

**Microplastics Occurrence, Health Effects, and Mitigation Policies:
An Evidence Review for the California Legislature**

Appendix C.1 Risk of Bias Ratings and Justification

Risk of bias assessments are completed for each individual study. There are 7 domains that must be rated using the following categories: low, probably low, probably high, or high. This document contains the ratings (with justifications) for the 7 digestive studies and 6 reproductive studies included in the evaluation of the evidence. The risk of bias ratings are then used to evaluate study quality, which informs the strength of the evidence rating and ultimate hazard identification conclusion.

Digestive Studies: Risk of Bias Ratings

Jin et al. 2019		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	There is insufficient information about the sequence generation process to permit a judgment of low risk of bias. Study authors make a simple statement such as ‘we randomly allocated’, but do not provide details regarding specific random components used in the sequence generation process.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Apical: AB-PAS/mucus secretion coverage ratio (colon): Low	Animals included in the analysis are exactly those who were randomized into the experiment. The number of animals allocated to treatment groups is reported for outcomes of interest and data are provided indicating adequate follow up of all animals from the beginning of the study.
Selective Outcome Reporting	Probably low	All of the study’s pre-specified outcomes outlined in the methods, that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Probably low	A conflict-of-interest statement denying financial interests is not provided, but associated funds and/or persons appear to be free of financial interests in study outcome and are unaffiliated with parties with a financial interest.

Other Potential Threats to Validity- Outcome Evaluation	Apical: AB-PAS/mucus secretion coverage ratio (colon): Probably low	Methods were clearly described but there is no mention of random selection or blinding for this analysis. They did select multiple colon sections and image fields for statistical testing. Pixel intensity and ratio by an image analyzer was used to assess this measurement. Therefore, there is no sufficient evidence of high risk of bias.
---	---	--

Lu et al. 2018		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	There is insufficient information about the sequence generation process to permit a judgment of low risk of bias, but study authors make a simple statement such as 'we randomly allocated', but do not provide details regarding specific random components used in the sequence generation process.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Apical: Decreased mucin secretion (colon): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome (i.e., no missing outcome data).
Selective Outcome Reporting	Probably low	All of the study's pre-specified outcomes outlined in the methods, that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Probably low	A conflict-of-interest statement denying financial interests is not provided, but associated funds and/or persons appear to be free of financial interests in study outcome and are unaffiliated with parties with a financial interest.
Other Potential Threats to Validity - Outcome Evaluation	Apical: Decreased mucin secretion (colon): Probably low	Mucus coverage ratio was measured by 6 sections of AB-PAS staining in each group. The text does not specify how many mice per group. The pixels were determined by an image analyzer.

B Li et al. 2020		
Domain	Rating	Justification for rating
Sequence Generation	Probably high	All of the study's pre-specified (primary and secondary) outcomes outlined in the methods, that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Induces chronic inflammation: IL-1 α , G-CSF, IL-2, IL-5, IL-6, IL-9, IP-10 and RANTES IL-6, IL-10, IL-1 (intestine): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome (i.e., no missing outcome data).
Selective Outcome Reporting	Probably low	All of the study's pre-specified (primary and secondary) outcomes outlined in the methods, that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity – Outcome Evaluation	Induces chronic inflammation: \ IL-1 α , G-CSF, IL-2, IL-5, IL-6, IL-9, IP-10 and RANTES: Probably low	IL-1 α , G-CSF, IL-2, IL-5, IL-6, IL-9, IP-10 and RANTES are a standard biomarkers for inflammation. They are measured either by ELISA kits or qt-PCR. All authors followed standard procedure to measure the biomarkers with enough samples to perform statistical analysis. Therefore, there is no evidence of potential high risk of bias.

