Supplemental Table 1 – List of excluded papers with reasons for exclusion

|  |  |  |
| --- | --- | --- |
| **Outcome** | **Reference** | **Reason for exclusion** |
| Anogenital distance | Ortega-Garcia JA, Olano-Soler HA, Martinez-Alvarez A, et al. Breastfeeding duration and anogenital distance in 2-year-old infants. Breastfeed Med. 2016;11:350-355. | No measurement of phthalates |
| Anogenital distance | Swan SH, Main KM, Liu F, et al. Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environ Health Perspect*. 2005;113:1056-1061. | Population overlap with a more recent study (Swan 2008) |
| Anogenital distance | Bustamante-Montes, Hernandez-Valero MA, Flores-Pimental D, et al. Prenatal exposure to phthalates is associated with decreased anogenital distance and penile size in male newborns. J Dev Orig Health Dis. 2013;4(4):300-306. | MEP was measured but detection rate was low (8 of 73 participants) and no analysis was conducted |
| Bone health | Shiue I. Urinary heavy metals, phthalates, phenols, thiocyanate, parabens, pesticides, polyaromatic hydrocrabons but not arsenic or polyfluorinated compounds are associated with adult oral health: USA NHANES, 2011-2012. Environ Sci Pollut Res. 2015;22:15636-15645. | Analysis of self-reported bone loss and phthalates does not appear to have been conducted; results were presented for bone loss and other chemicals |
| Oxidative stress | Park HY, Kim JH, Lim YH, Bae S, Hong YC. Influence of genetic polymorphisms on the association between phthalate exposure and pulmonary function in the elderly. Environ Res. 2013;122:18-24. | Did not measure MEP or MiBP |
| Oxidative stress | Asimakopoulos AG, Xue J, De Carvalho BP, et al. Urinary biomarkers of exposure to 57 xenobiotics and its association with oxidative stress in a population in Jeddah, Saudi Arabia. Environ Res. 2016;150:573-581. | Only appear to have assessed phthalates as a group; unclear if MEP and MiBP were examined separately |
| Oxidative stress | Zhang J, Liu L, Wang X, Huang Q, Tian M, Shen H. Low-level environmental phthalate exposure associates with urine metabolome alteration in a Chinese male cohort. Environ Sci Technol. 2016;50(11):5953-5960. | Authors state that a satisfactory model for MEP was not generated and therefore the metabolite was not analyzed; MiBP was not examined |
| Oxidative stress | Ferguson KK, Chen YH, VanderWeele TJ, McElrath TF, Meeker JD, Mukherjee B. Mediation of the relationship between maternal phthalate exposure and preterm birth by oxidative stress with repeated measurements across pregnancy. Environ Health Perspect. 2017;125:488-494. | Authors do not present results for the association between phthalates and isoprostane, although both were measured |
| Oxidative stress | Kataria A, Levine D, Werteneil S, et al. Exposure to bisphenols and phthalates and association with oxidant stress, insulin resistance, and endothelial dysfunction. Pediatr Res. 2017;81(6):857-864. | Only assessed MEP and MiBP as part of a low molecular weight variable |
| Oxidative stress | Choi YJ, Ha KH, Kim DJ. Exposure to bisphenol A is directly associated with inflammation in healthy Korean adults. Environ Sci Pollut Res Int. 2017;24(1):284-290. | Did not measure MEP or MiBP |
| Oxidative stress | Bai PY, Wittert G, Taylor AW, et al. The association between total phthalate concentration and non-communicable diseases and chronic inflammation in South Australian urban dwelling men. Environ Res. 2017;158:366-372. | Did not measure separate metabolites; only a single phthalate variable used |

