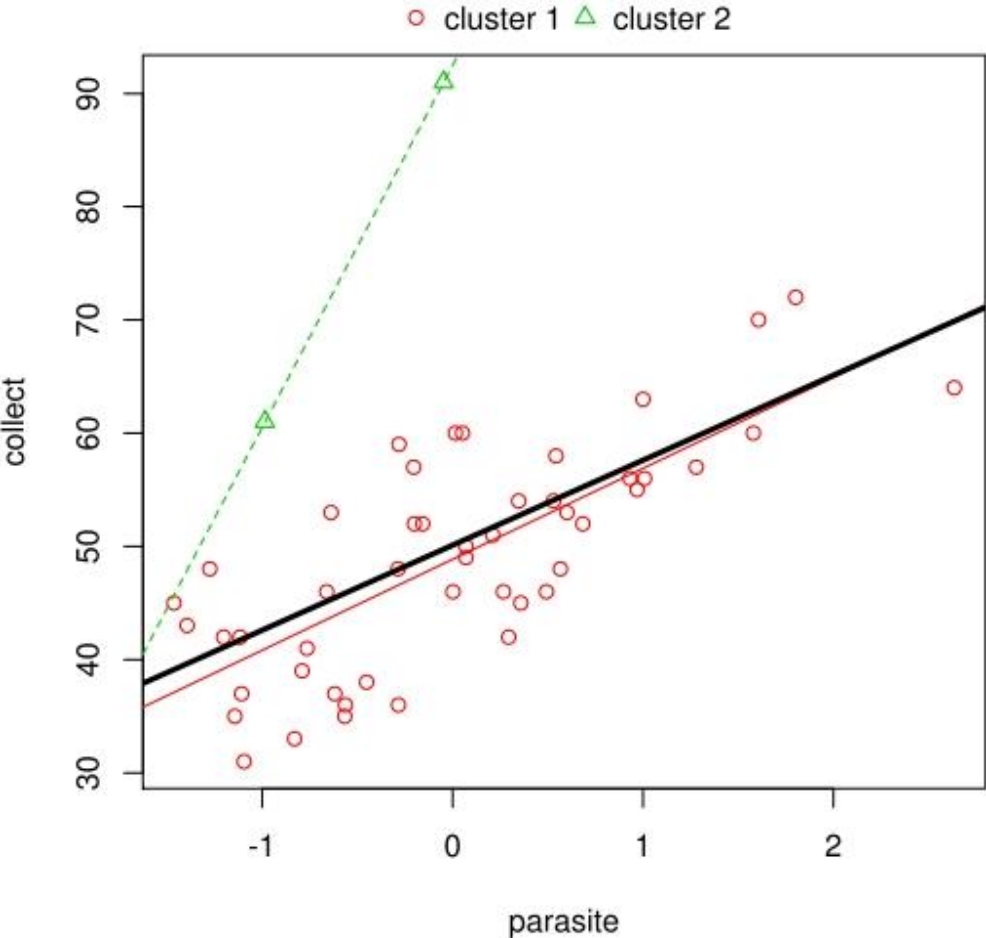
Figure S1.2. Detail of R output of cluster analysis. The horizontal axis shows Parasite Stress USA, the vertical axis shows Collectivism.
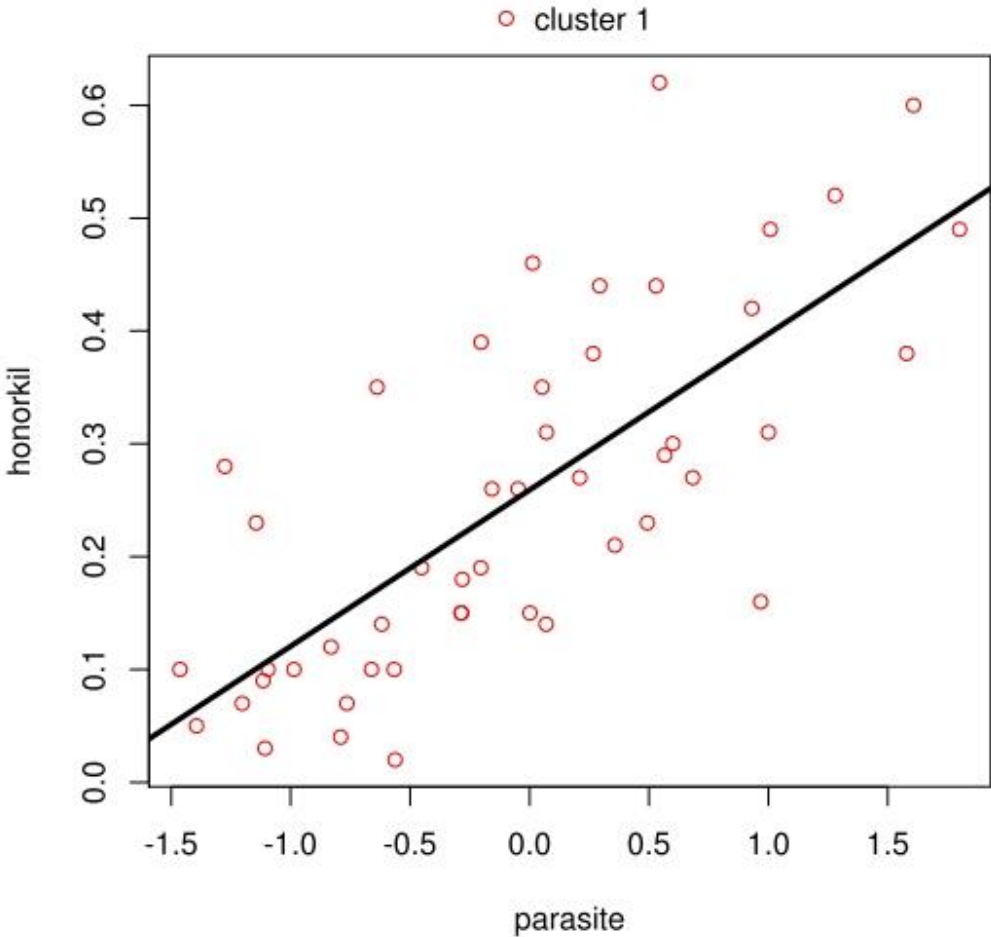
Figure S1.3. Detail of R output of cluster