Choi et al. 2021a		
Domain	Rating	Justification for rating
Sequence Generation	Probably high	Study authors do not make any statement about sequence generation and the review author does not find indirect evidence suggesting random sequence generation.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Induces chronic inflammation: TNF- α , IL-1 α , IL-6, iNOS, COX-2, NF-kB (intestine): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals).
	Induces oxidative stress: SOD activity, SOD expression relative levels (intestine): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals.
	Induces oxidative stress: ROS concentration (intestine): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals.
Selective Outcome Reporting	Probably low	All of the study's pre-specified (primary and secondary) outcomes outlined in the methods that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to

		indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity – Outcome Evaluation	Induces chronic inflammation: TNF- α , IL-1 α , IL-6, iNOS, COX-2, NF-kB (intestine): Probably low	TNF, IL-6, and IL-1 α are standard biomarkers for inflammation. They are measured either by ELISA kits or qt-PCR. All authors followed standard procedure to measure the biomarker with enough samples to perform statistical analysis. Therefore, there is no evidence of potential high risk of bias.
	Induces oxidative stress: SOD activity, SOD expression relative levels (intestine): Probably low	SOD measurement was performed by commercial kit. There is no evidence of potential high risk of bias. There were enough sample for statistical analysis and blinding does not affect the RoB for this outcome.
	Induces oxidative stress: ROS concentration, (intestine): Probably low	ROS measurement was performed by commercial kit. There is no evidence of potential high risk of bias. There were enough sample for statistical analysis and blinding does not affect the RoB for this outcome.

Choi et al. 2021b		
Domain	Rating	Justification for rating
Sequence Generation	Probably high	Study authors do not make any statement about sequence generation and the review author does not find indirect evidence suggesting random sequence generation.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.

Incomplete Outcome Data	Apical: Mucosa thickness, Muscle thickness, Flat luminal surface thickness, Crypt layer thickness (mid colon): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals.
	Modulates receptor-mediated effects: CCK concentration, Gastrin concentration (mid colon): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals.
	Apical: Charcoal transit ratio (mid colon): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals.
	Apical: Intestine length (mid colon): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals.
	Alters cell proliferation, cell death, or nutrient supply: Number of crypt of Lieberkuhn (mid colon): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals.
	Alters cell proliferation, cell death, or nutrient supply: Goblet cell counts (mid colon): Probably high	Study authors do not report numbers of animals allocated to treatment groups but do provide data indicating adequate follow up for a subset of animals.

Selective Outcome Reporting	Probably Low	All of the study's pre-specified outcomes outlined in the protocol, methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way, but study authors only report a subset of animals were examined for outcome of interest.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity – Outcome Evaluation	Apical: Mucosa thickness, Muscle thickness, Flat luminal surface thickness, Crypt layer thickness (mid colon): Probably high	H&E stained sections of mid-colon were used to measure the histopathological parameters. They were measured in duplicate for each slide. 4 to 6 mice per group were used for this analysis. They have examples of the measurements in the paper. However, it looks like the measurements can fluctuate depending on where the parameter is measured within the slide. This could lead to a high risk of user bias
	Modulates receptor-mediated effects: CCK concentration, Gastrin concentration (mid colon): Probably low	Hormone measure was performed on ELISA kit. There is not a 100% certainty of low risk of bias, but there is no evidence of high risk of bias for this measurement.
	Apical: Charcoal transit ratio (mid colon): Probably high	The charcoal transit ratio is measured by light microscopy and measuring the 'charcoal' section of the intestine. They do show example light images in the publication, but it is ambiguous when the 'charcoal' section begins or ends. There is no scale bar for comparison. The intestinal length is also not stretched completely in the image.
	Apical: Intestine length (mid colon): Probably high	Similar criticism as the charcoal transit ratio measurement. All measurements were done via optical imaging. The image does not stretch the intestine all the way and no measurement metric for comparison.
	Alters cell proliferation, cell death, or nutrient supply: Number of crypt of Lieberkuhn (mid colon): Probably high	The samples were prepped by fixing thin tissue sections onto grids and the measurement was done by TEM (transmission electron microscopy). The authors did not explicitly state the number of tissue sections or mention blinding so there could be selective bias.

