Of Beauty, Sex and Power

Too little attention has been paid to the statistical challenges in estimating small effects

Andrew Gelman and David Weakliem

In the past few years, Satoshi Kanazawa, a reader in management and research methodology at the London School of Economics, published a series of papers in the *Journal of Theoretical Biology* with titles such as "Big and Tall Parents Have More Sons" (2005), "Violent Men Have More Sons" (2006), "Engineers Have More Sons, Nurses Have More Daughters" (2005), and "Beautiful Parents Have More Daughters" (2007). More recently, he has publicized some of these claims in an article, "10 Politically Incorrect Truths About Human Nature," for *Psychology Today* and in a book written with Alan S. Miller, *Why Beautiful People Have More Daughters*.

However, the statistical analysis underlying Kanazawa's claims has been shown to have basic flaws, with some of his analyses making the error of controlling for an intermediate outcome in estimating a causal effect, and another analysis being subject to multiple-comparisons problems. These are technical errors (about which more later) that produce misleading results. In short, Kanazawa's findings are not statistically significant, and the patterns he analyzed could well have occurred by chance. Had the lack of statistical significance been noticed in the review process, these articles would almost certainly not have been published in the journal. The fact of their appearance (and their prominence in the media and a popular book) leads to an interesting statistical question: How should we think about research findings that are intriguing but not statistically significant? A quick answer would be to simply ignore them: After all, anyone armed with even a simple statistics package can go through public databases fishing for correlations to confirm a preexisting hypothesis. That dismissal would be too glib, however, because even nonsignificant findings can be suggestive. For example, an analysis might find the probability of a girl birth to be 5 percent more likely for attractive than for unattractive parents, but with a standard error of 4 percent. Not statistically significant, but if we had to guess whether girls are more likely to be born to beautiful or ugly parents, the data would suggest the former.

There are other substantive reasons why Kanazawa's hypothesis should not be dismissed out of hand, even though his results are not statistically significant. For example, his findings are motivated theoretically by a well-respected model put forward by Robert Trivers and Dan Willard in a classic 1973 paper. The Trivers-Willard Hypothesis suggests that if a heritable attribute is more beneficial to children of one sex than the other, then parents will bear relatively more offspring of that sex. In addition, research has found that beautiful people are more liked and more respected, and the sex of children has been found to influence the attitudes of parents, with, for example, politicians with girl children staking more liberal positions on women's issues, compared to politicians with boy children.

The present article focuses on the question of how to interpret nonsignificant results. We also touch on distortions of equivocal findings by the popular media. Considering the way that more careful statisticians and quantitative social scientists tend to surround their statements in clouds of qualifications and technical jargon, it is no surprise that reporters get influenced by brash overstatements. One reason this topic is important is that systematic errors such as overestimation of the magnitudes of small effects can mislead scientists and, through them, the general public.

Throughout, we use the term *statistically significant* in the conventional way, to mean that an estimate is at least two standard errors away from some "null hypothesis" or prespecified value that would indicate no effect present. An estimate is statistically insignificant if the observed value could reasonably be explained by simple chance variation, much in the way that a sequence of 20 coin tosses might happen to come up 8 heads and 12 tails; we would say that this result is not statistically significantly different from chance. More precisely, the observed proportion of heads is 40 percent but with a standard error of 11 percent—thus, the data are less than two standard errors away from the null hypothesis of 50 percent, and the outcome could clearly have occurred by chance. *Standard error* is a measure of the variation in an estimate and gets smaller as a sample size gets larger, converging on zero as the sample increases in size.

Classical and Bayesian Inference for Small Effects

We focus on Kanazawa's 2007 analysis of data from the National Longitudinal Study of Adolescent Health, which concluded that beautiful parents have more daughters. The study included interviewers' subjective assessments of respondents' attractiveness (on a 1–5 scale) along with data including the sexes of respondents' children (if any). For a group of just under 3,000

Andrew Gelman teaches in the Department of Statistics and the Department of Political Science, Columbia University, New York. David Weakliem teaches in the Department of Sociology, University of Connecticut. Address for Gelman: Columbia University, 1016 Social Work Building, New York, NY 10027. Internet: gelman@stat.columbia.edu

