Are the significant effects really significant?

Sir,

We would like to raise some questions about the statistical methods used by Torche and Kleinhaus (2012) in their valuable study of the effects of the Chile earthquake on gestational age and secondary sex ratio (SSR). They used significance tests to argue that they observed important effects. However, the conclusions would be different if the significance tests were performed correctly. We demonstrate the difference for the case of gestational age, but similar results follow for their analysis of effects on other response variables.

To assess the effect of the timing of earthquake exposure, Torche and Kleinhaus included nine indicator variables to indicate exposure in each of the 9 months of gestation and used an unexposed group as the reference category. The coefficients for the indicator variables thus reflected the mean difference in gestation duration between the unexposed baseline group and those exposed in that month.

In order to rule out the influence of any temporal trend affecting both earthquake and non-earthquake regions, they ran the regressions separately for the affected and unaffected regions. They pointed out this amounted to a ‘difference-in-difference’ approach. Using this approach, the net effect of earthquake exposure should be the difference between each coefficient in the affected region and the corresponding coefficient in the unaffected region. It is possible to construct a confidence interval or assess the statistical significance of this effect based on the fact that the standard error of the net effect is equal to the sum of standard errors of the two coefficients. However, the authors did not do this.

They claimed ‘if we find significant differences in the outcomes of interest between the exposed and unexposed period in the affected region, but no differences in the unaffected region, this will provide strong evidence for an effect of earthquake exposure on birth outcomes that cannot be attributed to confounding factors’ (Torche and Kleinhaus, 2012, p. 560). The ‘effect that cannot be attributed to confounding factors’ should be the net effect, which is the difference between the two coefficients. However, the interpretation ‘if X is statistically significant but Y is not, then the difference between X and Y is statistically significant’ is not necessarily true. Gelman and Stern (2006) give several examples in which this reasoning fails to hold. It fails in the same way in Torche and Kleinhaus (2012).

In Table II of their paper, the coefficient for the variable ‘Exposed m3’ was statistically significant in the affected region (coefficient = −0.271, \( P = 0.005 \)), but not in the unaffected region (coefficient = −0.104, \( P = 0.306 \)). A similar pattern held for the variable ‘Exposed m2’. The authors claimed that ‘earthquake exposure in Month 2 and 3 of exposure resulted in a significant decline in gestational age’ (Torche and Kleinhaus, p. 558). However, the net effect for Month 3 was −0.167 and, using the standard error calculation above, not statistically significant (standard error = \( \sqrt{0.12^2 + 0.12^2} = 0.14 \) and therefore \( P = 0.23 \)). Similarly, the net effect for Month 2 was −0.124 and not statistically significant (\( P = 0.38 \)). Therefore, after accounting for the confounding effects, there was no ‘strong evidence’ for the effects of exposure in Months 2 and 3 on gestational age according to the statistical significance criterion used by the authors. The net effects of Months 2 and 3 for female births were also not statistically significant. Interestingly, the authors performed the comparison correctly for the variable ‘Exposed m9’ when the response variable was SSR.

An alternative way to model the data would be to pool the data from both regions and use an indicator variable to specify the region. The net effects would then correspond to the coefficients for the region by month interaction terms. Furthermore, sampling from the unaffected region, as the authors did, simply discards useful data. There is no statistical reason to match the group sizes in this study.

Evidence is accumulating that stress influences both gestational age and SSR, though the exact mechanisms responsible are not clearly understood. This paper is valuable in reporting on yet another useful data. There is no statistical reason to match the group sizes in this study.

We would like to raise some questions about the statistical methods used by Torche and Kleinhaus (2012) in their valuable study of the effects of the Chile earthquake on gestational age and secondary sex ratio (SSR). They used significance tests to argue that they observed important effects. However, the conclusions would be different if the significance tests were performed correctly. We demonstrate the difference for the case of gestational age, but similar results follow for their analysis of effects on other response variables.

To assess the effect of the timing of earthquake exposure, Torche and Kleinhaus included nine indicator variables to indicate exposure in each of the 9 months of gestation and used an unexposed group as the reference category. The coefficients for the indicator variables thus reflected the mean difference in gestation duration between the unexposed baseline group and those exposed in that month.

In order to rule out the influence of any temporal trend affecting both earthquake and non-earthquake regions, they ran the regressions separately for the affected and unaffected regions. They pointed out this amounted to a ‘difference-in-difference’ approach. Using this approach, the net effect of earthquake exposure should be the difference between each coefficient in the affected region and the corresponding coefficient in the unaffected region. It is possible to construct a confidence interval or assess the statistical significance of this effect based on the fact that the standard error of the net effect is equal to the sum of standard errors of the two coefficients. However, the authors did not do this.

They claimed ‘if we find significant differences in the outcomes of interest between the exposed and unexposed period in the affected region, but no differences in the unaffected region, this will provide strong evidence for an effect of earthquake exposure on birth outcomes that cannot be attributed to confounding factors’ (Torche and Kleinhaus, 2012, p. 560). The ‘effect that cannot be attributed to confounding factors’ should be the net effect, which is the difference between the two coefficients. However, the interpretation ‘if X is statistically significant but Y is not, then the difference between X and Y is statistically significant’ is not necessarily true. Gelman and Stern (2006) give several examples in which this reasoning fails to hold. It fails in the same way in Torche and Kleinhaus (2012).

