

Supplemental Table 1: Quality assessment of included clinical and preclinical studies.

A) Quality assessment of included clinical studies using the Downs and Black Checklist (Downs and Black, 1998). The published checklist was adapted to remove items relating to an intervention.

Study	Items*																		Total
	1	2	3	5	6	7	10	11	12	15 [^]	16	17	18	20	21	22	25	27 ⁺	
Carter et al. (2014)	1	1	1	2	0	0	1	1	0	0	1	1	1	1	1	1	1	0	14
Håkonsen et al. (2014)	1	1	1	2	1	1	1	1	1	0	1	1	1	1	1	1	1	0	17
Ramlau-Hansen et al. (2010)	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	0	18
Robe et al. (1979)	1	1	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	0	6
Windham et al. (2004)	1	1	1	2	1	0	1	1	1	0	1	1	1	1	1	1	1	0	16

* Item numbers are as per Downs and Black (1998)(Downs and Black, 1998). The following items were removed as they refer to reporting of an intervention: 4, 8, 9, 13, 14, 19, 23, 24, and 26. Maximum possible score = 19. Scale of item score: for items 1 to 3 and 6 to 10: 1 = Yes, 0 = No; for item 5 – 2 = Yes, 1 = Partially, 0 = No; for items 11 to 27 - 1 = Yes, 0 = No or unable to determine. The Downs and Black Checklist covers the following areas: *Reporting* (items 1 -10; includes reporting of hypothesis/aim/objective, outcomes, patient characteristics, confounders/covariates, main findings, estimates of variability, confidence intervals and probabilities); *external validity* (items 11, 12; assessment of how representative the study participants are of the recruitable population); *internal validity* (bias, items 15-20; confounding, items 21-25); and *statistical power* (item 27). See Downs and Black (1998)(Downs and Black, 1998) for full details of the criteria for each item.

[^] For item 15, the ‘intervention’ was alcohol exposure.

⁺ A score of 1 was given if there was any mention of a power analysis being conducted to assess the required sample size to detect a statistically significant difference; otherwise score was 0.

B) Quality assessment of included pre-clinical studies using the ARRIVE guidelines (Kilkenny et al., 2010).

Study	Items																				Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Blanchard and Hannigan (1994)	2	1	1	1	0	1	1	1	2	0	0	2	1	0	1	2	0	2	0	2	20
Boggan et al. (1979)	2	1	1	1	0	1	1	1	2	0	0	1	0	0	0	0	0	1	1	2	15
Chen and Smith (1979)	1	1	1	1	0	1	1	1	0	0	0	2	1	0	1	2	0	1	0	1	15
Esquifino et al. (1986)	2	1	1	2	0	1	1	2	1	0	1	1	1	0	1	2	0	1	1	1	20
Hard et al. (1984)	2	1	1	0	0	1	1	1	1	0	0	2	1	0	1	1	0	1	0	1	15
Lan et al. (2009)	2	2	1	2	1	2	1	2	2	2	1	2	1	2	1	2	0	2	0	2	30
Lan et al. (2013)	2	2	2	2	2	2	1	2	1	1	1	2	2	2	0	2	0	2	2	2	32
McGivern (1987)	1	1	1	1	0	1	2	1	1	0	0	2	1	0	1	2	0	1	0	0	16
McGivern and Yellon (1992)	2	2	1	1	0	2	1	1	1	0	0	2	1	0	1	2	0	1	1	1	20
McGivern et al. (1992)	2	1	1	1	0	1	1	1	2	0	0	2	1	2	1	2	0	1	0	1	20
McGivern et al. (1995)	2	2	1	2	1	1	1	1	2	0	0	2	0	0	1	2	0	1	0	2	21
Murugan et al. (2013)	2	2	1	1	1	2	1	2	1	0	0	2	1	0	0	2	0	2	2	1	23
Polanco et al. (2010)	1	1	2	2	1	2	1	1	1	0	0	2	1	0	1	2	0	2	2	2	24
Polanco et al. (2010)	2	2	2	2	1	1	2	1	1	0	0	2	1	0	1	2	0	1	2	1	24
Shimamoto et al. (2006)	1	1	1	1	2	2	1	2	0	0	1	2	1	0	0	2	0	1	1	0	19
Sliwowska et al. (2016)	2	2	2	2	0	2	1	1	2	0	0	2	1	2^	0	2	0	1	1	2	25
Udani et al. (1985)	1	1	1	1	0	1	1	1	1	1	0	2	1	2	1	2	0	1	2	1	21
Ward et al. (1996)	2	1	1	2	0	1	1	2	2	1	1	2	1	2	0	2	0	1	0	1	23

Maximum possible score = 40. Scale of item score: 0 = no criteria covered; 1 = partial criteria covered; 2 = all criteria covered.