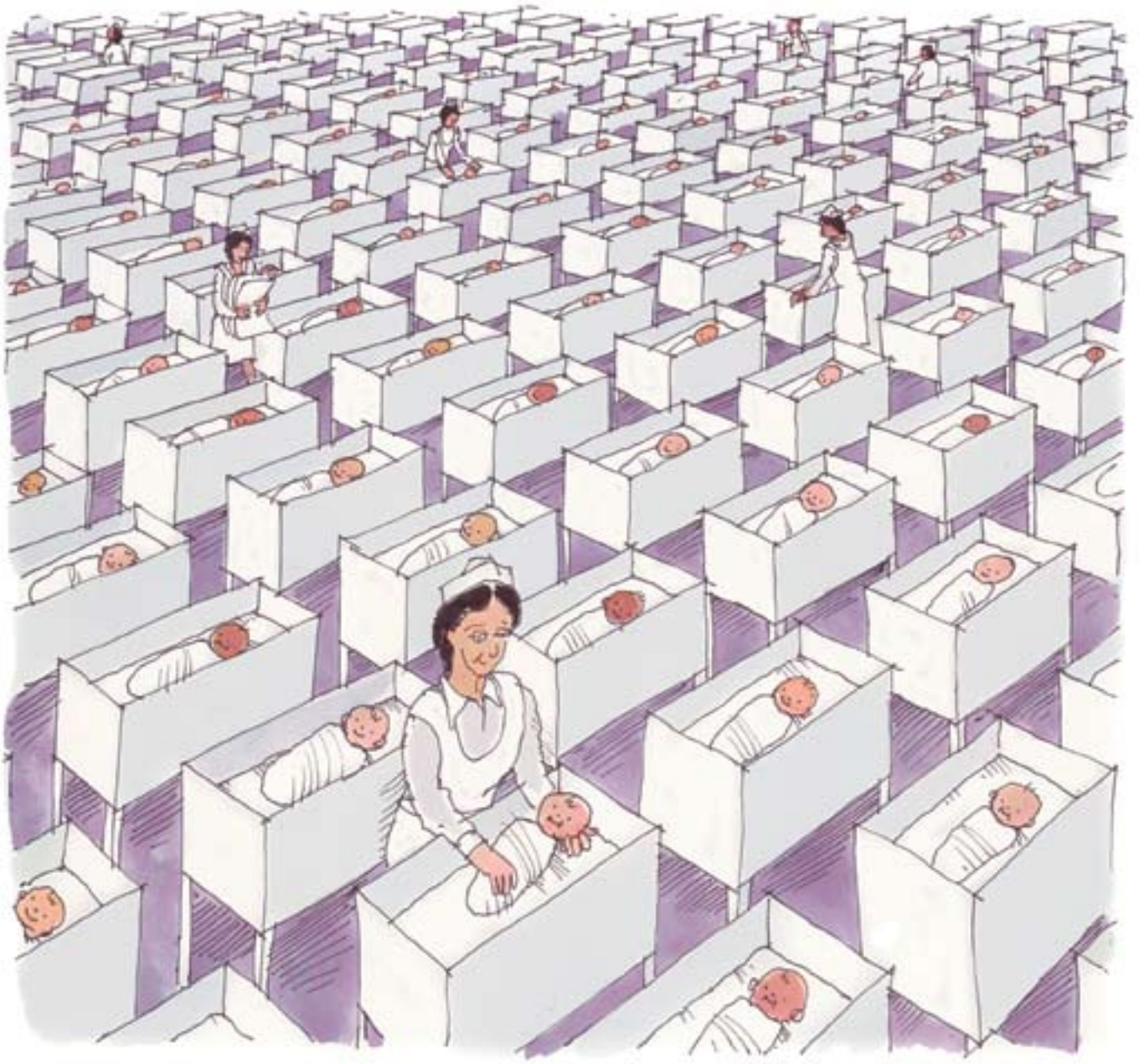

Figure 1. In 1973, Robert Trivers and Dan Willard hypothesized that sex ratio (boy births to girl births) was responsive to heritable traits that are selectively more advantageous to offspring of one sex than the other. Factors that affect sex ratio (race, maternal health and others) have been studied frequently. Any effects found are small—typically less than one percent. A series of recent sex-ratio studies reported much larger deviations, including a report that attractive parents were 26 percent less likely to have a son than a daughter as the first child. Published analyses followed, documenting methodological flaws in the studies and shrinking the findings below the level of statistical significance. The question arises: What is to be made of results that are suggestive but statistically insignificant, especially in an era when statistically insignificant findings may be widely reported as news?

parents, Kanazawa reported a statistically significant 8 percentage point difference—a 52 percent chance of girl births for the parents in the highest attractiveness category, compared to a 44 percent chance for the average of the four lower categories. However, as one of us (Gelman) explained in a letter published in the *Journal of Theoretical Biology*, comparing the top category to the bottom four is just one of the many possible comparisons that could be performed with these data—this is the multiple-comparisons problem mentioned earlier.

We are in the all-too-common situation of seeing a pattern in data that, in the jargon of social science, is suggestive without being statistically significant—that is, it could plausibly have occurred by chance alone, but it still provides some evidence in favor of a proposed model. As statisticians and social scientists, how can we frame the "suggestive but not statistically significant" problem? The key will be to think about possible effect sizes: As we shall discuss, based on the scientific literature it is just possible that beautiful parents are 1 percent more likely than others to have a girl baby, but it is implausible that the difference could be anything on the order of 5 percent. Using this as an example, we consider how the problem can be framed in two leading statistical paradigms: *classical inference,* which is based on hypothesis testing and statistical significance, and in which external scientific informa-