In Table II of their paper, the coefficient for the variable ‘Exposed m3’ was statistically significant in the affected region (coefficient = −0.271, \( P = 0.005 \)), but not in the unaffected region (coefficient = −0.104, \( P = 0.306 \)). A similar pattern held for the variable ‘Exposed m2’. The authors claimed that ‘earthquake exposure in Month 2 and 3 of exposure resulted in a significant decline in gestational age’ (Torche and Kleinhaus, p. 558). However, the net effect for Month 3 was −0.167 and, using the standard error calculation above, not statistically significant (standard error = \( \sqrt{0.12^2 + 0.12^2} = 0.14 \) and therefore \( P = 0.23 \)). Similarly, the net effect for Month 2 was −0.124 and not statistically significant (\( P = 0.38 \)). Therefore, after accounting for the confounding effects, there was no ‘strong evidence’ for the effects of exposure in Months 2 and 3 on gestational age according to the statistical significance criterion used by the authors. The net effects of Months 2 and 3 for female births were also not statistically significant. Interestingly, the authors performed the comparison correctly for the variable ‘Exposed m9’ when the response variable was SSR.

An alternative way to model the data would be to pool the data from both regions and use an indicator variable to specify the region. The net effects would then correspond to the coefficients for the region by month interaction terms. Furthermore, sampling from the unaffected region, as the authors did, simply discards useful data. There is no statistical reason to match the group sizes in this study.

Evidence is accumulating that stress influences both gestational age and SSR, though the exact mechanisms responsible are not clearly understood. This paper is valuable in reporting on yet another useful data. There is no statistical reason to match the group sizes in this study.

We would like to raise some questions about the statistical methods used by Torche and Kleinhaus (2012) in their valuable study of the effects of the Chile earthquake on gestational age and secondary sex ratio (SSR). They used significance tests to argue that they observed important effects. However, the conclusions would be different if the significance tests were performed correctly. We demonstrate the difference for the case of gestational age, but similar results follow for their analysis of effects on other response variables.

To assess the effect of the timing of earthquake exposure, Torche and Kleinhaus included nine indicator variables to indicate exposure in each of the 9 months of gestation and used an unexposed group as the reference category. The coefficients for the indicator variables thus reflected the mean difference in gestation duration between the unexposed baseline group and those exposed in that month.

In order to rule out the influence of any temporal trend affecting both earthquake and non-earthquake regions, they ran the regressions separately for the affected and unaffected regions. They pointed out this amounted to a ‘difference-in-difference’ approach. Using this approach, the net effect of earthquake exposure should be the difference between each coefficient in the affected region and the corresponding coefficient in the unaffected region. It is possible to construct a confidence interval or assess the statistical significance of this effect based on the fact that the standard error of the net effect is equal to the sum of standard errors of the two coefficients. However, the authors did not do this.

They claimed ‘if we find significant differences in the outcomes of interest between the exposed and unexposed period in the affected region, but no differences in the unaffected region, this will provide strong evidence for an effect of earthquake exposure on birth outcomes that cannot be attributed to confounding factors’ (Torche and Kleinhaus, 2012, p. 560). The ‘effect that cannot be attributed to confounding factors’ should be the net effect, which is the difference between the two coefficients. However, the interpretation ‘if X is statistically significant but Y is not, then the difference between X and Y is statistically significant’ is not necessarily true. Gelman and Stern (2006) give several examples in which this reasoning fails to hold. It fails in the same way in Torche and Kleinhaus (2012).

In Table II of their paper, the coefficient for the variable ‘Exposed m3’ was statistically significant in the affected region (coefficient = −0.271, \( P = 0.005 \)), but not in the unaffected region (coefficient = −0.104, \( P = 0.306 \)). A similar pattern held for the variable ‘Exposed m2’. The authors claimed that ‘earthquake exposure in Month 2 and 3 of exposure resulted in a significant decline in gestational age’ (Torche and Kleinhaus, p. 558). However, the net effect for Month 3 was −0.167 and, using the standard error calculation above, not statistically significant (standard error = \( \sqrt{0.12^2 + 0.12^2} = 0.14 \) and therefore \( P = 0.23 \)). Similarly, the net effect for Month 2 was −0.124 and not statistically significant (\( P = 0.38 \)). Therefore, after accounting for the confounding effects, there was no ‘strong evidence’ for the effects of exposure in Months 2 and 3 on gestational age according to the statistical significance criterion used by the authors. The net effects of Months 2 and 3 for female births were also not statistically significant. Interestingly, the authors performed the comparison correctly for the variable ‘Exposed m9’ when the response variable was SSR.

An alternative way to model the data would be to pool the data from both regions and use an indicator variable to specify the region. The net effects would then correspond to the coefficients for the region by month interaction terms. Furthermore, sampling from the unaffected region, as the authors did, simply discards useful data. There is no statistical reason to match the group sizes in this study.

Evidence is accumulating that stress influences both gestational age and SSR, though the exact mechanisms responsible are not clearly understood. This paper is valuable in reporting on yet another useful data. There is no statistical reason to match the group sizes in this study.