The ARRIVE criteria cover the following areas: (1) title, (2) abstract, (3) background, (4) objectives, (5) ethical statement, (6) study design, (7) experimental procedures, (8) experimental animals, (9) housing and husbandry, (10) sample size, (11) allocating animals to experimental groups, (12) experimental

outcomes, (13) statistical methods, (14) baseline data, (15) numbers analyzed, (16) outcomes and estimation, (17) adverse events, (18) interpretation/scientific implications, (19) generalizability/translation, (20) funding (see Kilkenny et al, 2010 (Kilkenny et al., 2010) for full details of criteria). For the sample size criterion (10), if clear n numbers were not provided for the experimental groups, a score of 0 was given.

^ Reported in Uban et al. (2013).

Notes on interpretation and application of the ARRIVE guidelines to scoring:

- For the title (1), if it was not clear that the study was investigating the offspring, a score of 1 was given. Also, the species should be provided to prevent confusion with a clinical study.
- For ethical statement (5), both the relevant guidelines for care and use of animals and ethical approval statement needed to be provided for a score of 2.
- For study design (6), where subjective bias could be introduced (e.g. when scoring for histology), evidence of blinding of researchers needed to be provided for a score of 2.
- For experimental procedures (7), score of 2 only given where justification for ethanol dose provided and method of euthanasia explicitly stated.
- For experimental animals (8), score of 2 only given if weight or age of dams to be treated was included, as well as age and sex of offspring.
- For housing and husbandry (9), details of housing and bedding material needed to be provided for a score of 2.
- For the sample size criteria (10, 15), if accurate n numbers were not provided for the experimental groups (i.e. only a range was given), a score of 0 was given. For criterion 10, sample size must be provided in the *methods* and number per litter indicated. For criterion 15, sample size analysed must be provided in the results (can be in figure legends). Only one study did a power analysis to determine sample size.
- For item (11), the order in which animals were treated was never provided, so maximum score was 1. If the method of randomisation was not given but the word 'random' (or something similar) was provided, a score of 1 was given. Previous experience indicates that the method of randomisation is rarely if ever given in preclinical studies.
- For experimental outcomes (12), these are rarely to never broken down into primary and secondary outcomes. As long as they were clearly stated, 2 points were given.
- For statistical methods (13), information on testing data to make sure it met the assumptions of the statistical test needed to be provided to receive a score of 2.
- For baseline data (14), note that this needed to be for the dams to be treated *before* treatment or testing. This was rarely provided.
- For item (18), did not consider if the authors discussed implications of their results for the 3Rs as previous experience has shown that this is never done in preclinical studies. This is assumed by the authors providing a statement of institutional ethics approval which requires that they address the 3Rs.

- For funding (20), the role of the funder needed to be provided for a score of 2. This could include indicating that the funds were for a specific author or indicating what roles the funders did *not* provide.