	<p>Alters cell proliferation, cell death, or nutrient supply: Goblet cell counts (mid colon): Probably high</p>	<p>H&E stained sections of mid-colon were used to measure the histopathological parameters. They were measured in duplicate for each slide. 4 to 6 mice per group were used for this analysis. They have examples of the measurements in the paper. The number of Goblet cells was also not normalized to any section of the tissue; therefore, it is difficult to compare across trials. There's some evidence of potential high risk of bias but not definitive.</p>
--	---	--

Djouina et al. 2022		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	Study authors make a simple statement such as 'we randomly allocated', but do not provide details regarding specific random component used in the sequence generation process
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	<p>Induces chronic inflammation: mRNA quantification of inflammatory cytokines TNF-α, Ifng, Il6, and Il1b (distal & proximal small intestine & colon) : Probably low</p>	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
	<p>Apical: Villus length (Distal and proximal small intestine): Probably low</p>	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
	<p>Villus/crypt ratio: Small intestine (proximal), Small intestine (distal): Probably low</p>	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals

Apical: AB/PAS positive area (Distal and proximal small intestine, colon – epithelium mucosal surface area) =: Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
Apical: Mucosal surface area (Colon-(epithelium): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
Is immunosuppressive: CD4 T lymphocytes (Proximal small intestine): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
Is immunosuppressive: CD8 T lymphocytes (proximal small intestine): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
Is immunosuppressive: Dendritic cells (proximal small intestine): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
Is immunosuppressive: Inflammatory monocytes: (proximal small intestine): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
Is immunosuppressive: Neutrophils (colon) Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
Is immunosuppressive: Anti-inflammatory macrophages (colon): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
Apical: Crypt depth: (colon, proximal small intestine): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.

Selective Outcome Reporting	Probably low	All of the study's pre-specified outcomes outlined in the methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way, but study authors report the number of animals analyzed for outcomes of interest as a range or report values for which numbers of animals analyzed need to be calculated by the review author.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity- Outcome evaluation	Induces chronic inflammation: mRNA quantification of inflammatory cytokines TNF- α , Il6, Il-b (distal & proximal small intestine, & colon): Probably low	Standard biomarkers for inflammation. They are measured either by ELISA kits or qt-PCR. All authors followed standard procedure to measure the biomarker with enough samples to perform statistical analysis. Therefore, there is no evidence of potential high risk of bias.
	Apical: Villus length: (proximal and distal small intestine) Probably low	These measurements were performed by microscopy and imaging software. Representative tissue samples were fixed with 4% paraformaldehyde. These are typical routine methods for these apical endpoints. However, they did not mention any blinding, which is critical for image-based measurements. They presented vague details of the number of tissues measured and number of mice measured in each group. However, "at least 5" mice is a sufficient number for statistical testing.
	Apical: Villus/crypt ratio: (proximal and distal small intestine) Probably low	These measurements were performed by microscopy and imaging software. Representative tissue samples were fixed with 4% paraformaldehyde. These are typical routine methods for these apical endpoints. However, they did not mention any blinding, which is critical for image-based measurements. They presented vague details of the number of tissues measured and number of mice measured in each group. However, "at least 5" mice is a sufficient amount for statistical testing.
	Apical: AB/PAS positive area (proximal and distal small intestine, colon - epithelium mucosal surface area) Probably low	These measurements were performed by microscopy and imaging software. Representative tissue samples were fixed with 4% paraformaldehyde. These are typical routine methods for these apical endpoints. However, they did not mention any blinding, which is critical for image-based measurements. They presented vague details of the number of tissues measured and number of mice measured in each group. However, "at least 5" mice is a sufficient amount for statistical testing.