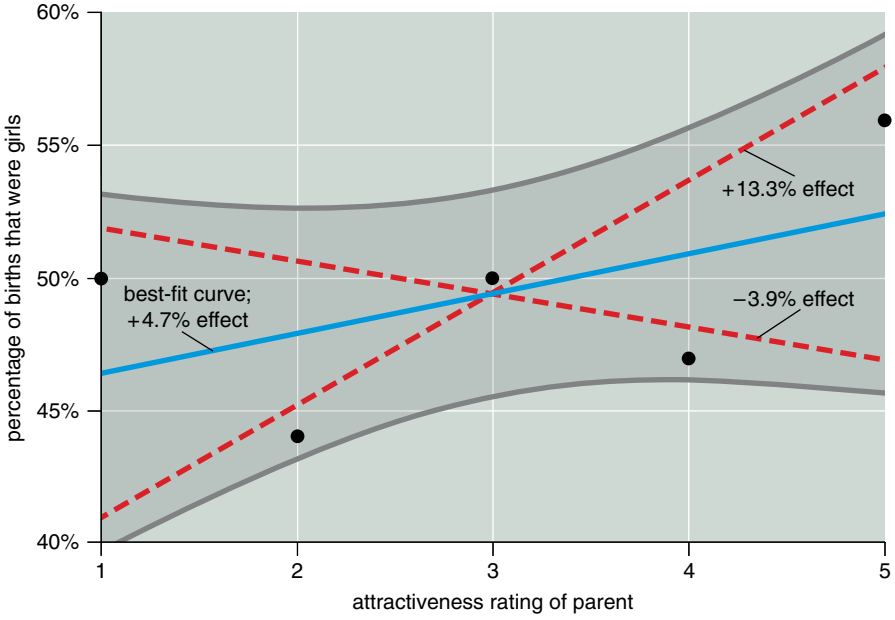

Figure 2. Using a sample of 2,972 respondents from the National Longitudinal Study of Adolescent Health, each of whom had been rated on a five-point scale of attractiveness, Satoshi Kanazawa compared the most attractive group to the average of the other four groups and concluded that there was a higher probability that the first-born child of the most attractive respondents would be a girl. Using the same data, a more usual methodology determines the best-fit curve for all of the data, with a standard error based on the size of the study. A classical 95 percent confidence interval reveals that effects as low as –3.9 percent and as high as 13.3 percent are consistent with these data. Thus, the finding of a 4.7 percent effect delivers little information about the association of parental beauty with the sex ratio of offspring in the population.

tion is encoded as a family of "null hypotheses"; and *Bayesian inference,* in which external information is encoded as a "prior distribution."

Taking the data presented in Kanazawa's article, we followed a standard analysis predicting the probability of girl births from the numerical attractiveness measure and found that more attractive parents were 4.7 percent more likely to have girls, with a standard error of 0.043, or 4.3 percent. The challenge is to interpret this finding, which is consistent with an existing hypothesis but is not statistically significant.

First we must recognize that the effects being studied are likely to be small. There is a large literature on variation in the sex ratio of human births, and the effects that have been found have been on the order of 1 percentage point (for example, the probability of a girl birth shifting from 48.5 percent to 49.5 percent). Variation attributable to factors such as race, parental age, birth order, maternal weight, partnership status and season of birth is estimated at from less than 0.3 percentage points to about 2 percentage points, with larger changes (as high as 3 percentage points) arising under economic conditions of poverty and famine. That extreme deprivation increases the proportion of girl births is no surprise, given reliable findings that male fetuses (and also male babies and adults) are more likely than females to die under adverse conditions. Based on our literature review, we would expect any effects of beauty on the sex ratio to be less than 1 percentage point, which represents the range of natural variation under normal conditions.

Classical Inference

Returning to our example: With an estimate of 4.7 percent and a standard error of 4.3 percent, the classical 95 percent confidence interval for the difference in probability of a girl birth, comparing attractive to unattractive parents, is [−3.9 percent, 13.3 percent]. To put it another way, effects as low as −3.9 percent or as high as +13.3 percent are roughly consistent with the data. Given that we only expect to see effects in the range of ±1 percent, we have essentially learned nothing from this study.

Another way to frame this is to consider what would happen if repeated independent studies were performed with the same precision, and thus, approximately the same standard error of 4.3 percent. Working with a 95 percent confidence interval, there is at minimum a 5 percent chance of obtaining a statistically significant result, which would imply an estimate of 8.4 percent or larger in either direction (1.96 standard deviations from zero). If multiple tests are performed, the chance of finding something statistically significant increases. In any case, though, the estimated effect—at least 8.4 percent—is much larger than anything we would realistically think the effect size could be. This is a Type M (magnitude) error: The study is constructed in such a way that any statistically significant finding will almost certainly be a huge overestimate of the true effect. In addition there will be Type S (sign) errors, in which the estimate will be in the opposite direction of the true effect. We get a sense of the probabilities of these errors by considering three scenarios of studies with standard errors of 4.3 percentage points:

1. *True difference of zero.* If there is no correlation between parental beauty and sex ratio of children, then a statistically significant estimate will occur 5 percent of the time, and it will always be misleading.

2. *True difference of 0.3 percent.* If the probability of girl births is actually 0.3 percent higher among attractive than among unattractive parents, then there is a 3 percent probability of seeing a statistically significant positive result—and a 2 percent chance of seeing a statistically significant negative result. In either case, the estimated effect, of at least 8.4 percentage points, will be over an order of magnitude higher than the true effect, and with a 2/5 chance of going in the wrong direction. If the result is not statistically significant, the chance of the estimate being in the wrong direction (a Type S error) is 47.5 percent, so close to 50 percent that the direction of the estimate provides almost no information on the sign of the true effect.

