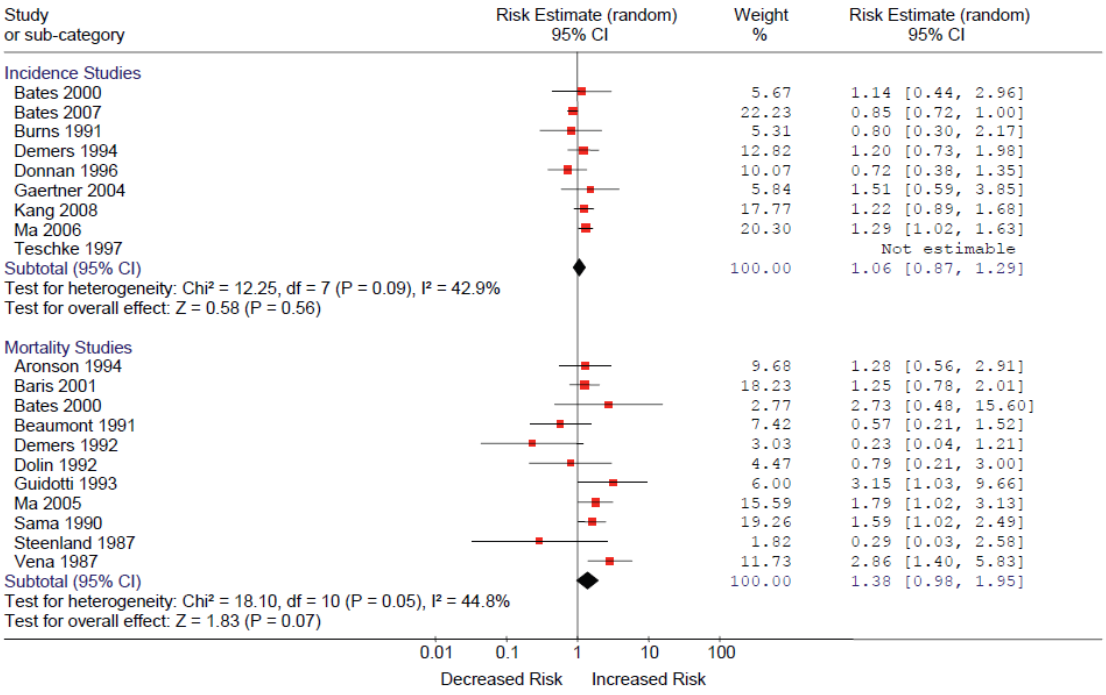
APPENDIX C – Forest Plots for different Cancer sites included in the Meta-Analysis

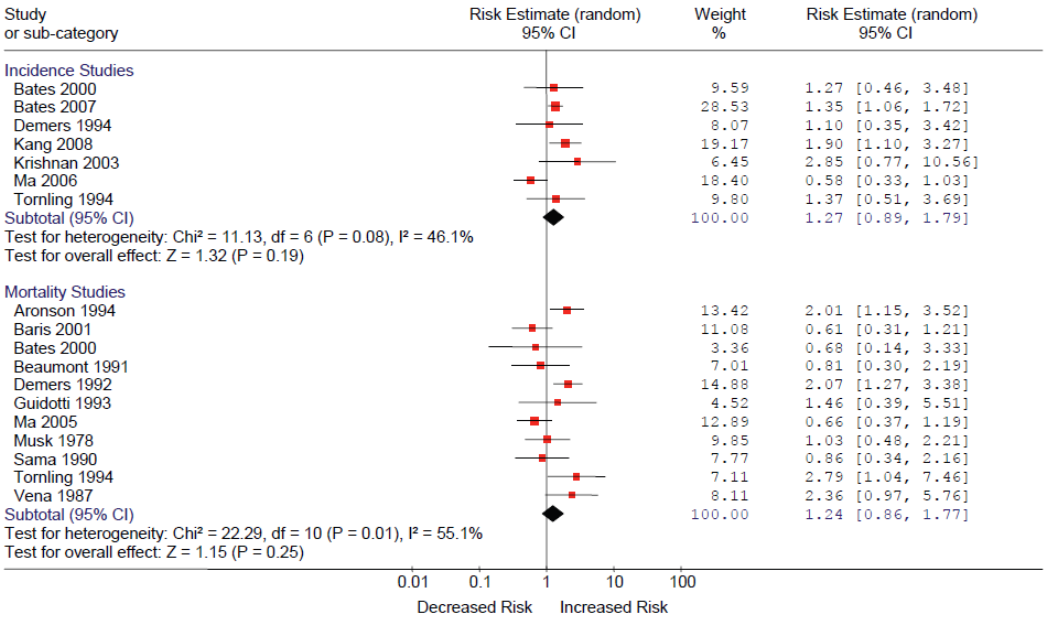
Bladder Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

For bladder cancer studies, the p-value for the Q test for heterogeneity was 0.005 for all good and adequate studies, indicating that there was significant heterogeneity between studies. Heterogeneity was noticeable in the forest plot as many of the confidence intervals did not overlap. I^2 for good and adequate studies showed 50.8% of moderate heterogeneity and mortality studies had more inconsistency than incidence studies (44.8% vs. 42.9%). The findings suggest that 50.8% of the variability across good and adequate studies for bladder cancer is due to heterogeneity rather than chance. Some of the observed heterogeneity might be explained by different study design, variations in study population as well as different methods to verify bladder cancers.

