Title
The human sex ratio in New York City did not change after 11 September 2001.

Journal
Human reproduction (Oxford, England), 22(4)

ISSN
0268-1161

Author
Lydersen, Stian

Publication Date
2007-04-04

Peer reviewed

decreased significantly (i.e. $P < 0.01$, two-tailed test) in New York City after 11 September 2001. First, the *P*-value quoted in the summary seems to be a misprint. The main finding in the article has $P < 0.05$, two-tailed test. Second, the result is misleading, due to inadequate statistical analysis (lack of adjustment for multiple tests). In fact, the sex ratios reported in the article have the mean and variation to be expected from history. They show no evidence of change after 11 September 2001. Fortunately, the authors have reported their methods and data in a way that make a closer look at the results possible.

Figure 1 in the article shows the observed sex ratio for 91 periods of 28 days, including the 10 periods after 11 September 2001. The main finding is a low sex ratio of 0.9995 in the period 1–28 January 2002. On the average, there are about 7900 births per period. Assuming a constant expected sex ratio of 1.0510, the amount of variation in Figure 1 is exactly what is expected from a binomial distribution. The two-sided probability of observing a ratio of 0.9995 or more extreme in one period is 0.026. An observation like that among 91 periods is to be expected. The authors hypothesize a lower than expected sex ratio in one or more of the 3 months centred on December 2001, based on other studies. Hence, a proper analysis of this hypothesis must account for all three periods, that is, adjusting for multiple tests. The three periods have an observed average sex ratio of 1.042 (1.049, 1.077 and 0.9995, respectively) read from Figure 1 in the article. Assuming a constant expected sex ratio of 1.0510, the probability of observing an average of 1.042 or more extreme is 0.49.

The authors analyse the data differently, using a time-series model, to account for possible correlations between periods or seasonal variations. Their Table I reports the results for each of the 10 periods after 11 September 2001. Only one of them, namely 1–28 January 2002, exhibits $P < 0.05$. The exact *P*-values are not given, but may be calculated from the reported estimates. The *z*-value for the period 1–28 January 2002 is $-0.0484/0.0207 = -2.34$, and the corresponding two-sided $P = 0.019$. An appropriate way to combine the 10 periods is to use the normalized average *z*-value, which is $(0.0240/0.0191 + \cdots + (-0.0144)/0.0190)/\sqrt{10} = 0.17$, giving $P = 0.86$. This corresponds to the Stouffer method for combining *P*-values. Similarly, combining the 3 hypothesized months centred on December 2001 gives $P = 0.68$.

Hence, making proper adjustment for multiple tests, there is no evidence of change in sex ratio in New York city after 11 September 2001. On the contrary, the observed sex ratios are as expected from before the event.

Reference

Catalano R, Bruckner T, Marks AR and Eskenazi B (2006) Exogenous shocks to the 2 human sex ratio: the case of 11 September 2001 in New York City, Hum Reprod. Advance 3 Access Published Online on 26 August 2006. Human Reprod, 4, doi:10.1093/humrep/del283.

Stian Lydersen[1]

Unit for Applied Clinical Research,
Department of Cancer Research and Molecular Medicine,
The Norwegian University of Science and Technology (NTNU),
The Cancer Building, 5th floor,
St Olavs Hospital, N-7006 Trondheim,
Norway

[1]To whom correspondence should be addressed at:
E-mail: stian.lydersen@ntnu.no

doi:10.1093/humrep/del475
Advance Access publication January 4, 2007

Reply: The human sex ratio in New York City did not change after 11 September 2001

Sir,

We thank Professor Lydersen for carefully considering our work (Catalano *et al.*, 2006). He argues that we misled readers when we concluded that the sex ratio of 0.9995 observed in New York City in January 2002 (the lowest among the 91 monthly cohorts we studied) would not be expected from history. He bases his argument, as we understand it, on the assumption that we repeated our test three, perhaps more, times and should not cite the *P*-value from only one test as support for our hypothesis. He contends that we should have combined the *P*-values, as suggested by Stouffer (Stouffer *et al.*, 1949), for at least three of the tests. Applying the Stouffer method to the three tests leads Professor Lydersen to the inference that we should have expected the observed value of 0.9995 in January 2002.