3. *True difference of 1 percent.* If the probability of girl births is actually 1 percent higher among attractive than among unattractive parents—which, based on the literature, is on the high end of possible effect sizes—then there is a 4 percent chance of a statistically significant positive result,

and still over a 1 percent chance of a statistically significant result in the wrong direction. Overall there is a 40 percent chance of a Type S error: Again, the estimate gives us little information about the sign or the magnitude of the true effect.

4. *True difference of 3 percent.* Even if the true difference were as high as 3 percent, which we find implausible from the literature review, there is still only a 10 percent chance of obtaining statistical significance, and the overall Type S error rate is 24 percent.

A study of this size is thus not fruitful for estimating variation on the scale of one or a few percent. This is one reason that successful studies of the human sex ratio use much larger samples, typically from demographic databases where the sample size can be millions.

Bayesian Inference

We can also redo Kanazawa's analysis using a Bayesian prior distribution. In Bayesian inference, the prior distribution represents information about a problem from sources external to the data currently being analyzed. A diffuse prior distribution is one that conveys essentially no information beyond the data at hand. To start with, given a sufficiently diffuse prior distribution, the posterior distribution would be approximately normal with a mean of 4.7 percent and a standard error of 4.3 percent, which would imply about an 86 percent probability that the true effect is positive. In general, the more concentrated the prior distribution around zero (a presumption based on the sex-ratio literature that the true effect is likely to be small), the closer the posterior probability will be to 50 percent.

For example, consider a bell-shaped distribution with center zero and with a shape such that the true difference in percentage of girls, comparing beautiful and ugly parents, is most likely to be near zero, with a 50-percent chance of being in the range [−0.3 percent, 0.3 percent], a 90-percent chance of being in the range [−1 percent, 1 percent], and a 94-percent chance of being less than 3 percentage points in absolute value. We center the prior distribution at zero because, ahead of time, we have no particular reason to believe that the true difference in the probability of girl births, comparing attractive and unattractive parents in the general population, will be positive or negative.

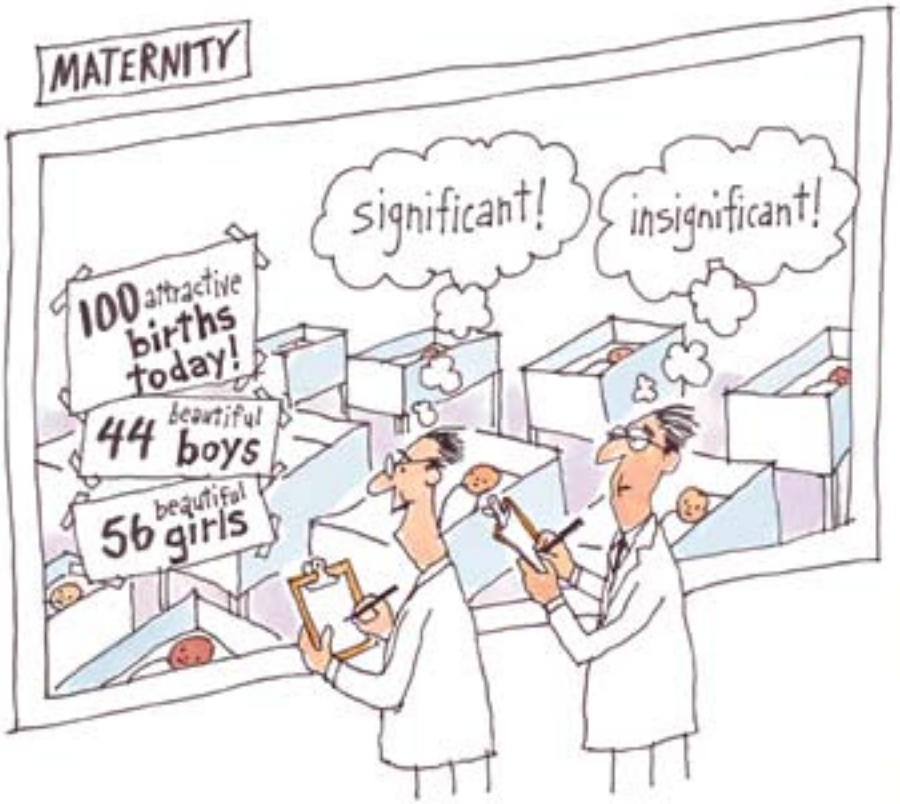

Figure 3. "Statistical methods of analysis are intended to aid the interpretation of data that are subject to appreciable haphazard variability," wrote D. R. Cox and D. V. Hinkley in their book *Theoretical Statistics*. Statistical findings are not in the eye of the beholder, but they require awareness of statistical reasoning. Uncertainty is inevitable and is summarized by a probability distribution. Fifty-six is decidedly more than 44, but 56 girls and 44 boys could happen by chance in any nursery.