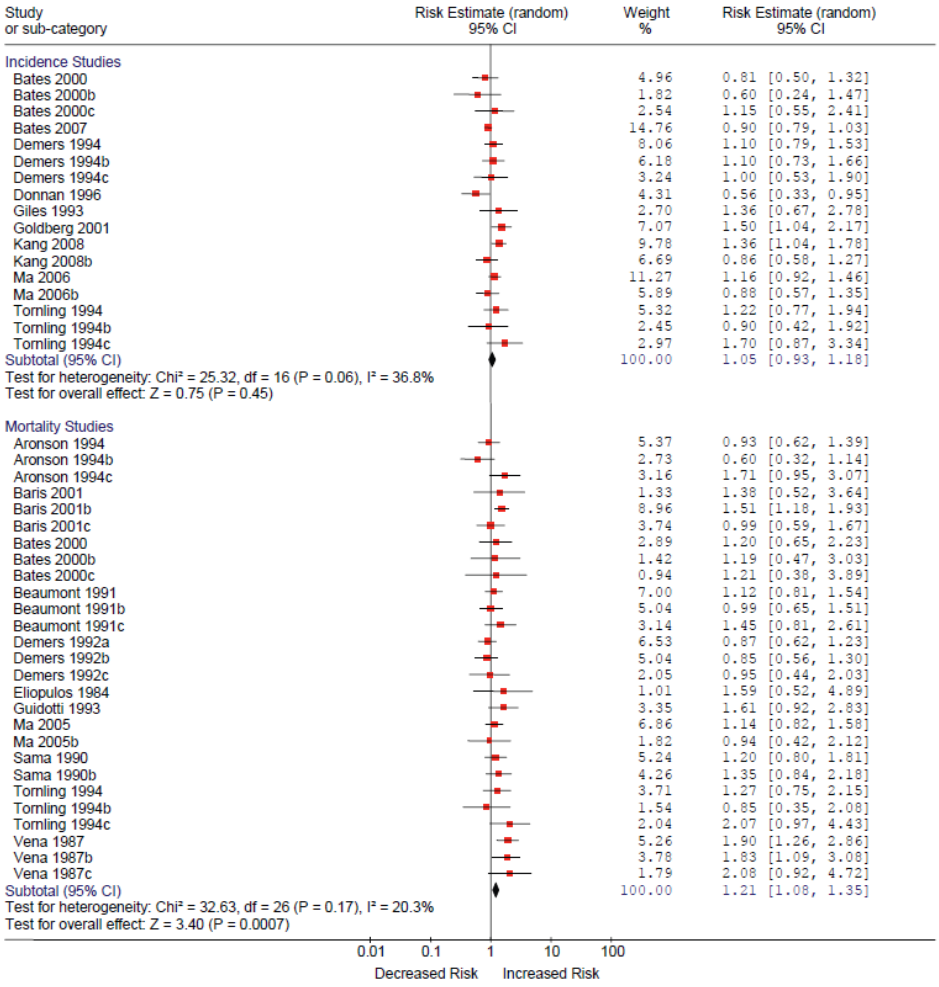
Brain and Central Nervous System Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was marginally significant heterogeneity for all good and adequate studies. The Cochran's Q test had a p-value of 0.10 and the corresponding I^2 statistics was 49.3%, both indicating moderate inconsistency between studies. Mortality studies showed more variability between studies than incidence studies (55.1% vs. 46.1%). Total variation across studies is interpreted as 55% for mortality studies and 46% for incidence studies for brain cancer ($p < 0.10$ for both incidence and mortality). This noticeable inconsistency in brain cancer primarily resulted from the studies of Ma (2006), Ma (2005), and Baris (2001), which were also identified as outliers in the Galbraith plot.

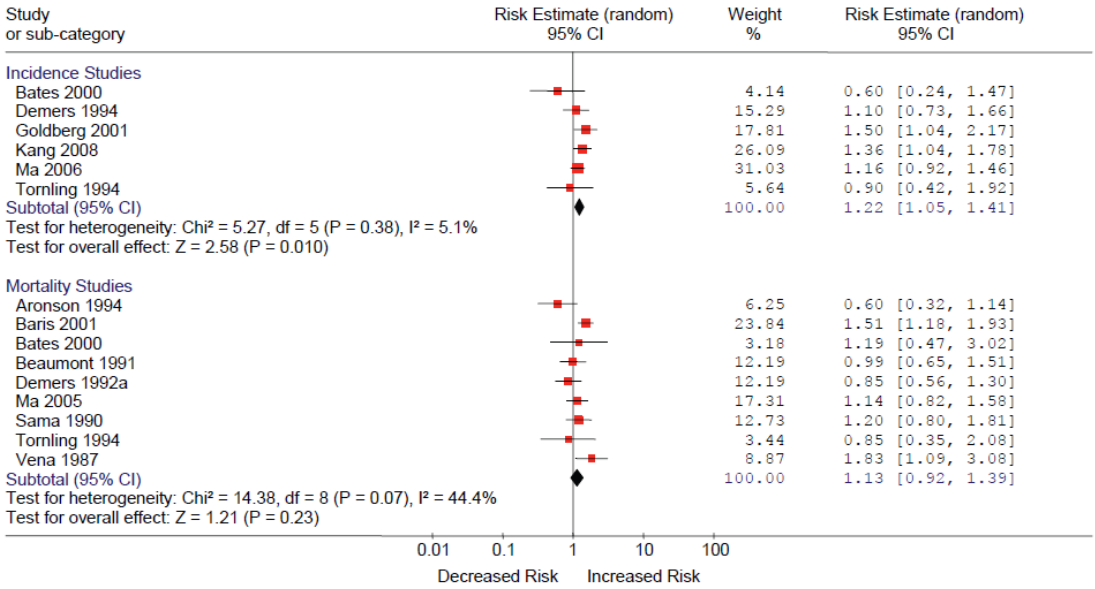
Colorectal Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

The Q test value for heterogeneity for all colorectal cancer studies had a p-value of 0.01 and the corresponding I_2 statistics was 34.7%, indicating moderate variability between studies. Incidence studies showed more inconsistency (p = 0.06), which was significant and a moderate degree of inconsistency across studies (I^2 = 36.8%). Although the risk estimates of mortality studies appeared to vary across studies, there was no significant evidence of heterogeneity (I^2 = 20.3%; p = 0.17) indicating little variability between studies that can not be explained by chance.

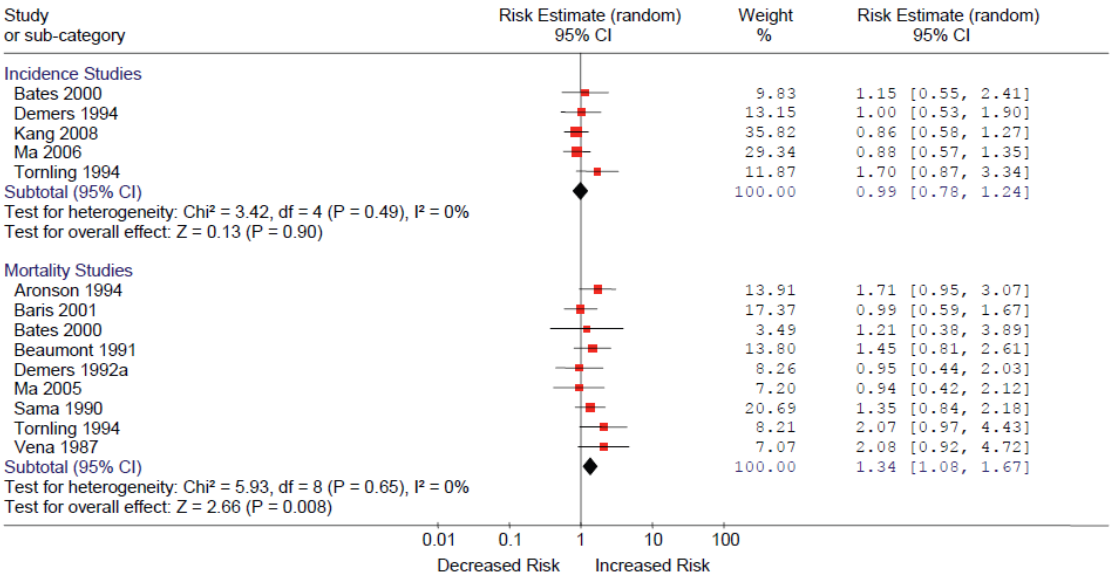
Colon Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

No considerable heterogeneity was found across all studies of good or adequate quality (I^2 = 29%; p = 0.14) or studies for incidence risk (I^2 = 5.1%; p = 0.38). However, mortality studies showed a p-value of 0.07 with the corresponding I_2 statistics being 44.4%, both suggesting moderate inconsistency between studies. The source of heterogeneity in colon cancer mortality may be attributed to the diverse study design, the different outcome measures, and potential confounders. Particularly, a high probability of misclassification between colon and rectal cancers on death certificates could have some effect on the heterogeneity across mortality studies. Aronson's study (1994) was identified as an outlier in the Galbraith plot, shown by the points outside the expected 95% confidence interval lines. It reported the lowest SMR for colon cancer at 0.60 (95% CI: 0.30, 1.08) based on 11 cases from death records.