analysis. The horizontal axis shows Parasite Stress USA, the vertical axis shows Honor killings.
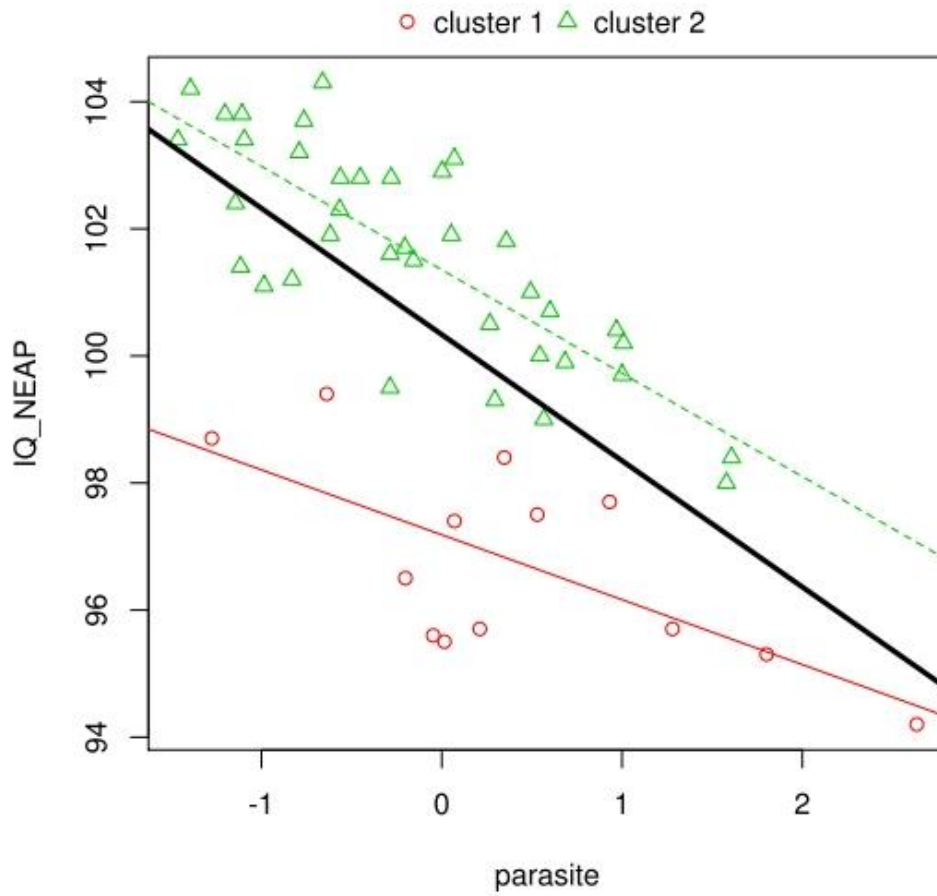
Figure S1.4. Detail of R output of cluster analysis. The horizontal axis shows Parasite Stress USA, the vertical axis shows IQ.