The next step is to perform the calculations of the probability of different effect sizes, given the prior distribution and the data. Briefly, the resulting posterior distribution gives a probability that the difference is positive—that beautiful parents actually have more daughters—of only 58 percent—and even if the effect is positive, there is a 78-percent chance it is less than 1 percentage point. This analysis depends on the prior distribution but not to an extreme extent; for example, if we broaden the distribution curve, increasing the range of inclusion for outliers, the posterior probability that the true difference is positive is still only 65 percent. Switching between families of distribution curves has little effect on the results. The key is that effects are likely to be small, and in fact the data are consistent with small results.

The ideal for scientific understanding about a quantity (in this case, the correlation between beauty of parents and sex ratio of children) is to have a recognized uncertainty that can be summarized by a probability distribution. Individual researchers can collect data or creatively analyze existing sources (as was done by Kanazawa) and publish their results, and then occasional meta-analyses can be done to review the results. This procedure smooths some of the variation that is inherent in these small-sample studies, where the probability of a positive effect can jump from 50 percent to 58 percent, then perhaps down to 38 percent with the next study, and so forth.

The 50 Most Beautiful People

One way to calibrate our thinking about Kanazawa's results is to collect more data. Every year, *People* magazine publishes a list of the 50 most beautiful people, and, because they are celebrities, it is not difficult to track down the sexes of their children, which we did for the years 1995–2000. Data were collected from Wikipedia, the Internet Movie Database and celebrities' personal Web pages, using a cutoff date of August 2007. Information was missing for two beautiful people in 1995, two in 1996, three in 1997, six in 1998, three in 1999, and two in 2000.

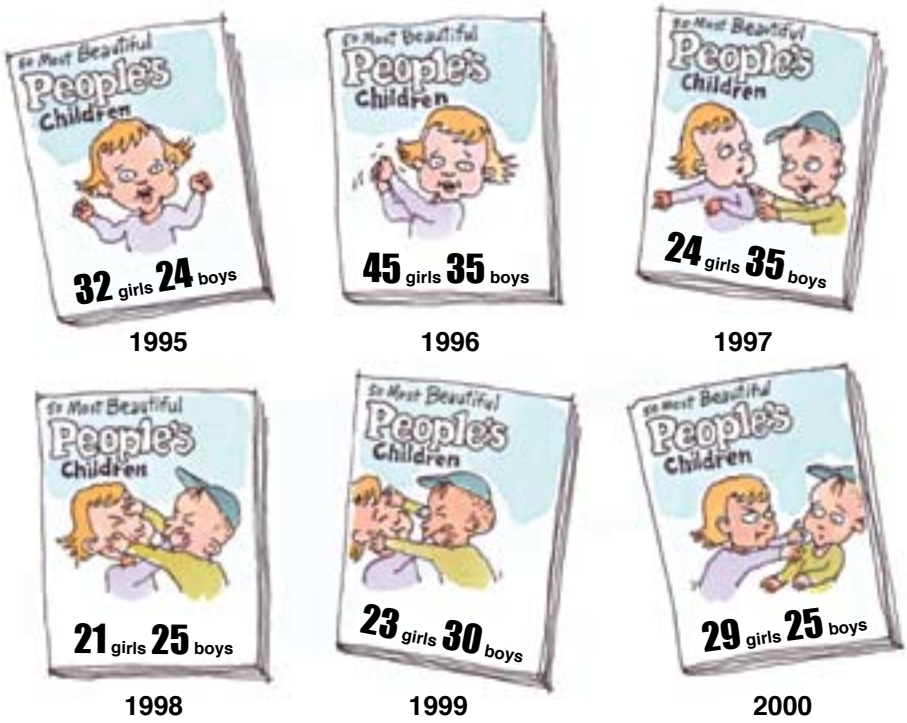

Figure 4. The authors performed a sex-ratio study of the offspring of the most beautiful people in the world as selected by *People* magazine between 1995 and 2000. The girls started strong in 1995 with 32 girls to 24 boys. Girls continued strong in 1996. However, as the sample size grew, the ratio converged on the population frequency, concluding with 157 girls and 172 boys, or 47.7 percent girls, approximately the same as the population frequency of 48.5 percent.

The data are available for download at http://www.stat.columbia.edu/~gelman/research/beautiful/

As of 2007, the 50 most beautiful people of 1995 had 32 girls and 24 boys, or 57.1 percent girls, which is 8.6 percentage points higher than the population frequency of 48.5 percent. This sounds like good news for the hypothesis. But the standard error is $0.5/\sqrt{(32 + 24)} = 6.7$ percent, so the discrepancy is not statistically significant. Let's get more data.

The 50 most beautiful people of 1996 had 45 girls and 35 boys: 56.2 percent girls, or 7.8 percent more than in the general population. Good news! Combining with 1995 yields 56.6 percent girls—8.1 percent more than expected—with a standard error of 4.3 percent, tantalizingly close to statistical significance. Let's continue to get some confirming evidence.

The 50 most beautiful people of 1997 had 24 girls and 35 boys—no, this goes in the wrong direction, let's keep going…For 1998, we have 21 girls and 25 boys, for 1999 we have 23 girls and 30 boys, and the class of 2000 has had 29 girls and 25 boys. Putting all the years together and removing the duplicates, such as Brad Pitt, *People*'s most beautiful people from 1995 to 2000 have had 157 girls out of 329 children, or 47.7 percent girls (with a standard error of 2.8 percent), a statistically insignificant 0.8 percentage points lower than the population frequency. So nothing much seems to be going on here. But if statistically insignificant effects were considered acceptable, we could publish a paper every two years with the data from the latest "most beautiful people."