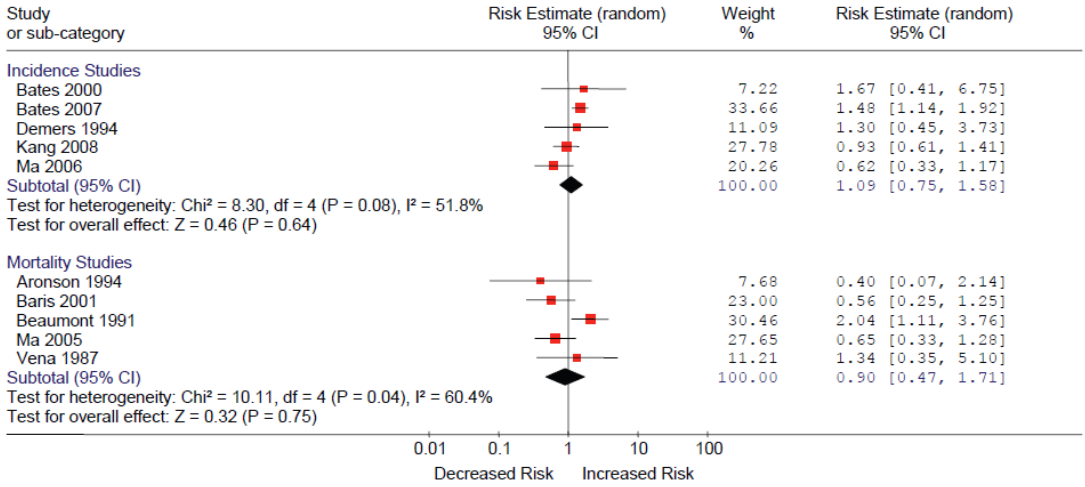
Rectal Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

Although the risk estimates appear to vary across the studies, there was no statistically significant evidence of heterogeneity (p = 0.45) for all studies with combined outcome. The corresponding I_2 statistics for all studies was 0%, also supporting no variability between studies. Neither incidence nor mortality studies showed any evidence of inconsistency among studies. All studies fell within confidence bounds in the Galbraith plot.

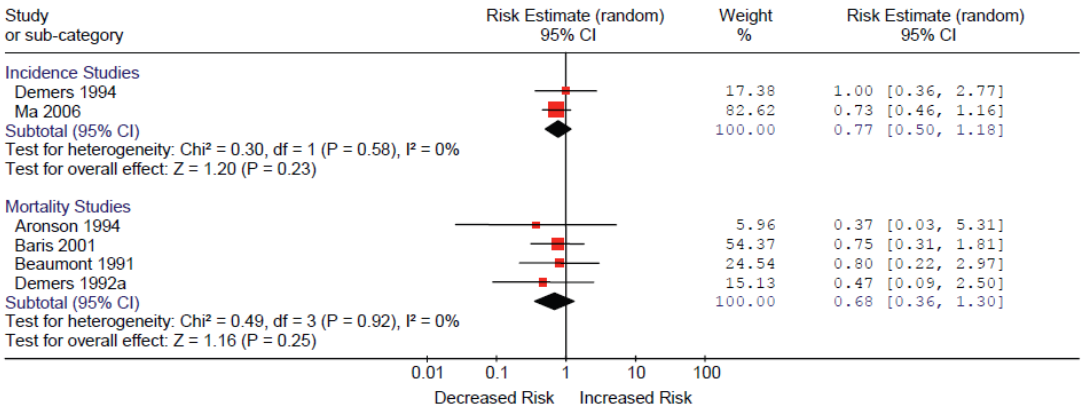
Esophageal Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was significant heterogeneity for all good or adequate studies. The Cochran's Q test had a p-value of 0.0.2 and the corresponding I^2 statistics was 54%, both indicating moderate inconsistency between studies. The degree of heterogeneity was similar for incidence and mortality studies showing 51.8% and 60.4%, respectively, of I^2 statistics ($p < 0.10$ for both incidence and mortality). It means that over half of the total variation across studies for esophageal cancer is not due to chance. Usual sources of heterogeneity could be the different study design, the small number of cases, the reference population, or the control of confounding. Potential confounders might have contributed to between study variations in esophageal cancer, given that most selected studies failed to adjust for smoking or diet. However, influence analysis did not show significant change of risk estimates, indicating that no individual study had a significant effect on the pooled risk estimate by itself.

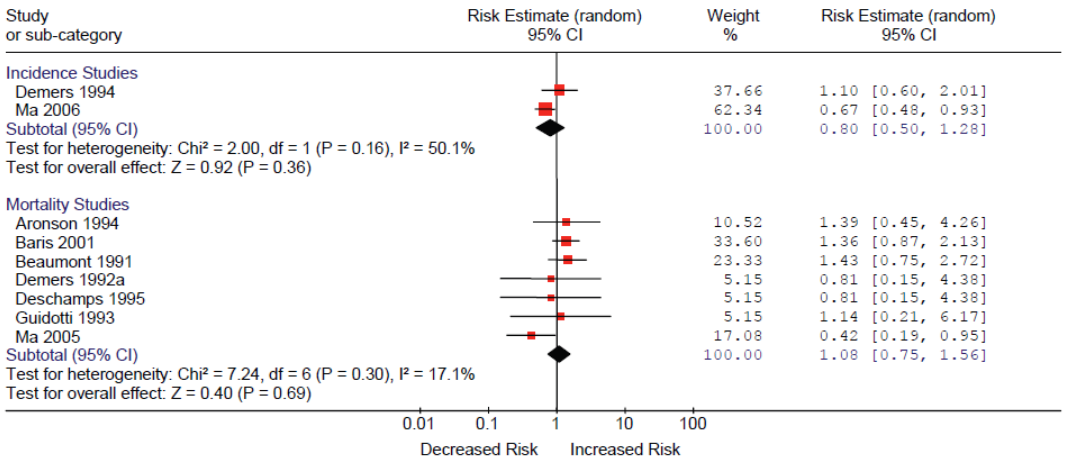
Laryngeal Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was no evidence of significant heterogeneity for all good or adequate studies. The Cochran's Q test had a p-value of 0.97 and the corresponding I^2 statistics was 0%, both indicating no inconsistency between studies. Heterogeneity statistics were not calculated for incidence studies because there were only two studies. Mortality studies did not show the evidence of inconsistency between studies (I^2 =0%, p=0.92).