<p>Apical: Mucosal surface area (Colon- epithelium) Probably low</p>	<p>These measurements were performed by microscopy and imaging software. Representative tissue samples were fixed with 4% paraformaldehyde. These are typical routine methods for these apical endpoints. However, they did not mention any blinding, which is critical for image-based measurements. They presented vague details of the number of tissues measured and number of mice measured in each group. However, "at least 5" mice is a sufficient amount for statistical testing.</p>
<p>Is immunosuppressive: CD4 T lymphocytes (proximal small intestine) Probably low</p>	<p>Measurements were performed via cell isolation and flow cytometry with standard methodology. There is not a 100% certainty of low risk of bias, but there is no evidence of high risk of bias for this measurement.</p>
<p>Is immunosuppressive: CD8 T lymphocytes (Proximal small intestine) Probably low</p>	<p>Measurements were performed via cell isolation and flow cytometry with standard methodology. There is not a 100% certainty of low risk of bias, but there is no evidence of high risk of bias for this measurement.</p>
<p>Apical: Crypt depth (Colon, proximal small intestine) Probably low</p>	<p>These measurements were performed by microscopy and imaging software. Representative tissue samples were fixed with 4% paraformaldehyde. These are typical routine methods for these apical endpoints. However, they did not mention any blinding, which is critical for image-based measurements. They presented vague details of the number of tissues measured and number of mice measured in each group. However, "at least 5" mice is a sufficient amount for statistical testing.</p>
<p>Is immunosuppressive: Inflammatory monocytes (proximal small intestine) Probably low</p>	<p>Measurements were performed via cell isolation and flow cytometry with standard methodology. There is not a 100% certainty of low risk of bias, but there is no evidence of high risk of bias for this measurement.</p>
<p>Is immunosuppressive: Neutrophils (colon) Probably low</p>	<p>Measurements were performed via cell isolation and flow cytometry with standard methodology. There is not a 100% certainty of low risk of bias, but there is no evidence of high risk of bias for this measurement.</p>
<p>Is immunosuppressive: Anti-inflammatory macrophages (colon) Probably low</p>	<p>Measurements were performed via cell isolation and flow cytometry with standard methodology. There is not a 100% certainty of low risk of bias, but there is no evidence of high risk of bias for this measurement.</p>

Wen et al. 2022		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	Study authors make a simple statement such as ‘we randomly allocated’, but do not provide details regarding specific random component used in the sequence generation process.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Alters cell proliferation, cell death, or nutrient supply: Goblet cell counts (colon): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
	Oxidative stress: Colonic glutathione (GSH, SOD, MDA) (colon): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
	Apical: Colon length (colon) Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
	Induces chronic inflammation: TNF- α , IL-6, and IL-10 (colon): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
	Muscular layer width, Crypt depth (colon, proximal small intestine): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
Selective Outcome Reporting	Probably Low	All of the study’s pre-specified outcomes outlined in the methods, abstract, that are of interest in the review have been reported in the pre-specified way, but study authors only report a subset of animals were examined for outcome of interest.

Conflict of Interest	Low	Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations without financial interest in the treatments studied.
Other Potential Threats to Validity- Outcome Evaluation	Alters cell proliferation, cell death, or nutrient supply: Goblet cell counts (colon): Probably high	The muscular layer width, crypt depth was measured using ImageJ 6.0, goblet cells were counted under an optical microscope. There is no mention of blinding. N = 3 for each trial tested. There's no guidance for the reader on how they conclude their measurements. No example measurement images. There is some evidence of potentially high risk of bias, but not definitive.
	Oxidative stress: Colonic glutathione (GSH, SOD, MDA) (colon): Probably high	GSH, SOD, MDA measurements were performed with commercial kits, but the liver and colon tissue samples were homogenized together. Therefore, there is some evidence of potential risk of bias because of the blending between liver and colon. There were enough samples for statistical analysis. Sample blinding does not influence RoB for this outcome.
	Apical: Colon length (colon): Probably low	Colon shortening was measured by a ruler and microscope. Example colon was clearly laid out in one of the figures with a ruler for comparison. There is no sufficient evidence of potential risk of bias.
	Induces chronic inflammation: Pro-inflammation cytokines (TNF- α , IL-6, and IL-10) (colon): Probably low	Standard biomarkers for inflammation. They are measured either by ELISA kits or qt-PCR. All authors followed standard procedure to measure the biomarker with enough samples to perform statistical analysis. Therefore, there is no evidence of potential high risk of bias.
	Muscular layer width, Crypt depth (colon, proximal small intestine): Probably high	The muscular layer width, crypt depth was measured using ImageJ 6.0, goblet cells were counted under an optical microscope. There is no mention of blinding. N = 3 for each trial tested. There's no guidance for the reader on how they conclude their measurements. No example measurement images. There is some evidence of potentially high risk of bias, but not definitive.