Why Is This Important?

Why does this matter? Why are we wasting our time on a series of papers with statistical errors that happen not to have been noticed by a journal's reviewers? We have two reasons: First, as discussed in the next section, the statistical difficulties arise more generally with findings that are suggestive but not statistically significant. Second, as we discuss presently, the structure of scientific publication and media attention seem to have a biasing effect on social science research.

Before reaching *Psychology Today* and book publication, Kanazawa's findings received broad attention in the news media. For example, the popular Freakonomics blog reported,

> A new study by Satoshi Kanazawa, an evolutionary psychologist at the London School of Economics, suggests . . . there are more beautiful women in the world than there are handsome men. Why? Kanazawa argues it's because good-looking parents are 36 percent more likely to have a baby daughter as their first child than a baby son—which suggests, evolutionarily speaking, that beauty is a trait more valuable for women than for men. The study was conducted with data from 3,000 Americans, derived from the National Longitudinal Study of Adolescent Health, and was published in the *Journal of Theoretical Biology*.

Publication in a peer-reviewed journal seemed to have removed all skepticism, which is noteworthy given that the authors of Freakonomics are themselves well qualified to judge social science research.

In addition, the estimated effect grew during the reporting. As noted above, the 4.7 percent (and not statistically significant) difference in the data became 8 percent in Kanazawa's choice of the largest comparison (most attractive group versus the average of the four least attractive groups), which then became 26 percent when reported as a logistic regression coefficient, and then jumped to 36 percent for reasons unknown (possibly a typo in a newspaper report). The funny thing is that the reported 36 percent signaled to us right away that something was wrong, since it was 10 to 100 times larger than reported sex-ratio effects in the biological literature. Our reaction when seeing such large estimates was not "Wow, they've found something big!" but, rather, "Wow, this study is underpowered!" Statistical power refers to the probability that a study will find a statistically significant effect if one is actually present. For a given true effect size, studies with larger samples have more power. As we have discussed here, "underpowered" studies are unlikely to reach statistical significance and, perhaps more importantly, they drastically overestimate effect size estimates. Simply put, the noise is stronger than the signal.

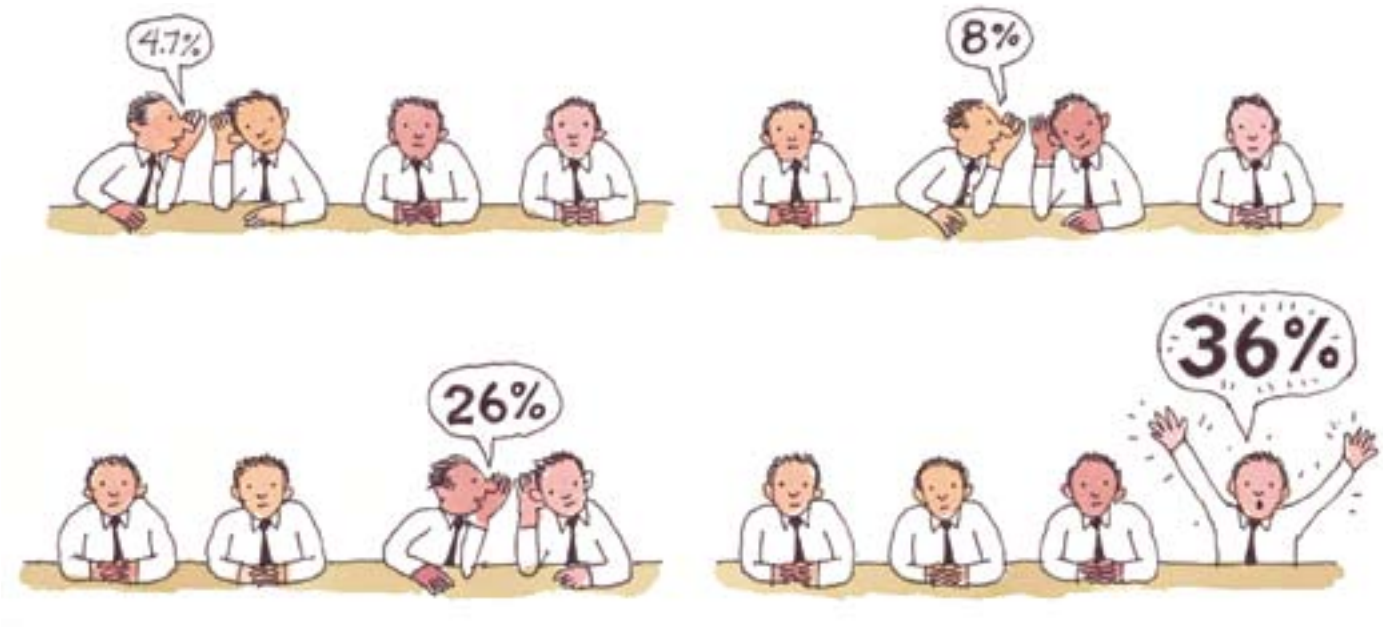

Figure 5. "A statistical analysis, properly conducted, is a delicate dissection of uncertainties," wrote M. J. Moroney in *Facts from Figures*, published in 1951. Moroney's comment is especially true for small effects. Random findings can be misidentified as significant. Methodological problems can be overlooked by reviewers. And popular reporting can misrepresent or even exaggerate the original findings. The dangers are especially present if the finding is suggestive and the study is underpowered for the job of resolving the correlations in the data.