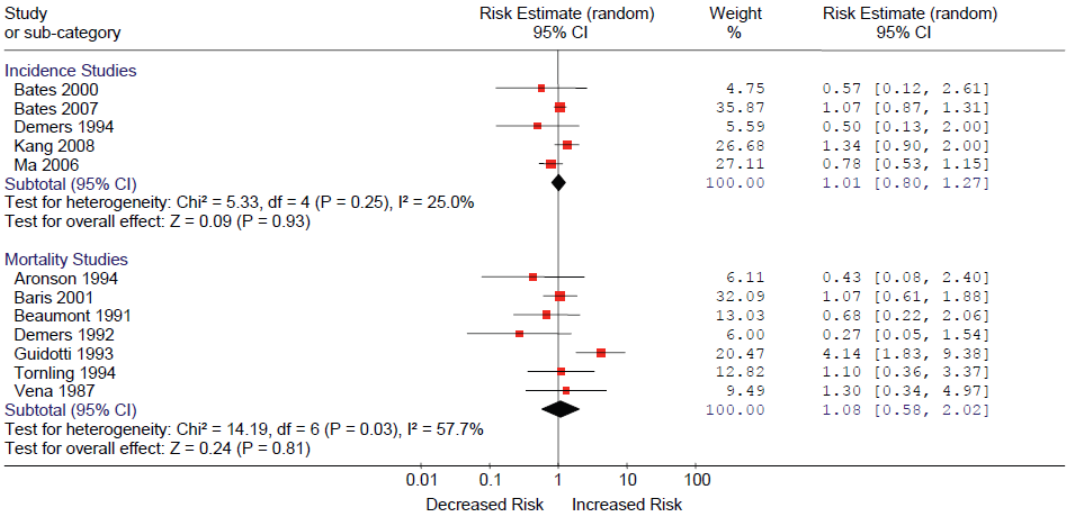
Oral and Pharyngeal Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was a low to moderate level of heterogeneity for all good or adequate studies, although it was not statistically significant. The Cochran's Q test had a p-value of 0.12 and the corresponding I^2 statistics was 37%, indicating that 37% of total variation across studies for oral & pharyngeal cancer was not due to chance (inconsistency between studies). Heterogeneity statistics were not calculated for incidence studies because there were only two studies. Inconsistency among mortality studies was not observed (I^2 = 17%, p = 0.30). Eliminating weak studies did not appreciably change the mortality summary risk estimate.

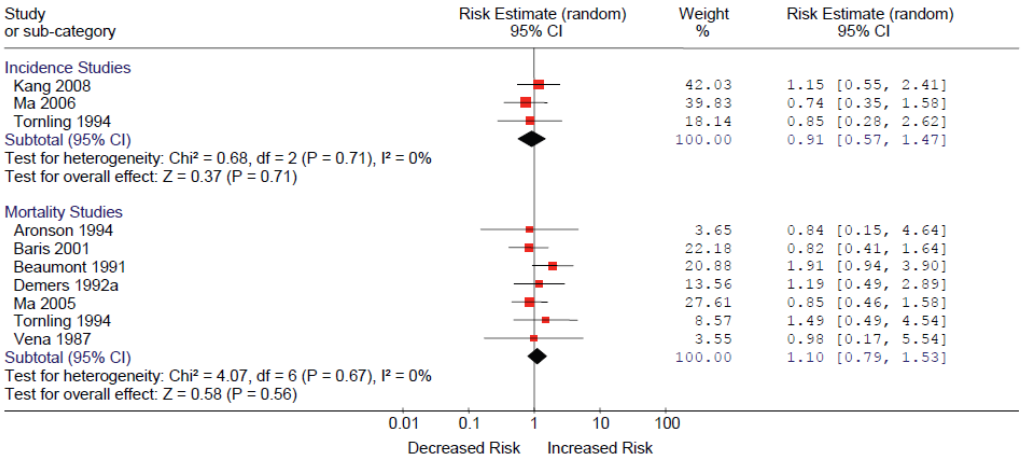
Kidney Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was a noticeable inconsistency among all good or adequate studies. The Cochran's Q test had a p-value of 0.04 and the corresponding I^2 statistics was 46%, both indicating moderate inconsistency between studies. Mortality studies showed more variability between studies than incidence studies (58% vs. 25%). This indicates that 58% of total variation across studies for kidney cancer is not due to chance ($p < 0.10$ for mortality studies). The sensitivity analysis by omitting weak quality studies changed the summary incidence risk estimate from 1.25 (95% CI: 1.02, 1.53) to 1.05 (95% CI: 0.79, 1.38) for combined outcomes.

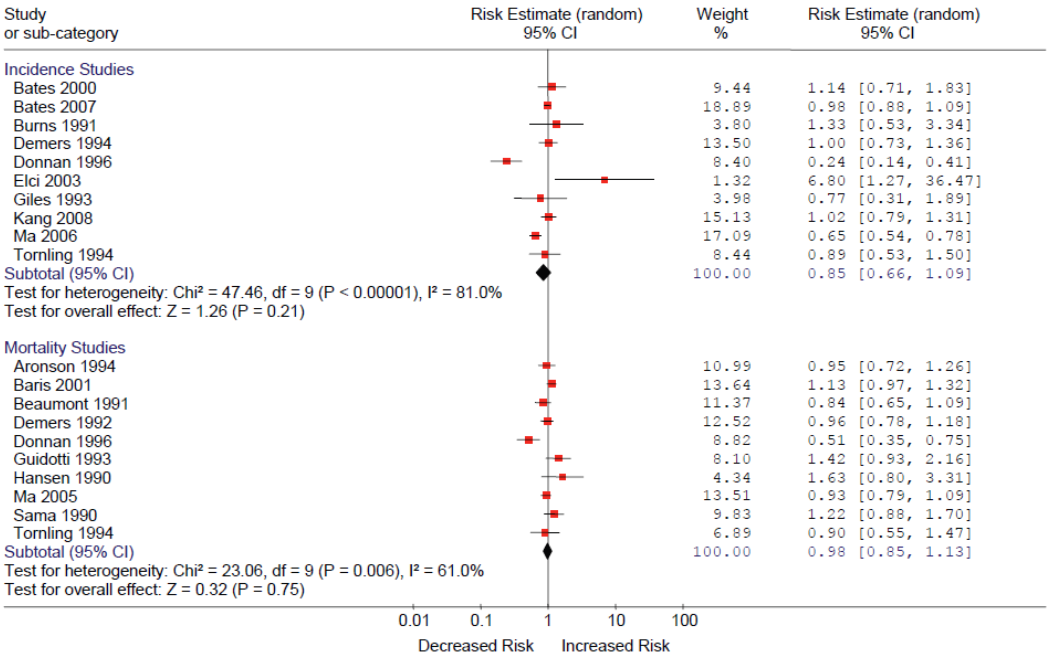
Liver and Gallbladder Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was no evidence of significant heterogeneity across studies either in incidence or mortality studies. The Cochran's Q test was equal to 0.82 with a p value of 0.62, and the corresponding I^2 statistic was 0%. All studies fell into within the confidence bound in the Galbraith plot, which can be translated that there is no major contribution to heterogeneity. Sensitivity analysis by study quality did not show qualitative change of the summary risk estimate.