Reproductive Studies: Risk of Bias Ratings

An et al. 2021		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	Study authors make a simple statement such as ‘we randomly allocated’, but do not provide details regarding specific random component used in the sequence generation process.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Apical: Number of growing follicles (ovaries): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
	Alters hormone receptor signaling; alters reproductive hormone production, secretion, or metabolism: AMH levels (ovaries): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
Selective Outcome Reporting	Probably low	All of the study’s pre-specified outcomes outlined in the methods that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity-	Apical: Number of growing follicles (ovaries): Probably high	The number of growing follicles were observed by 5 visual fields under a microscope. Since it is a visual observation and no mention of blinding, there is evidence of potential high risk of bias.

Outcome Evaluation	Alters hormone receptor signaling; alters reproductive hormone production, secretion, or metabolism: AMH levels (ovaries): Probably low	AMH levels were measured with ELISA kits. Because these are standard markers and they followed manufacturer's direction, these outcomes have no evidence of high risk of bias.
--------------------	---	--

J Hou et al. 2021		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	There is insufficient information about the sequence generation process to permit a judgment of low risk of bias, but study authors make a simple statement such as 'we randomly allocated', but do not provide details regarding specific random components used in the sequence generation process.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Apical: Number of growing follicles (ovaries): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
	Alters hormone receptor signaling; alters reproductive hormone production, secretion, or metabolism: AMH levels (pg/ml) IL-18 (pg/ml) IL-1 β (pg/ml) (serum): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for all animals, only a subset of animals.
Selective Outcome Reporting	Probably low	All of the study's pre-specified outcomes outlined in the methods that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.

Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity-Outcome Evaluation	Apical: Number of growing follicles (ovaries): Probably high	The number of growing follicles were observed by 5 visual fields under a microscope. Since it is a visual observation and no mention of blinding, there is evidence of potential high risk of bias.
	Alters hormone receptor signaling; alters reproductive hormone production, secretion, or metabolism: Hormone level changes: AMH levels (pg/ml) IL-18 (pg/ml) IL-1 β (pg/ml) (serum): Probably low	AMH, IL-18 & IL-1 β levels were measured with ELISA kits. Because these are standard markers and they followed manufacturer's direction, these outcomes have no evidence of high risk of bias.

Huang et al. 2022		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	There is insufficient information about the sequence generation process to permit a judgment of low risk of bias, but there is indirect evidence that suggests the sequence generation process was random.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Apical: Sperm count (epididymis): Probably low	The number of animals allocated is reported and matches the number of animals reported for each outcome (i.e., no missing outcome data). However, the originally, 32 pregnant mice were randomized into four groups (8 dams each), but only six dams and their male pups per group were followed during the experimental period.

Selective Outcome Reporting	Probably low	All of the study's pre-specified (primary and secondary) outcomes outlined in the methods that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity-Outcome Evaluation	Apical: Sperm count (epididymis): Probably high	They used a hemocytometer to count the number of sperm according to manufacturer instructions. There was no blinding during this process. Only a brief description of the method. Therefore, there is some evidence of potential high risk of bias.

B Hou et al. 2021		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	There is insufficient information about the sequence generation process to permit a judgment of low risk of bias. Study authors make a simple statement such as 'we randomly allocated', but do not provide details regarding specific random components used in the sequence generation process.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Apical: Rate of living sperm (Testis): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome (i.e., no missing outcome data).
	Apical: Sperm malformation (Testis): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome (i.e., no missing outcome data).
Selective Outcome Reporting	Probably low	All of the study's pre-specified outcomes outlined in the methods that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in

		the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity-Outcome Evaluation	Apical: Rate of living sperm (Testis): Probably high	A drop of sperm is dropped onto a slide and the living sperm was counted. One major risk of bias is the timing between preparing the sample and analyzing the sperm count. A longer processing time could lead to more sperm dying and then skewing the result. There was no mention of blinding which is important for this outcome.
	Apical: Sperm malformation (Testis): High	Sperm shape was analyzed by a microscope and observation. There is no mention how they got these results based on metrics. There are only a few descriptor words based on the shape. Therefore, this result has sufficient evidence of high risk of bias. There are example photographs but no explanation on how they get the result. There was no blinding.

