Appendix B. Adapted Newcastle-Ottawa Quality Assessment Scales.

1.1 Cross-sectional studies

Selection (maximum 3 points)

1) Representativeness of the sample: 1 point was given if the sample was truly representative of the average in the target population (all subjects or random sampling) or somewhat representative (non-random sampling);

2) Sample size: 1 point was given if sample size was justified and satisfactory;

3) Non-included subjects: 1 point was given if comparability between included and non-included subjects was established, and if the inclusion rate was satisfactory.

Comparability (maximum 2 points)

Subjects in different outcome groups are comparable, based on study design or analysis. Confounding factors are controlled.

1) One point was given if there was adequate adjustment of anthropometric measures for age and gender, and adequate adjustment of endocrine disrupting chemical urinary levels for urinary creatinine levels (or of endocrine disrupting serum levels for serum lipid levels).

2) One additional point was given if there was adequate adjustment for other factors influencing measures of body weight or adiposity, such as race/ethnicity, poverty-to-income ratio, caloric intake, educational level and exercise habits.

Outcome (maximum 3 points)

1) Assessment of the outcome: 1 points were given if the measures of body weight (body mass index, waist circumference) or body fat were assessed independently and blindly or from record linkage (database records), and 1 point was given if the outcome was assessed by self-report;

2) Statistical test: 1 point was given if the statistical test used to analyze the data was clearly described and appropriate, and the measurement of the association was presented, including confidence intervals and the probability level (p value).

Studies that scored a total of 8 or 7 points was considered to have a low risk of bias; 6 points were considered to have a medium risk of bias; 5 points or less were considered to have a high risk of bias. With respect to selection, studies were considered to have a low, medium or high risk of bias if they scored 3, 1-2 or 0 point, respectively. With respect to comparability, studies were considered to have a low, medium or high risk of bias if they scored 2, 1 or 0 point, respectively. With respect to outcome, studies were considered to have a low, medium or high risk of bias if they scored 3, 2 or 1 point, respectively.
1.2 Cohort studies

Selection (maximum 4 points)
1) Representative of the exposed cohort: 1 point was given if the participants were truly or somewhat representative of the average in the community;

2) Selection of the non-exposed or less-exposed cohort: 1 point was given if the non-or less-exposed participants were selected from the same community as the exposed cohort;

3) Ascertainment of exposure: 1 point was given if the methods to assess exposure to endocrine disrupting chemicals were clearly described;

4) Demonstration that the outcome of interest was not present at the start of study: 1 point was given if baseline measures of body weight or adiposity were explicitly described for all patients.

Comparability (maximum 2 points):

Subjects in different outcome groups are comparable, based on study design or analysis. Confounding factors are controlled.

1) One point was given if there was adequate adjustment of anthropometric measures for age and gender, and adequate adjustment of endocrine disrupting chemical urinary levels for urinary creatinine levels (or of endocrine disrupting serum levels for serum lipid levels).

2) One additional point was given if there was adequate adjustment for other factors influencing measures of body weight or adiposity, such as race/ethnicity, poverty-to-income ratio, caloric intake, educational level and exercise habits.

Outcome (maximum 3 points)

1) Assessment of outcome: one point was given if the measures of body weight (body mass index, waist circumference) or body fat were assessed independently and blindly or from record linkage (database records);

2) Time of follow-up: 1 point was given if the time of follow up was clearly stated and was of 6 months or more);

3) Adequacy of follow up: 1 point was given if complete follow up was reached and all subjects were accounted for, or if subjects lost to follow up were explicitly described and unlikely to introduce bias.

Studies that scored a total of 8 or 9 points were considered to have a low risk of bias; 7 or 6 points were considered to have a medium risk of bias; 5 points or less were considered to have a high risk of bias. With respect to selection, studies were considered to have a low, medium or high risk of bias if they scored 4, 2-3 or 1 point, respectively. With respect to comparability, studies were considered to have a low, medium or high risk of bias if they scored 2, 1 or 0 point, respectively. With respect
to outcome, studies were considered to have a low, medium or high risk of bias if they scored 3, 2 or 1 point, respectively.