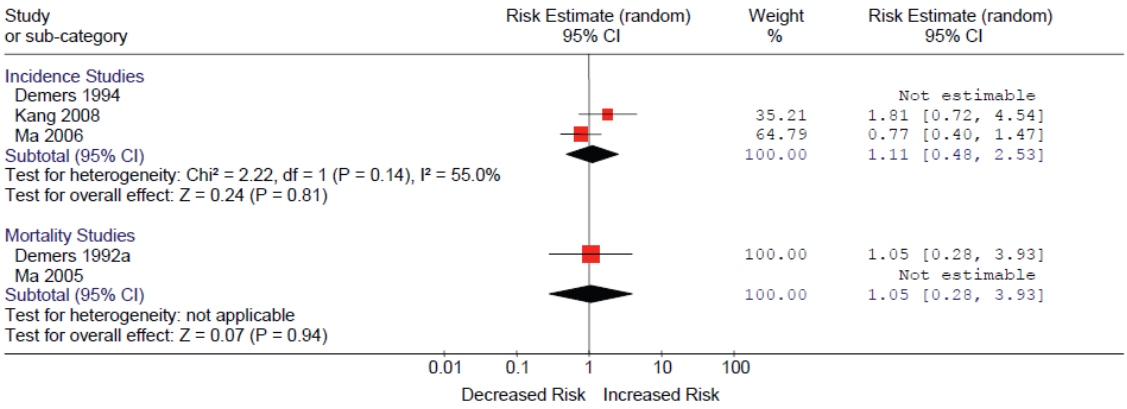
Lung Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was a large degree of heterogeneity observed among good and adequate studies. The Cochran's Q test had a p value of 0.001 and the corresponding I^2 statistics was 74%, both indicating large inconsistency between studies. Incidence studies showed more variability between studies than mortality studies (81% vs. 61%). The above indicate that 81% of total variation across studies for lung cancer incidence is not due to chance ($p < 0.10$ for both incidence and mortality). Donnan (1996), Elci (2003), and Ma (2006) were identified as outliers in the Galbraith plot, as shown by the points outside the expected 95% confidence interval lines. Potential explanations for this heterogeneity are the inclusion of various case-control studies, regional differences, possible misclassification of respiratory cancer code, or lack of adjustments for confounding. Sensitivity analyses by omitting a study at a time to examine the influence of individual studies did not appreciably change the summary risk estimate.

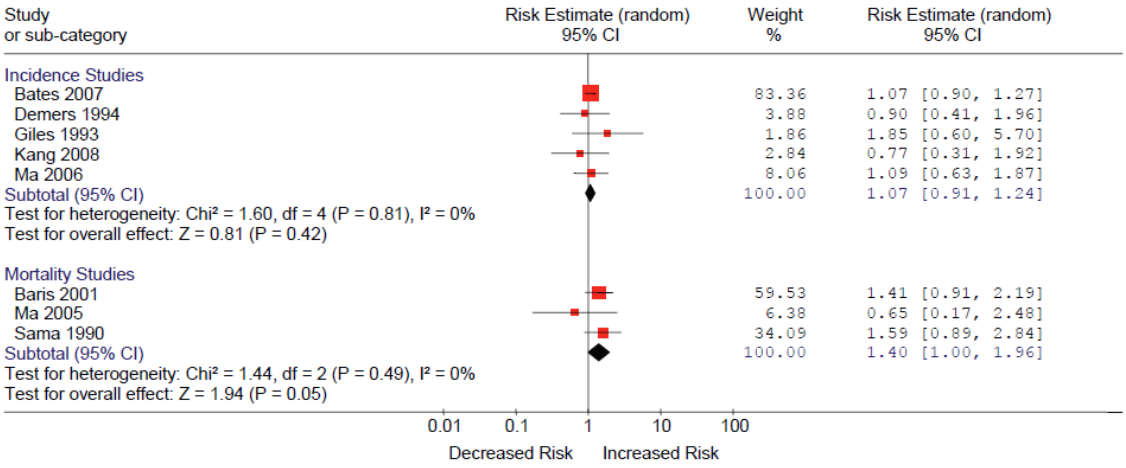
Hodgkin's Lymphoma



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software. Note: Studies which reported a zero confidence interval (Demers 1994; Ma 2005) were not estimable with respect to obtaining a summary risk estimate using the Review manger software; thereby the estimate shown in the figure has a discrepancy compared to the risk estimate calculated by STATA.

There was no significant heterogeneity for all good or adequate studies ($I^2 = 0$; $p = 0.57$) or among all mortality studies regardless of study quality ($I^2 = 0$; $p = 0.75$). However, the number of good or adequate studies was not enough to test heterogeneity.

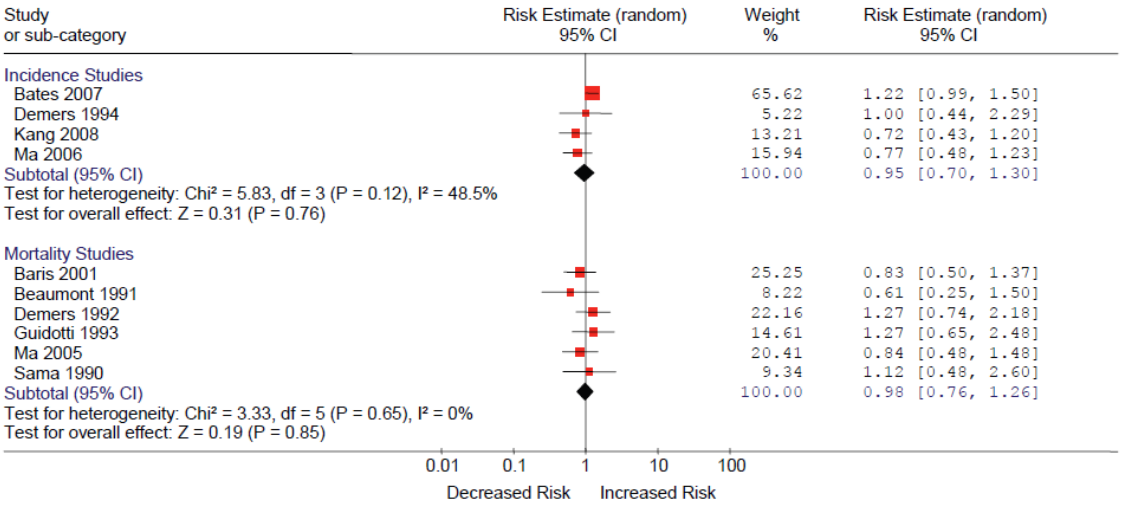
Non-Hodgkin's Lymphoma



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was little evidence of significant heterogeneity for all, good or adequate studies. The Cochran's Q test had a p-value of 0.39 and the corresponding I^2 statistics was 5%, both indicating no inconsistency between studies. Both incidence and mortality studies also were found to have no variability between studies (I^2 = 0% for both incidence and mortality studies). Sensitivity analysis to examine the influence of each study showed that no one study had influence on substantial change of the summary risk estimates.

