Endocrine-Disruption Evaluation of Oxybenzone (Benzophenone-3)

Weight-of-Evidence Assessment for EATS Modalities

Insilica LLC

December 19, 2025

CONFIDENTIAL DRAFT — NOT FOR DISTRIBUTION

EARLY ALPHA VERSION — FOR INTERNAL REVIEW ONLY

This document is a **preliminary output** from an automated toxicological assessment pipeline under active development. It is provided for internal review and demonstration purposes only.

Important Notices:

- This tool and its outputs are in an **early alpha stage** and have not been validated for regulatory or decision-making use
- The methodology relies on large language models (LLMs) whose outputs have not been systematically benchmarked against expert curation
- All findings should be considered **hypothesis-generating** rather than definitive conclusions
- This document is **confidential** and should not be shared, distributed, or cited outside of authorized review contexts
- Do not use this report as the basis for regulatory submissions, product decisions, or public communications

Document Version: [redacted] **Generated:** December 19, 2025

Contents

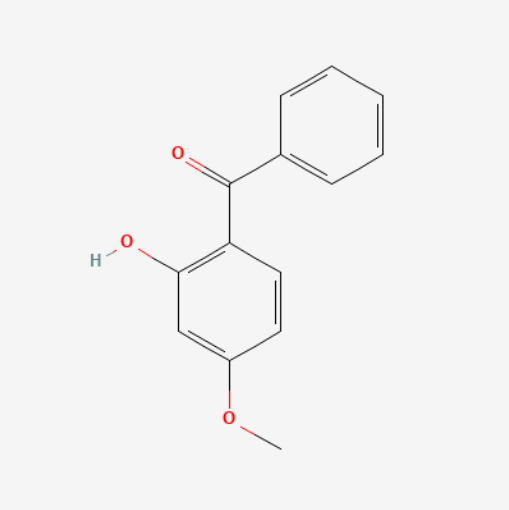
List of Tables

List of Figures

Chemical Identity

Chemical Identification	
Chemical Name	Oxybenzone
Synonyms	Benzophenone-3, BP-3, 2-Hydroxy-4-methoxybenzophenone
CAS Number	131-57-7
Molecular Formula	C14H12O3
Molecular Weight	228.24 g/mol
DTXSID	DTXSID3022405
PubChem CID	4632
Use	UV filter in sunscreens and cosmetics

Structure Identifiers	
SMILES	`COC1=CC(=C(C=C1)C(=O)C2=CC=CC=C2)O`
InChIKey	`DXGLGDHPHMLXJC-UHFFFAOYSA-N`



2D structure of Oxybenzone

Primary Metabolite: Benzophenone-1 (CAS 131-56-6). Primary metabolite via O-demethylation, enhanced estrogenic activity.

Executive Summary

Final Assessment

An LLM-generated weight-of-evidence conclusion synthesizing all pipeline outputs is available in Part II: Conclusion & Assessment. This automated hazard characterization integrates literature findings, database evidence, and mechanistic analysis to provide a comprehensive determination of oxybenzone's endocrine disruption potential.

This report presents an automated review evaluating whether oxybenzone (benzophenone-3, BP-3) may cause endocrine disruption. Oxybenzone is a UV-absorbing compound widely used in sunscreens and personal care products. The analysis pipeline employs large language models (LLMs) and machine learning classifiers to process scientific literature at scale, extracting quantitative findings and mechanistic evidence relevant to endocrine-mediated toxicity. This framework is designed to be exposure- and outcome-agnostic, enabling future extension to other chemical-hazard pairs with minimal reconfiguration. Detailed methods are provided in Part III.

Key Finding

Oxybenzone demonstrates **estrogen receptor agonist** and **androgen receptor antagonist** activity in high-throughput screening. Tox21 data (19/42 ER assays active) and computational predictions (ADMET-AI: 97th percentile; OPERA CERAPP: 0.80–0.87 confidence) consistently indicate direct ER binding. OPERA's CoMPARA model clarifies that AR activity represents antagonism (0.80 confidence), not agonism (0.98 confidence negative), explaining the anti-androgenic effects via SHBG elevation. The compound is readily absorbed through human skin and detected in urine, blood, and breast milk. Of particular concern is that oxybenzone's Phase I metabolite, benzophenone-1 (BP-1), demonstrates similar or enhanced estrogenic activity in vitro.