Supplemental Table 2 – Downs and Black (1998) Quality Assessment Tool

| No. | Original question text | Point allocation | Applicable study designs | Modifications |
| --- | --- | --- | --- | --- |
| 1 | Is the hypothesis/objective of the study clearly described? | Yes = 1 point, No = 0 points | All | None |
| 2 | Are the main outcomes to be measured clearly described in the Introduction or Methods section? | Yes = 1 point, No = 0 points | All | None |
| 3 | Are the characteristics of the patients included in the study clearly described? | Yes = 1 point, No = 0 points | All | None |
| 4 | Are the interventions of interest clearly described? | Yes = 1 point, No = 0 points | None/RCT only | None |
| 5 | Are the distributions of principal confounders in each group of subjects to be compared clearly described? | Yes = 2 points, Partially = 1 point, No = 0 points | All | None |
| 6 | Are the main findings of the study clearly described? | Yes = 1 point, No = 0 points | All | None |
| 7 | Does the study provide estimates of the random variability in the data for the main outcomes? | Yes = 1 point, No = 0 points | All | None |
| 8 | Have all important adverse events that may be a consequence of the intervention been reported? | Yes = 1 point, No = 0 points | RCT only | None |
| 9 | Have the characteristics of patients lost to follow-up been described? | Yes = 1 point, No = 0 points | Cohort  | None |
| 10 | Have actual probability values been reported (e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001? | Yes = 1 point, No = 0 points | All | None |
| 11 | Were the subjects asked to participate in the study representative of the entire population from which they were recruited? | Yes = 1 point, No = 0 points, Unable to determine = 0 | All | None |
| 12 | Were the subjects who were prepared to participate representative of the entire population from which they were recruited? | Yes = 1 point, No = 0 points, Unable to determine = 0 | All | None |
| 13 | Were the staff, places, and facilities where the patients were treated representative of the treatment the majority of the patients receive? | Yes = 1 point, No = 0 points, Unable to determine = 0 | RCT only | None |
| 14 | Was an attempt made to blind study subjects to the intervention they have received? | Yes = 1 point, No = 0 points, Unable to determine = 0 | RCT only | None |
| 15 | Was an attempt made to blind those measuring the main outcomes of the intervention? | Yes = 1 point, No = 0 points, Unable to determine = 0 | RCT only | None |
| 16 | If any of the results of the study were based on ‘data dredging,’ was this made clear? | Yes = 1 point, No = 0 points, Unable to determine = 0 | All | None |
| 17 | In trials and cohort studies, do the analyses adjust or different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls? | Yes = 1 point, No = 0 points, Unable to determine = 0 | Cohort; NCC | Cohort and nested case-control studies are the only designs that will have potentially measured the exposure (phthalates) prior to the outcome. For self-reported exposures, this question could also apply to case-control studies if the participants are asked about a time period prior to enrollment in the study.  |
| 18 | Were the statistical tests used to assess the main outcomes appropriate? | Yes = 1 point, No = 0 points, Unable to determine = 0 | All | None |
| 19 | Was compliance with the intervention/s reliable? | Yes = 1 point, No = 0 points, Unable to determine = 0 | RCT only | None |
| 20 | Were the main outcome measures used accurate (valid and reliable)? | Yes = 1 point, No = 0 points, Unable to determine = 0 | All | None |
| 21 | Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population? | Yes = 1 point, No = 0 points, Unable to determine = 0 | All | None  |
| 22 | Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time? | Yes = 1 point, No = 0 points, Unable to determine = 0 | All | None |
| 23 | Were study subjects randomized to intervention groups? | Yes = 1 point, No = 0 points, Unable to determine = 0 | RCT only | None |
| 24 | Was the randomized intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable? | Yes = 1 point, No = 0 points, Unable to determine = 0 | RCT only | None |
| 25 | Was there adequate adjustment for confounding in the analyses from which the main findings were drawn? | Yes = 1 point, No = 0 points, Unable to determine = 0 | All | If Q5 = 2, Q25 = 1; if Q5 = 1, Q25 = 0; if Q5 = 0, Q25 = 0 |
| 26 | Were losses of patients to follow-up taken into account? | Yes = 1 point, No = 0 points, Unable to determine = 0 | Cohort | None |
| 27 | Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%? | 0-5 possible points | All | “Did the study acknowledge that the analysis had adequate power and/or provide a power calculation, or was at least one significant result observed (indicating adequate power)?” – Yes = 1, No = 1, Unable to determine = 0 |

Abbreviations: RCT – randomized controlled trial; NCC – nested case-control study