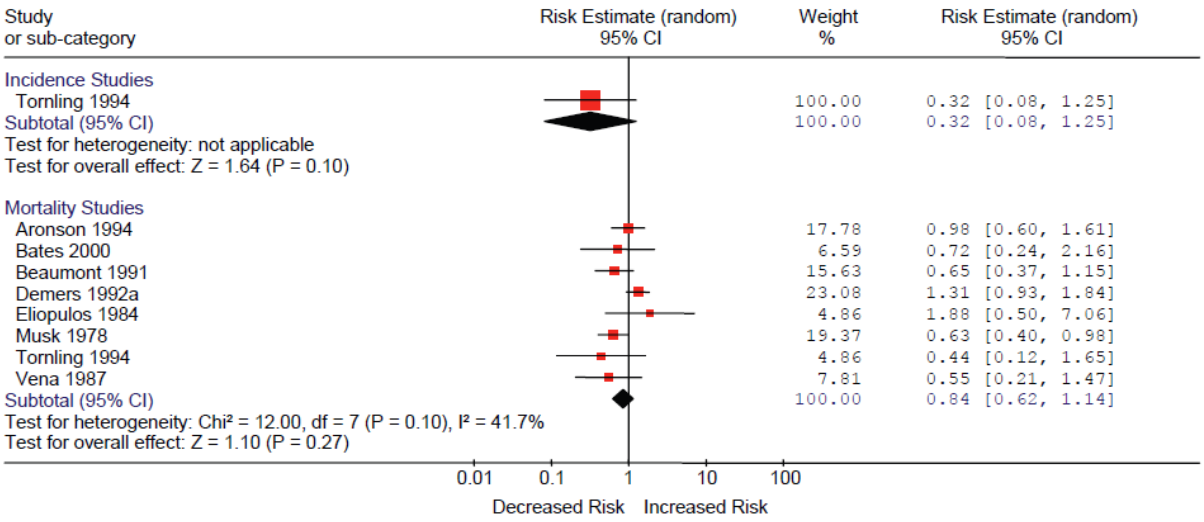
Leukemia



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was little evidence of significant heterogeneity for good or adequate studies. The Cochran's Q test for all studies had a p-value of 0.39 and the corresponding I^2 statistics was 5%, both indicating no inconsistency between studies. Mortality studies also were found to have no variability between studies (I^2 = 0%, p = 0.65). Incidence studies showed a moderate level of variability between studies, suggesting that 49% of total variation across studies for leukemia incidence risk is not due to chance, though it was marginally significant (p = 0.12). Potential sources of heterogeneity in incidence study could be the different study designs used, the small number of cases, or the different methods used to verify leukemia. However, when one study at a time was omitted from the sensitivity analysis to examine the influence of each study, no one study was found to have any substantial effect on the summary risk estimates.

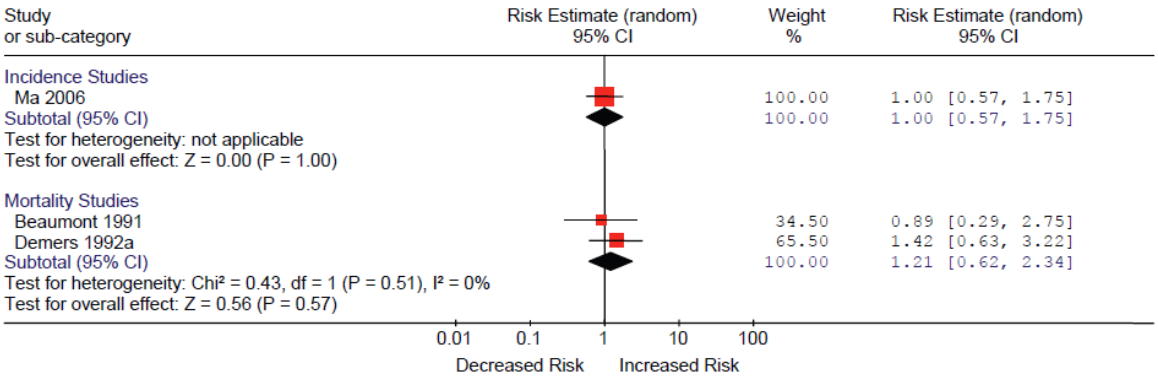
Lymphatic and Hematopoietic Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

Heterogeneity assessment was not available for incidence because there was only one study. Eight mortality studies showed moderate study heterogeneity with marginal significance. I^2 for heterogeneity test showed that 42% of variance was due to heterogeneity between study variance ($p = 0.10$). Demers (1992) was identified as an outlier in the Galbraith plot.

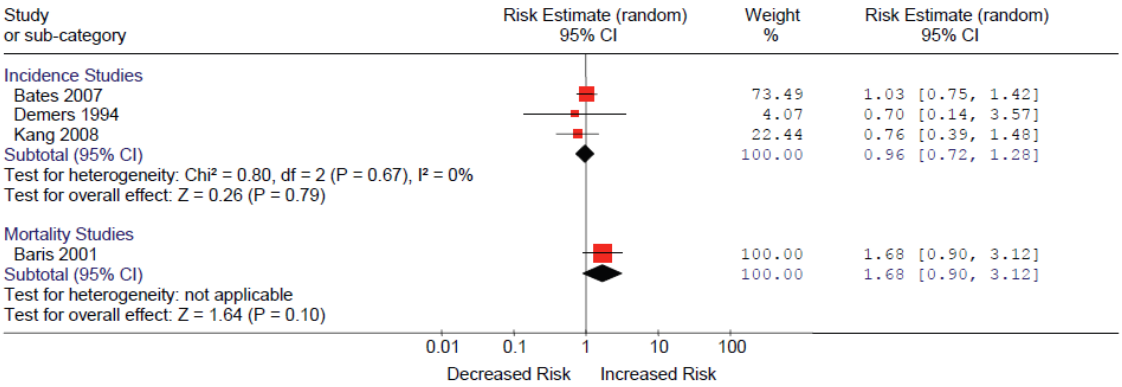
Lymphoma and Reticulosarcoma



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

The number of good or adequate studies on incidence did not allow assessing heterogeneity. For all studies, there was no significant heterogeneity across mortality studies ($I^2 = 0$, $p = 0.89$). All studies fell into within the confidence bound in the Galbraith plot without an outlier.