This problem will occur again and again, and is worth thinking about now. To start with, most of the low-hanging fruit in social science research has presumably been plucked, leaving researchers to study small effects. Sex ratios are of inherent interest to all of us who have or are considering having babies, as well as for their implications for the organization of society. Miller and Kanazawa's billing of their result as a "politically incorrect truth" hints at the connection to live political issues such as abortion, parental leave policies and comparable-worth laws that turn upon judgments of the appropriate roles for men and women in society.

As we discussed earlier in this article, studies with insufficient statistical power will spit out random results that will occasionally be statistically significant and, even more often, be suggestive, as in Kanazawa's beauty-and-births studies. It is tempting to interpret the directions of these essentially random findings without recognizing the fragility of the explanations we construct. Evolutionary psychology could be used to explain a result in the opposite direction, using the following sort of argument: Persons judged to be beautiful are, one could claim, more likely to be healthy, affluent and from dominant ethnic groups, more generally having traits that are valued in the society at large. (Consider, for example, Miss Americas, who until recent decades were all white.) Such groups are more likely to exercise power, a trait that, in some sociobiological arguments, is more beneficial for men than women—thus it would be natural for more attractive parents to be more likely to have boys. We are not claiming this is true; we are just noting that the argument could go in either direction, which puts a particular burden on the data analysis. The ability of this theory to explain findings in any direction was pointed out in 2007 by Jeremy Freese in the *American Journal of Sociology*, who describes this sort of argument as "more 'vampirical' than 'empirical'—unable to be killed by mere evidence."

In statistics, you can't prove a negative. "Beautiful parents have more daughters" is a compelling headline; the sounder statement is a less appealing headline: "There is no compelling evidence that beautiful parents are more or less likely to have daughters." As a result, public discourse can get cluttered with unproven claims, which perhaps will lead to a general skepticism that will, in boy-who-cried-wolf fashion, unfairly discredit more convincing research.

The result is a sort of asymmetrical warfare, with proponents of sex differences and other "politically incorrect" results producing a series of empirical papers that, for reasons of inadequate statistical power, give essentially random clues about true population patterns, and opponents of this line of research being reduced to statements such as "the data are insufficient." The aforementioned Freakonomics article concluded, "It is good that Kanazawa is only a researcher and not, say, the president of Harvard. If he were, that last finding about scientists may have gotten him fired." It should be possible to criticize large unproven claims in biology and social science without dismissing the entire enterprise.

Why Is This Not Obvious?

The natural reaction of a competent quantitative researcher to the statistics in this article is probably, Duh. But if this is so obvious, why did the mistake result in not one, but several papers in the *Journal of Theoretical Biology,* a prominent publication with an impressive name and a respectable impact factor of 2.3 (higher than any of the three top journals in statistics, *Journal of the American Statistical Association,* the *Annals of Statistics* and the *Journal of the Royal Statistical Society*)? One problem, of course, is that referees are busy, and statistical errors can be subtle and easily overlooked. But another problem is the confusing connection between statistical significance and sample size. It is well known that, with a large enough sample size, one can just about always find statistically significant, if small, effects. But it is not so well realized that,

when effects truly are small, there is little point in trying to find them with underpowered studies.

These problems are not new, even in the field of sex ratios. For example, in a 1957 book with the unfortunate title of *Probability, Statistics, and Truth,* Richard von Mises studied the sex ratios of births in the 24 months of 1907–1908 in Vienna and found less variation than would be expected from chance alone. He attributed this to different sex ratios in different ethnic groups. In fact, however, the variance, though less than expected by chance, was not statistically significantly less. There seems to be a human desire to find more than pure randomness in sex ratios, despite there being no convincing evidence that sex ratios vary much at all except under extraordinary conditions.

Realistically, a researcher on sex ratios has to make two arguments: a statistical case that observed patterns represent real population effects and cannot be explained simply by sampling variability, and a biological argument that effects on the order of 1 percent are substantively important. The claimed effect size of 26 percent should have aroused suspicion in comparison to the literature on human sex ratios; in addition, though, the papers managed to survive the review process because reviewers did not recognize that the power of the studies was such that only very large estimated effects could make it through the statistical-significance filter. The result is essentially a machine for producing exaggerated claims, which of course only become more exaggerated when they hit the credulous news media (with an estimate of 4.7 percent ± 4.3 percent being ramped up to 26 percent and then reported as 36 percent).

