BMJ

BMJ 2012;344:e3971 doi: 10.1136/bmj.e3971 (Published 13 June 2012)

ENDGAMES

STATISTICAL QUESTION

Meta-analyses: tests of heterogeneity

Philip Sedgwick *senior lecturer in medical statistics*

Centre for Medical and Healthcare Education, St George's, University of London, Tooting, London, UK

Researchers investigated the association between consumption of white rice and type 2 diabetes. They performed a meta-analysis of prospective cohort studies that reported the relative risk of type 2 diabetes by intake of white rice (high *v* low). In total, four publications were identified that included seven distinct prospective cohort analyses in Asian and Western populations. Rice intake and type 2 diabetes were identified through self report. A total of 13 284 incident cases of type 2 diabetes were ascertained among 352 384 participants with follow-up periods ranging from four to 22 years.[1]

For each study the researchers identified the relative risk of type 2 diabetes for high consumption of white rice compared with low intake. Statistical tests of heterogeneity were undertaken across the seven sample estimates (Cochran's Q test, P=0.001; I^2=72.2%). The overall relative risk was 1.27 (95% confidence interval 1.04 to 1.54). The researchers concluded that higher consumption of white rice was associated with a significantly higher risk of type 2 diabetes.

Which of the following statements, if any, are true for the statistical test of heterogeneity?

a) Null hypothesis: heterogeneity exists between the sample relative risks as estimates of the population parameter

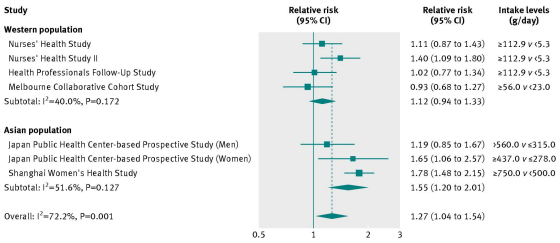
b) Statistical heterogeneity existed between the seven sample estimates of the population relative risk

c) A random effects model was appropriate for the calculation of the overall relative risk

Answers

Statement *b* and *c* are true, whereas *a* is false.

The meta-analysis combined the seven sample estimates for the population parameter of the relative risk of type 2 diabetes for high consumption of white rice compared with low intake. The overall estimate of the population relative risk for type 2 diabetes was more precise than any of the individual sample estimates. The forest plot for the meta-analysis is shown (figure). The overall relative risk of 1.27 (1.04 to 1.54) is displayed at the bottom of the plot against the line "Overall: I^2=72.2%, P=0.001." The forest plot shows how low and high intake of white rice were categorised for each study.



Pooled random effects relative risk (95% confidence interval) of type 2 diabetes comparing high consumption of white rice with low consumption. P values were calculated using Cochran's Q test for heterogeneity

It was essential that the meta-analysis incorporated a statistical test of heterogeneity to assess the extent of variation between the seven sample estimates. Statistical homogeneity would exist if the sample relative risks were similar in size and if variation between them was no more than expected when taking samples from the same population—that is, there was minimal variation between them. If statistical homogeneity did not exist, then statistical heterogeneity was present, and the sample estimates would differ substantially. Variation between sample estimates may occur for a variety of reasons. The population parameter may have differed in size between subgroups—for example, between ethnic groups. The result of the statistical test of heterogeneity influenced how the total overall result was obtained.

The traditional statistical test for heterogeneity is Cochran's Q test. The test is performed in a similar way to traditional statistical hypothesis testing, there being a null hypothesis and an alternative hypothesis. Hypothesis testing starts at the position of statistical homogeneity. For the above meta-analysis, the null hypothesis states that homogeneity exists between the sample estimates of the population parameter (*a* is false); any variation that did exist resulted from differences between studies when sampling from the same population, or possibly minor differences between studies in methodology. The alternative hypothesis states that heterogeneity exists between the sample estimates. The P value associated with the Cochran Q test was

p.sedgwick@sgul.ac.uk

0.001. Therefore, the null hypothesis was rejected in favour of the alternative at the 5% critical level of significance. It was concluded that statistical heterogeneity existed between the sample estimates (*b* is true). The P value for Cochran's Q test is displayed on the forest plot in the line with the title: "Overall: I^2=72.2%, P=0.001."

Cochran's Q test may not always accurately detect heterogeneity in sample estimates. Because of this, Higgins I^2 statistic is often used as well. This statistic represents the percentage of variation between the sample estimates as a result of heterogeneity. It can take values from 0% to 100%, with 0% indicating that statistical heterogeneity does not exist. Significant heterogeneity is typically considered to be present if I^2 is 50% or more. The I^2 for the overall effect is shown on the forest plot in the line with the title: "Overall: I^2=72.2%, P=0.001," and this corroborates the inference of the Cochran Q test that statistical heterogeneity existed (*b* is true).

A so called random effects meta-analysis was performed because of the presence of statistical heterogeneity (*c* is true). If statistical heterogeneity had not existed (that is, if statistical homogeneity had existed), a fixed effects meta-analysis would have been undertaken. The difference between these approaches is the methodology used to calculate the total overall effect. In the presence of heterogeneity, a random effects meta-analysis produces a wider confidence interval for the total overall effect than a fixed effects meta-analysis, resulting in a less accurate overall effect size.

Although random effects methodology accounted for the presence of heterogeneity when calculating the overall estimate, it is questionable whether the seven sample estimates should have been combined into a single overall effect. The presence of heterogeneity suggested that the association between consumption of white rice and type 2 diabetes may have differed between subgroups in the population. The studies were split into two subgroups—Asian and Western populations—with the aim of establishing whether homogeneity existed between the sample estimates within these subgroups (figure). A difference between Western and Asian populations in the association between consumption of white rice and risk of type 2 diabetes was reported. The association for Asian populations (relative risk 1.55, 1.20 to 2.01) was stronger than for Western populations (1.12, 0.94 to 1.33). The researchers concluded that heterogeneity of sample estimates did not exist for either subgroup. Cochran's Q test was not statistically significant for either stratum (0.127 and 0.172, respectively). For Western populations, Higgin's I^2 statistic was 40.0%, whereas for Asian populations it was indicative of weak heterogeneity (I^2=51.6%). Although statistical homogeneity was a reasonable assumption for both subgroups, a random effects model was performed. In the presence of homogeneity, a random effects model produced the same subtotal estimates and 95% confidence intervals as a fixed effects model.

Competing interests: None declared.

1 Hu EA, Pan A Malik V, Sun Q. White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. *BMJ* 2012;344:e1454.

Cite this as: *BMJ* 2012;344:e3971