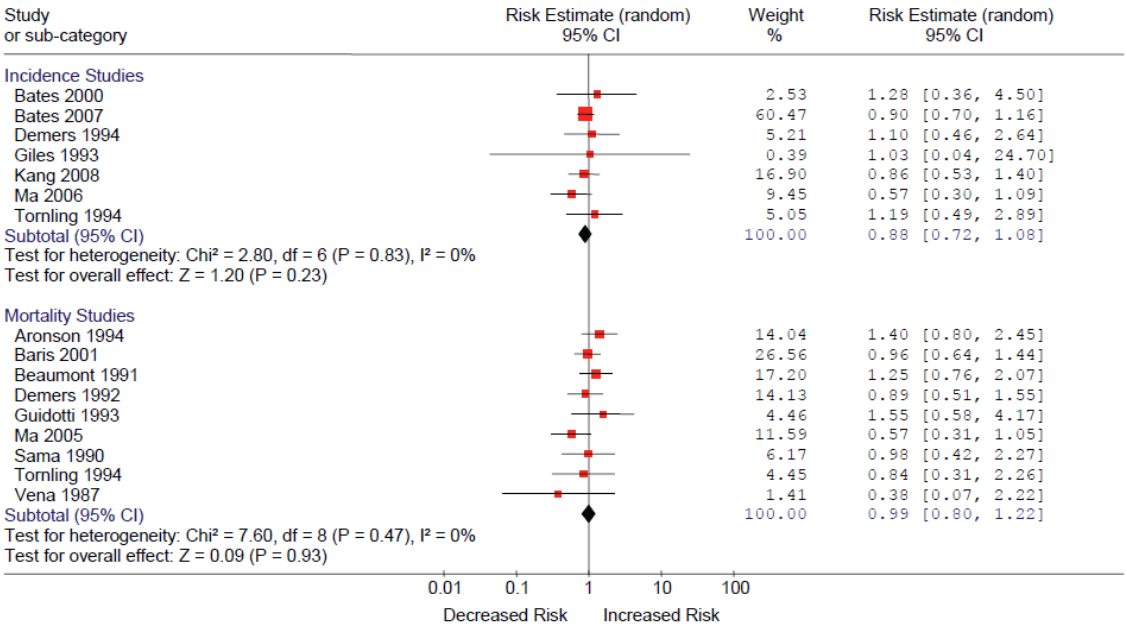
Multiple Myeloma



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

The number of studies in good or adequate studies for multiple myeloma was too few to adequately assess heterogeneity. Q statistic and I_2 measure was tested for all studies, including the ones of poor quality, but there was no evidence of inconsistency across studies ($I^2 = 0$, $p = 0.57$). There was no significant heterogeneity across incidence studies ($I^2 = 0$, $p = 0.67$).

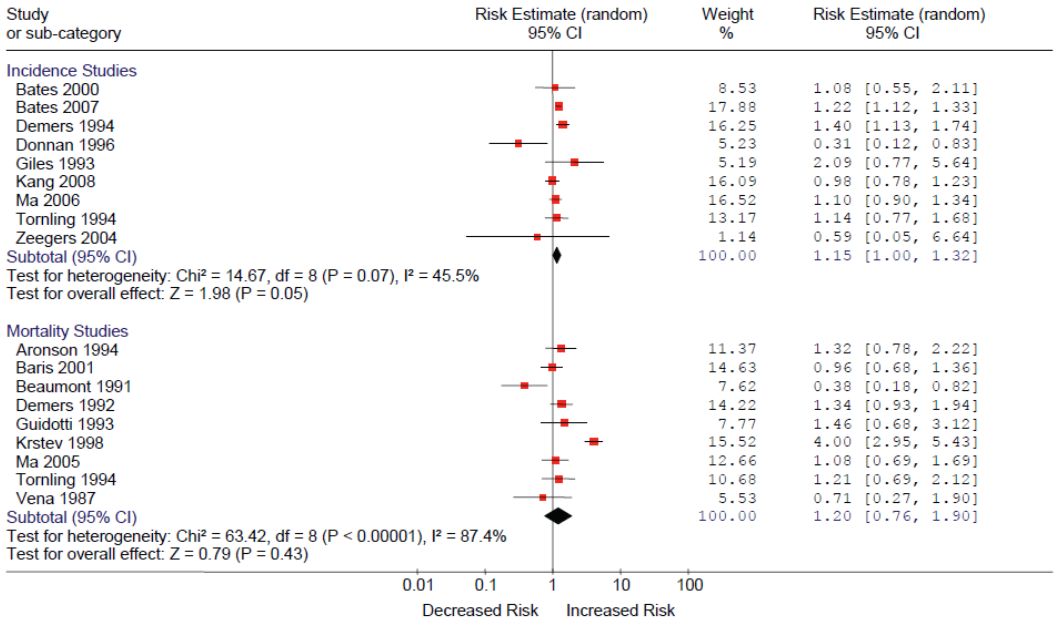
Pancreatic Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was no evidence of significant heterogeneity for all good or adequate studies (I^2=0%, p=0.77). Neither incidence studies nor mortality studies showed any variability between studies (I^2=0%, p=0.83 for incidence; I^2=0%, p=0.47 for mortality). Sensitivity analysis examining the influence of each study did not make a substantial effect on the summary risk estimates.

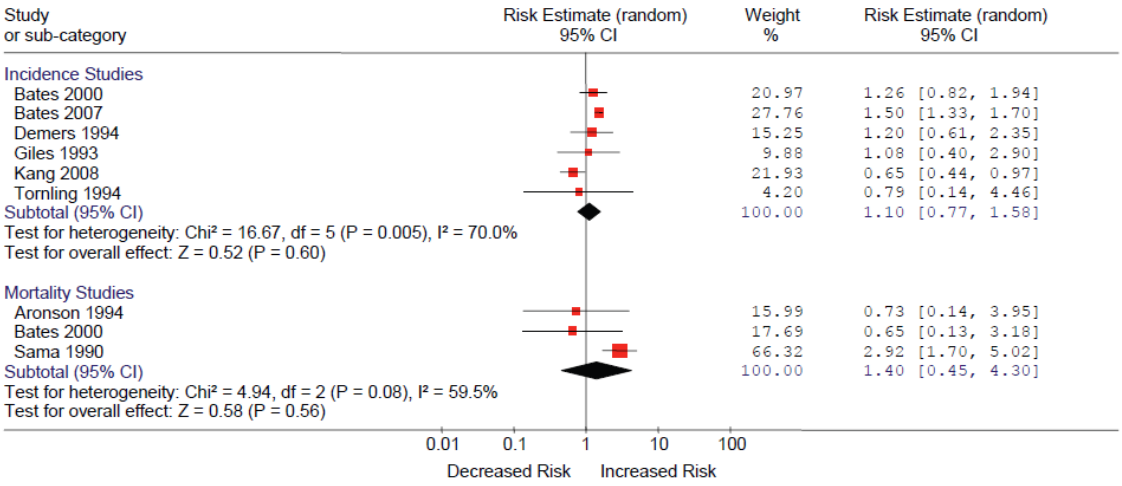
Prostate Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was a high level of heterogeneity for all good or adequate studies. The Cochran's Q test had a p value of 0.001 and the corresponding I^2 statistics was 80%, both indicating moderate inconsistency between studies. Mortality studies showed more variability between studies compared to incidence studies (87% vs. 46%), indicating that 87% of total variation across incidence studies for prostate cancer is not due to chance ($p < 0.10$ for both incidence and mortality). This noticeable inconsistency may be primarily attributed to the results of Donnan (1996), Beaumont (1991) and Krstev (1998). These 3 studies were identified as outliers in the Galbraith plot. Donnan (1998) made a major contribution to heterogeneity (0.31, 95% CI: 0.10, 0.72) for prostate cancer incidence. Krstev (1998) reported OR with a case-control study design based on a small number of observed cases. Beaumont (1991) reported an incidence RR with a weak quality study. Grimes (1991) reported an RR in mortality which was based on fewer than five cases using a surveillance study design. Donnan (1996) and Krstev (1998) were found to have a substantial effect on the change of the summary risk estimates when one study at a time was omitted from the sensitivity analysis.