Statisticians should take some of the blame here. Statistics textbooks are clear enough on the concepts of statistical significance and power, but they don't provide much guidance on how to think about what to do when you get implausibly large estimates from small samples. Classical significance calculations do not make use of prior knowledge of effect sizes, and Bayesian analyses are often not much better. Textbook treatments of Bayesian inference almost entirely use noninformative prior distributions and essentially ignore issues of statistical power. Conversely, power calculations are commonly used in designing studies (to indicate how large a study should be) but are rarely used to enlighten data analyses. And theoretical concepts such as Type S and Type M errors have not been integrated into statistical practice.

The modern solution to difficulties of statistical communication is to have more open exchange of methods and ideas. More transparency is apparently needed, however: For example, *Psychology Today* did not seem to notice the published critique of Kanazawa's findings in the *Journal of Theoretical Biology* or several other methodological criticisms that have appeared in sociology journals. We hope that a more systematic way of understanding estimates of small effects will provide a clearer framework for open communication.

Acknowledgments

We thank Kelly Rader and Lucy Robinson for data collection and analysis, and David Dunson, Thomas Volscho, David Judkins and Jeremy Freese for helpful comments.

References

Almond, D., and L. Edlund. 2006. Trivers-Willard at birth and one year: Evidence from U.S. natality data 1983–2001. *Proceedings of the Royal Society B* 274:2491–2496.

Cagnacci, A., A. Renzi, S. Arangino, C. Alessandrini, and A. Volpe. 2005. Interplay between maternal weight and seasons in determining the secondary sex ratio of human offspring. *Fertility and Sterility* 84:246–248.

Carlin, B. P., and T. A. Louis. 2001. *Bayes and Empirical Bayes Methods for Data Analysis*, second edition. London: CRC Press.

Catalano, R. A. 2003. Sex ratios in the two Germanies: A test of the economic stress hypothesis. *Human Reproduction* 18:1972–1975.

Chahnazarian, A. 1988. Determinants of the sex ratio at birth: Review of recent literature. *Social Biology* 35:214–235.

Dubner, S. J. 2006. Why do beautiful women sometimes marry unattractive men? http://freakonomics.blogs.nytimes.com/ 2 August, 9:44 am.

Fawcett, T. W., B. Kuijper, I. Pen, and F. J. Weissing. 2007. Should attractive males have more sons? *Behavioral Ecology* 18:71–80.

Freese, J. 2007. The problem of predictive promiscuity in deductive applications of evolutionary reasoning to intergenerational transfers: Three cautionary tales. In *Caring and Exchange Within and Across Generations,* ed. A. Booth *et al.* Washington, D.C.: Urban Institute Press.

Freese, J., and B. Powell. 2001. Making love out of nothing at all? Null findings and the Trivers-Willard hypothesis. *American Journal of Sociology* 106:1776–1788.

Gelman, A. 2007. Letter to the editor regarding some papers of Dr. Satoshi Kanazawa. *Journal of Theoretical Biology* 245:597–599.

Gelman, A. 2007. Scaling regression inputs by dividing by two standard deviations. Technical report, Department of Statistics, Columbia University.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin. 2003. *Bayesian Data Analysis,* second edition. London: CRC Press.

Gelman, A., and A. Jakulin. 2007. Bayes: Liberal, radical, or conservative? *Statistica Sinica* 17:422–426.

Gelman, A., and F. Tuerlinckx. 2000. Type S error rates for classical and Bayesian single- and multiple comparison procedures. *Computational Statistics* 15:373–390.

Hamermesh, D. S., and J. E. Biddle. 1994. Beauty and the labor market. *AmericanEconomic Review* 84:1174–1194.

James, W. H. 1987. The human sex ratio. I: A review of the literature. *Human Biology* 59:721–752.

Kanazawa, S. 2005. Big and tall parents have more sons: Further generalizations of the Trivers-Willard hypothesis. *Journal of Theoretical Biology* 233:583–590.

Kanazawa, S., and G. Vandermassen. 2005. Engineers have more sons, nurses have more-daughters: An evolutionary psychological extension of Baron-Cohen's extreme male brain theory of autism. *Journal of Theoretical Biology* 233:589–599.

Kanazawa, S. 2006. Violent men have more sons: Further evidence for the generalized Trivers-Willard hypothesis. *Journal of Theoretical Biology* 239:450–459.

Kanazawa, S. 2007. Beautiful parents have more daughters: A further implication of the generalized Trivers-Willard hypothesis. *Journal of Theoretical Biology* 244:133–140.

Miller, A. S., and S. Kanazawa. 2007. Ten politically incorrect truths about human nature. *Psychology Today,* July–August.

Miller, A. S., and S. Kanazawa. 2007. *Why Beautiful People Have More Daughters*. New York: Perigee.

Pinker, S. 2007. In defense of dangerous ideas. *Chicago Sun-Times,* July 15.

Rubin, D. B. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* 12:1151–1172.

Trivers, R. L, and D. E. Willard. 1973. Natural selection of parental ability to vary the sex ratio of offspring. *Science* 179:90–92.

Volscho, T. W. 2005. Money and sex, the illusory universal sex difference: Comment on Kanazawa. *Sociological Quarterly* 46:716–736.

von Mises, R. 1957. *Probability, Statistics, and Truth,* second edition. New York: Dover.

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/issues/id.79/past.aspx