S2

African American CG cases strongly influenced the numerator of CG rates$_{total}$ (and by implication Parasite Stress USA), but do African American population sizes have an equally strong influence on the denominator? As can be seen in Table S2, removing the African American population from the total population did not substantially reduce the correlations of CG rates$_{total}$ with Parasite Stress USA or %Black. By contrast, removing the non-Hispanic White population from the total population did substantially reduce the correlations. Together with the results reported in Table 2, we see that African Americans had a dominant influence on the numerator of CG rates$_{total}$ (and by implication Parasite Stress USA) but little influence on the denominator.

Table S2. Correlations among Parasite Stress USA, population percentage African American (%Black), and CG rates$_{total}$. Modified CG rates$_{total}$ were produced by removing African American or non-Hispanic White components of the population. The formulas for various calculations for CG rates are also shown.

	CG rates$_{total}$	No pop$_{black}$	No pop$_{white}$
	$CG_{total} /$ pop$_{total}$	$CG_{total} /$ $(pop_{total} - pop_{black})$	$CG_{total} /$ $(pop_{total} - pop_{white})$
Parasite Stress USA	.96***	.87***	.53***
%Black	.80***	.97***	.62***
CG rates$_{total}$	-	.98***	.72***

S3

       Table S3 shows that modified incarceration rates behaved the same way as modified CG rates (see Table 2) indicating that incarceration rates$_{total}$ is confounded with %Black due to misaggregation. Thus, incarceration rates demonstrate the same indicators of misaggregation as CG rates$_{total}$. This provides support for our claim that cross-race misaggregation is important for variables other than CG rates.

Table S3. Correlations among Parasite Stress USA, population percentage African American (%Black), and incarceration rates$_{total}$. Modified incarceration rates$_{total}$ was produced by excluding African American or non-Hispanic White incarceration cases from the total number of cases, or including only African American or non-Hispanic White incarceration cases. The formulas for various calculations of incarceration rates are shown.

	incarceration rates$_{total}$	No cases$_{black}$	No cases$_{white}$	Only cases$_{black}$	Only cases$_{white}$
	cases$_{total}$ / pop$_{total}$	(cases$_{total}$ − cases$_{black}$) / pop$_{total}$	(cases$_{total}$ − cases$_{white}$) / pop$_{total}$	cases$_{black}$ / pop$_{total}$	cases$_{white}$ / pop$_{total}$
%Black	.65***	-.35*	.81***	.96***	-.12
incarceration rates$_{total}$	-	.45**	.86***	.76***	.48**

$^{+} p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$