Li et al. 2021		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	There is insufficient information about the sequence generation process to permit a judgment of low risk of bias. Study authors make a simple statement such as 'we randomly allocated', but do not provide details regarding specific random components used in the sequence generation process.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably high	Study authors do not make any statement about blinding and the review author does not find indirect evidence suggesting blinding.
Incomplete Outcome Data	Apical: Sperm abnormality (Testis): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome.
	Apical: Sperm concentration (Testis): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome.
	Apical: Sperm motility (Testis): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome.

Selective Outcome Reporting	Probably low	All of the study's pre-specified outcomes outlined in the methods that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity-Outcome Evaluation	Apical: Sperm motility (Testis): Probably low	The sperm abnormality was measured by an automatic sperm analyzer. There is no evidence of potential high risk of bias.
	Apical: Sperm concentration (Testis): Probably high	The sperm concentration was measured via microscope. The sperm suspension was smeared on a glass slide and five smears for each animal and 12 fields were randomly collected for observation. Blinding was never mentioned. Because the measurement was based on user observation, that could be high risk of bias.
	Apical: Sperm abnormality (Testis): Probably high	The sperm abnormality was measured via microscope. The sperm suspension was smeared on a glass slide and five smears for each animal and 12 fields were randomly collected for observation. Blinding was never mentioned. Because the measurement was based on user observation, that could be high risk of bias. They clearly explain the definition of sperm abnormality.

Jin et al. 2022		
Domain	Rating	Justification for rating
Sequence Generation	Probably low	There is insufficient information about the sequence generation process to permit a judgment of low risk of bias. Study authors make a simple statement such as 'we randomly allocated', but do not provide details regarding specific random components used in the sequence generation process.
Allocation Concealment	Probably high	Study authors do not make any statement about allocation concealment and the review author does not find indirect evidence suggesting allocation concealment.
Blinding of personnel and outcome assessors	Probably low	Study authors describe blinding for one experiment, and the methods for a second experiment are similar, but do not specifically mention blinding.

Incomplete Outcome Data	Apical: Viability of sperm (Testis): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome (i.e., no missing outcome data).
	Apical: Sperm deformity (Testis): Low	The number of animals allocated is reported and matches the number of animals reported for each outcome (i.e., no missing outcome data).
	Alters production and levels of reproductive hormones OR Alters hormone receptor levels/functions: Testosterone LH levels (ng/ml) FSH levels (ng/ml) Concentrations of testosterone (ng/ml) (serum): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals.
	Apical: Seminiferous tubular diameter, Germinal epithelium thickness (Testis): Probably low	Study authors report number of animals allocated to treatment groups, but do not provide data indicating adequate follow up for a subset of animals or only provide a qualitative statement about missing outcome data.
Selective Outcome Reporting	Probably low	All of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way (i.e., the outcomes outlined in the methods section match what is reported in the results section and vice versa), and the number of animals analyzed for outcomes of interest is provided.
Conflict of Interest	Low	The study did not receive support from a company, study author, or other party having a financial interest in the outcome of the study. A conflict-of-interest statement is provided to indicate the study authors have no financial interests and there is evidence of the parties not having a financial interest.
Other Potential Threats to Validity-Outcome Evaluation	Apical: Viability of sperm (Testis): Probably low	They used a hemocytometer to count the number of sperm according to manufacturer instructions. The person was blinded. There is no evidence of potential high risk of bias.
	Apical: Sperm deformity (Testis): Probably high	The person was blinded and assessed percentage of 1000 sperm that were malformed. Sperm deformity included specific shapes and morphologies manifested but still depends on user observation, which has some bias. There is some evidence of potential high risk of bias.
	Alters production and levels of reproductive hormones OR Alters hormone receptor levels/functions:	Testosterone, LH, and FSH were measured with ELISA kits. Because these are standard markers and they followed manufacturer's direction, these outcomes have no evidence of high risk of bias.

	<p>Testosterone LH levels (ng/ml); FSH levels (ng/ml); Concentrations of testosterone (serum) (ng/ml): Probably Low</p>	
	<p>Apical: Testicular Damage: Seminiferous tubular diameter, Germinal epithelium thickness (Testis): Probably high</p>	<p>Sufficient number of slides for statistical analysis; used image-based software for measurement; no mention of blinding; 'round' or 'nearly round' is not clearly defined so could have potentially high selection bias.</p>