

## Answers

Statements *a*, *b*, and *c* are all false.

The primary aim of the trial was to investigate the effects of a low glycaemic index diet in pregnancy, when compared to no intervention, on the incidence of macrosomia. Participants were a group of women who were at risk of fetal macrosomia. The secondary outcome was difference in gestational weight gain, measured at 24, 28, 34, and 40 weeks. Because women were followed during pregnancy, the trial had a longitudinal design, and the measurements of weight gain are called repeated measures or serial data. Treatment groups were compared at each time point using the independent samples *t* test, which has been described in a previous question.<sup>(2)</sup> However, this approach presents statistical problems.

The serial observations on gestational weight gain would not have been independent of each other, but would have correlated with one another (*a* is false). If two measurements are independent of each other, knowing the value of one provides no information about the other. This would not have been true for gestational weight gain. A woman's weight gain at—for example, 24 weeks—would be correlated with her weight gain at 28 weeks; that is, her weight gain at 24 weeks would predict her weight gain at 28 weeks.

Treatment groups were compared in gestational weight gain at 24, 28, 34, and 40 weeks using an independent samples *t* test. The purpose of each test was to make inferences about the population based on the sample. However, if the sample was not representative of the population, errors could have occurred when hypothesis testing. Two types of error were possible, type I and II, described in a previous question.<sup>(3)</sup> A type I error would have occurred if the null hypothesis was incorrectly rejected in favour of the alternative—that is, if there was a difference in gestational weight gain between treatment groups in the trial but not in the population. A type II error would have occurred if the null hypothesis was not rejected in favour of the alternative when it should have been—that is, there was a difference between treatments in gestational weight gain in the population but the difference was not seen in the trial. Type I and II errors would have occurred because of sampling error—that is, because only a proportion of the population was studied.

As described in a previous question,<sup>(3)</sup> the probability that a type I error will occur for a single hypothesis test is 0.05; however, when multiple hypothesis tests are performed the probability is increased (*b* is false). Care must be taken when research papers undertake a large number of statistical tests—ultimately some of these may result in type I errors. However, we will not know which significant findings are a type I error. The researchers in this example reported many tests that compared treatments groups, not just for gestational

weight gain at 24, 28, 34, and 40 weeks. Because the measurements of gestational weight gain were not independent of each other, the significance tests performed at each gestational age would not be independent of each other. If one of the tests was significant, then tests at the other time points would probably be significant too simply because the measurements of gestational weight gain were correlated; this would further increase the probability of a type I error.

The fact that the measurements of gestational weight gain were correlated may have resulted in spurious significant differences between treatment groups at each of the gestational ages. Undertaking four separate statistical tests did not permit a general comparison between treatment groups across pregnancy. Therefore, it cannot be concluded that there was a significant difference between treatments in gestational weight gain across pregnancy ( $c$  is false). A more sophisticated statistical approach, such as a longitudinal analysis incorporating a multilevel model or generalised estimating equations, would have been needed to compare treatment groups in gestational weight gain across pregnancy at the different gestational ages. Such analyses would investigate the change in gestational weight gain in women during pregnancy while accounting for the non-independence of data. Although the researchers in this example constructed a statistical linear model that examined total weight gain and controlled for weight at baseline, they presented the results of statistical testing at each gestational age, which may have been misleading because it ignored important statistical issues described above.

Competing interests: None declared.

- 1 Walsh JM, McGowan CA, Mahony R, Foley ME, McAuliffe FM. Low glycaemic index diet in pregnancy to prevent macrosomia (ROLO study): randomised control trial. *BMJ* 2012;345:e5605.
- 2 Sedgwick P. Independent samples t test. *BMJ* 2012;340:c2673.
- 3 Sedgwick P. Multiple significance tests: the Bonferroni correction. *BMJ* 2012;344:e509.

**Cite this as:** *BMJ* 2013;346:f363