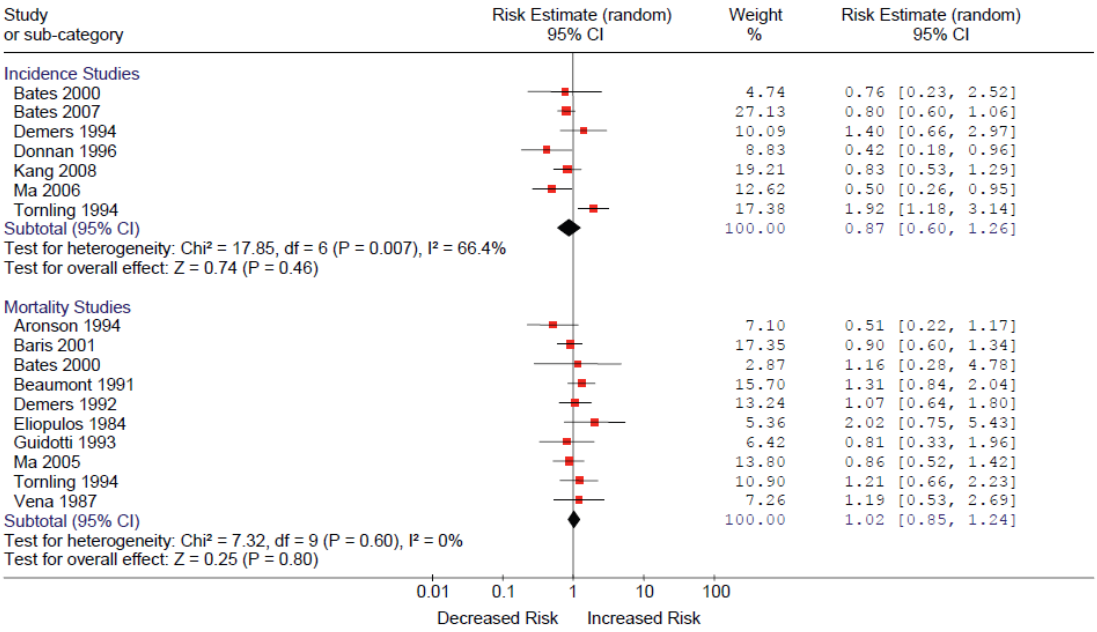
Skin Melanoma



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

Moderate to large heterogeneity was observed for good or adequate studies. For incidence studies, the Cochran's Q test had a p value of 0.005 and the corresponding I^2 statistics was 70.0%, both indicating considerable inconsistency between studies. Mortality studies showed a similar degree of between-study variability (I^2 = 59.5%; p = 0.08). A potential source of this noticeable heterogeneity could be possible misclassification of skin melanoma, different methods of verifying cases, different study designs and populations. Sama et al. (1990) and Kang et al. (2008) were both identified as outliers in the Galbraith plot.

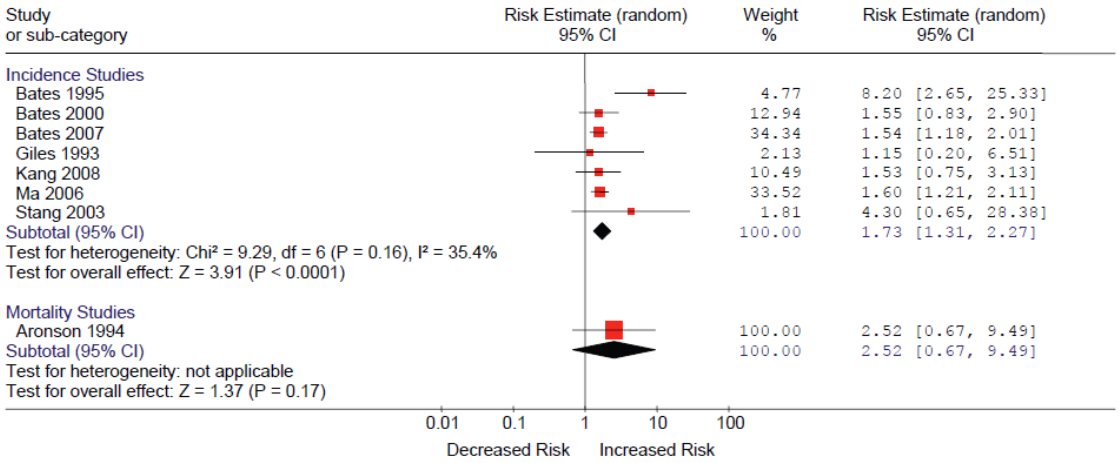
Stomach Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was low to moderate heterogeneity for all good or adequate studies. The Cochran's Q test had a p-value of 0.04 and the corresponding I^2 statistics was 39%, both indicating moderate inconsistency between studies. Among incidence studies there was a 66% of total variation across studies for stomach cancer not due to chance (p = 0.007). Tornning et al. (1994) was identified as an outlier in the Galbraith plot, as shown by a point outside the expected 95% confidence interval lines. Mortality studies did not show any evidence of heterogeneity. Sensitivity analysis by study quality did not change substantially the summary risk estimates.

Testicular Cancer



* Figure is presented only for good and adequate studies. Risk estimates, confidence intervals and graphics were generated by the Review Manager Software.

There was no evidence of significant heterogeneity for all good or adequate studies. The Cochran's Q test had a p-value of 0.32 and the corresponding I^2 statistics was 13%, both indicating low level of inconsistency between studies. Incidence studies showed similar variability between studies (I^2 = 35%; p = 0.16).