References

- Blanchard BA & Hannigan JH 1994 Prenatal ethanol exposure: Effects on androgen and nonandrogen dependent behaviors and on gonadal development in male rats. *Neurotoxicology and Teratology* **16** 31-39.
- Boggan WO, Randall CL & Dodds HM 1979 Delayed sexual maturation in female C57BL/6J mice prenatally exposed to alcohol. *Research Communications in Chemical Pathology and Pharmacology* **23** 117-125.
- Carter RC, Jacobson JL, Dodge NC, Granger DA & Jacobson SW 2014 Effects of prenatal alcohol exposure on testosterone and pubertal development. *Alcoholism: Clinical and Experimental Research* **38** 1671-1679.
- Chen JJ & Smith ER 1979 Effects of perinatal alcohol on sexual differentiation and open-field behavior in rats. *Hormones and Behavior* **13** 219-231.
- Downs SH & Black N 1998 The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health* **52** 377-384.
- Esquifino AI, Sanchis R & Guerri C 1986 Effect of prenatal alcohol exposure on sexual maturation of female rat offspring. *Neuroendocrinology* **44** 483-487.
- Håkonsen LB, Brath-Lund ML, Hounsgaard ML, Olsen J, Ernst A, Thulstrup AM, Bech BH & Ramlau-Hansen CH 2014 In utero exposure to alcohol and puberty in boys: A pregnancy cohort study. *BMJ Open* **4** e004467.
- Hard E, Dahlgren IL, Engel J, Larsson K, Liljequist S, Lindh AS & Musi B 1984 Development of sexual behavior in prenatally ethanol-exposed rats. *Drug and Alcohol Dependence* **14** 51-61.
- Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG & Nc3rs Reporting Guidelines Working Group 2010 Animal research: Reporting in vivo experiments: The ARRIVE guidelines. *British Journal of Pharmacology* **160** 1577-1579.
- Lan N, Vogl AW & Weinberg J 2013 Prenatal ethanol exposure delays the onset of spermatogenesis in the rat. *Alcoholism: Clinical and Experimental Research* **37** 1074-1081.
- Lan N, Yamashita F, Halpert AG, Sliwowska JH, Viau V & Weinberg J 2009 Effects of prenatal ethanol exposure on hypothalamic-pituitary-adrenal function across the estrous cycle. *Alcoholism: Clinical and Experimental Research* **33** 1075-1088.
- Mcgovern RF 1987 Influence of prenatal exposure to cimetidine and alcohol on selected morphological parameters of sexual differentiation: a preliminary report. *Neurotoxicology and Teratology* **9** 23-26.
- Mcgovern RF, Mcgeary J, Robeck S, Cohen S & Handa RJ 1995 Loss of reproductive competence at an earlier age in female rats exposed prenatally to ethanol. *Alcoholism: Clinical and Experimental Research* **19** 427-433.
- Mcgovern RF, Raum WJ, Handa RJ & Sokol RZ 1992 Comparison of two weeks versus one week of prenatal ethanol exposure in the rat on gonadal organ weights, sperm count, and onset of puberty. *Neurotoxicology and Teratology* **14** 351-358.
- Mcgovern RF & Yellon SM 1992 Delayed onset of puberty and subtle alterations in GnRH neuronal morphology in female rats exposed prenatally to ethanol. *Alcohol* **9** 335-340.

- Murugan S, Zhang C, Mojtahedzadeh S & Sarkar DK 2013 Alcohol exposure in utero increases susceptibility to prostate tumorigenesis in rat offspring. *Alcoholism: Clinical and Experimental Research* **37** 1901-1909.
- Polanco TA, Crismale-Gann C, Reuhl KR, Sarkar DK & Cohick WS 2010 Fetal alcohol exposure increases mammary tumor susceptibility and alters tumor phenotype in rats. *Alcoholism: Clinical and Experimental Research* **34** 1879-1887.
- Ramlau-Hansen CH, Toft G, Jensen MS, Strandberg-Larsen K, Hansen ML & Olsen J 2010 Maternal alcohol consumption during pregnancy and semen quality in the male offspring: two decades of follow-up. *Human Reproduction* **25** 2340-2345.
- Robe LB, Robe RS & Wilson PA 1979 Maternal heavy drinking related to delayed onset of daughters menstruation. *Currents in Alcoholism* **7** 515-520.
- Shimamoto A, Liu J, Kozawa S & Fujimiya T 2006 Determination of endogenous testosterone in rat tissues following fetal alcohol exposure using HPLC with UV detection. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences* **836** 69-73.
- Sliwowska JH, Comeau WL, Bodnar TS, Ellis L & Weinberg J 2016 Prenatal alcohol exposure and pair feeding differentially impact puberty and reproductive development in female rats: Role of the kisspeptin system. *Alcoholism: Clinical and Experimental Research* **40** 2368-2376.
- Uban KA, Comeau WL, Ellis LA, Galea LA & Weinberg J 2013 Basal regulation of HPA and dopamine systems is altered differentially in males and females by prenatal alcohol exposure and chronic variable stress. *Psychoneuroendocrinology* **38** 1953-1966.
- Udani M, Parker S, Gavalier J & Van Thiel DH 1985 Effects of in utero exposure to alcohol upon male rats. *Alcoholism: Clinical and Experimental Research* **9** 355-359.
- Ward IL, Ward OB, Mehan D, Winn RJ, French JA & Hendricks SE 1996 Prenatal alcohol and stress interact to attenuate ejaculatory behavior, but not serum testosterone or LH in adult male rats. *Behavioral Neuroscience* **110** 1469-1477.
- Windham GC, Bottomley C, Birner C & Fenster L 2004 Age at menarche in relation to maternal use of tobacco, alcohol, coffee, and tea during pregnancy. *American Journal of Epidemiology* **159** 862-871.