We suggest that readers consider Professor Lydersen's criticism with two circumstances in mind. First, the Stouffer method should not be applied to our results. The method reconciles differing *P*-values from an experiment repeated on multiple samples drawn to represent the same population. Repetitions of an experiment can yield differing *P*-values because results from random samples of a population only approximate those that would be observed if the entire population were subjected to the experimental manipulation.

We did not repeat an experiment three times on three different samples of the same population. The experiment we described occurred, mercifully, only once. We, moreover, did not use samples. We, rather, tested our hypothesis with *all* the City's live births in each of three cohorts reasonably assumed in the fifth, sixth and seventh month of gestation on 11 September 2001 and, therefore, in different stages of development. We chose these cohorts because, as described in our article, research suggests that fetuses respond to maternal stress hormones roughly at the 20th week of gestation. That work does not, however, specify when that response peaks. Our earlier research in California (Catalano *et al.*, 2005)

found a low sex ratio in the cohort born in December 2001 and led us, *a priori*, to hypothesize a low sex ratio in New York City among cohorts at a similar stage of development. The California test used calendar months, whereas we used constant 28-day periods defined such that the 75th period began on 11 September 2001. We, therefore, tested the three cohorts centred roughly on December 2001. Unlike the circumstance to which the Stouffer method applies, our theory does not imply, nor did we argue, that each of these three populations, not samples, would yield the same *P*-value.

Professor Lydersen's comments make it clear that reporting our results as we did invited his criticisms. More specifically, showing results for all cohorts in gestation on 11 September in our tables may have led readers to believe that we would have inferred support for our hypothesis had low values appeared in any of these cohorts. As explained in the text, however, we had no such decision rule. We added the cohorts at the request of a reviewer who wanted to see if the data offered any support for the argument that exogenous shocks to a population lower the secondary sex ratio by lowering the sex ratio at conception. If this were true, ratios 8, 9 and 10 months after 11 September would have been low. As the fetal loss argument would predict, we did not find lower sex ratios in these cohorts.

We, in hindsight, should have warned the reader that *P*-values have ambiguous implications for our test because we analysed the population, not samples. We did not want to use page space to reprise the controversy over this ambiguity. We, instead, used the Box–Jenkins (Box *et al.*, 1994) conventions for arriving at expected values and their confidence intervals in time-series analyses. These conventions have been well developed and widely disseminated over more than three decades of use.

In the absence of randomly assigned control groups, judgment and argument, as well as statistical control, inform expectations of the dependent variable under the null hypothesis. We, accordingly, ask readers familiar with sex ratios in populations as large as New York City's how often they would expect to observe a ratio as low as 0.9995 in monthly birth cohorts? We observed only one that low among the 91 months we studied, and it appeared in a cohort that our theory suggested would have lost males *in utero*.

Readers who share Professor Lydersen's concerns, which we cannot dismiss, may wish to view our work as more exploratory, and less confirmatory, than we intended. We suggest, however, that these readers consider the results as real differences among the populations about which we theorized rather than, as the Stouffer method assumes, estimates from samples. If the reader feels compelled to combine *P*-values, we suggest that he or she consider that we have found low sex ratios in separate populations (not samples of the same population), one geographically close to and another far from, the terror attacks on New York City. These low ratios appeared, moreover, at times consistent with our theory.

References

Box G, Jenkins G and Reinsel G (1994) *Time Series Analysis: Forecasting and Control*, 3rd edn. Prentice-Hall, .

Catalano R, Bruckner T, Gould J, Eskenazi B and Anderson E (2005) Sex ratios in California following the terrorist attacks of September 11, 2001. Hum Reprod 20, 1221–1227.

Catalano R, Bruckner T, Marks AR and Eskenazi B (2006) Exogenous shocks to the human sex ratio: the case of 11 September 2001 in New York City. Hum Reprod. Advance Access published online on 26 August 2006. Human Reprod, 4, doi:10.1093/humrep/del283.

Stouffer SA, Suchman EA, DeVinney LC, Star SA and Williams Jr RM (1949) The American Soldier, Vol. 1. *Adjustment During Army Life*. Princeton University Press, .

R.Catalano[1], T.Bruckner, A.Marks and B.Eskenazi

School of Public Health, University of California, Berkeley, USA

[1]To whom correspondence should be addressed at: E-mail: rayc@berkeley.edu

doi:10.1093/humrep/del476
Advance Access publication January 4, 2007