The following sections summarize the key stages and analyses conducted in this assessment:

Metabolite Identification. We identified 4 major metabolites of oxybenzone through computational prediction (SyGMa) corroborated by literature mining and Neptune knowledge graph annotations. Unlike many xenobiotics, oxybenzone metabolism is primarily Phase II conjugation: **oxybenzone glucuronide** (potentially phototoxic), **oxybenzone sulfate** (primary urinary excretion product), and **2,4'-dimethoxybenzophenone** (O-methylation product). The Phase I metabolite **benzophenone-1 (BP-1)**, formed via O-demethylation, is toxicologically significant as it demonstrates stronger estrogenic activity than the parent compound. PubChem annotations provided 1,462 records for oxybenzone (CID 4632), with 49 metabolism-related annotations supporting the identified metabolic pathways.

Literature Review. Automated literature mining queried PubMed, CTD, and PubTator3 databases using complementary search strategies. A total of **2,006 papers** were identified and assessed, with **62 papers** proceeding to full bias assessment. From these, **175 key events** were extracted across the endocrine disruption pathway.

Global Document Review. Across the literature corpus, 8,303 extracted events were consolidated into 4,312 canonical events, and 4,492 mechanistic relationships were identified with an average evidence strength of 0.72.

Four primary mechanistic pathways were characterized: (1) oxidative stress leading to multi-system toxicity (52 connected pathways), (2) ERα activation leading to carcinogenesis and reproductive toxicity (23 pathways), (3) HPG axis suppression leading to developmental toxicity (25 pathways), and (4) androgen system perturbation leading to male reproductive failure (16 pathways).

Computational Tool Review. Eight computational toxicology tools were evaluated, including ADMET-AI, OPERA (CERAPP/CoMPARA), ChEMBL Multitask, StopTox, Admetica, SolTranNet, Mordred, and PubChemPy. [redacted] predicts high estrogen receptor activity (NR-ER: 0.75, 97th percentile among approved drugs). [redacted] predicts ER agonism (0.80 confidence), ER antagonism (0.86 confidence), and ER binding (0.87 confidence). Critically, [redacted] predicts no AR agonism (0.98 confidence) but AR antagonism (0.80 confidence), clarifying the mechanism of oxybenzone's anti-androgenic activity.

Important Papers Review. High-impact publications were identified and ranked by human relevance, study design quality, statistical significance, and endocrine disruption relevance. Key studies demonstrate ERα and ERβ binding activity, anti-androgenic effects via SHBG elevation, and detectable systemic exposure from dermal application.

SCCS Regulatory Review. Automated extraction from SCCS [redacted] (2021) identified **32 endocrine endpoints** from guideline studies. The SCCS established a NOAEL of [redacted] based on spermatocyte effects in male rat offspring [redacted]. Low-dose effects were observed at [redacted] in mice (AGI changes, mammary gland ERα+ cell fraction). The SCCS noted effects on spermatocytes "may be due to an ED effect of BP-3," aligning with OPERA CoMPARA's prediction of AR antagonism.

Database Review. Systematic database queries (ChemHarmony, CTDbase, ToxValDB, ToxRefDB, ICE, Tox21, PubChem) returned 4,477 total activity records. Oxybenzone showed **19/42 active results** [redacted]. BP-1 [redacted]). Thyroid pathway showed 0/27 active, indicating pathway specificity. The compound is classified H361 ("Suspected of damaging fertility or the unborn child") in REACH registrations. Notable data gap: ToxRefDB contains zero OECD guideline-compliant DART studies.

Hazard Characterization. The integrated weight-of-evidence assessment concludes that oxybenzone **meets criteria for classification as an endocrine-active substance with concern**, demonstrating activity through receptor-mediated pathways. The primary mechanism is ERα/ERβ agonism, with secondary anti-androgenic activity via AR antagonism. Concern level is rated **Moderate to High** based on widespread human exposure (60–97% US population), demonstrated mechanism, metabolite amplification, and vulnerable population exposure. Confidence is **High** for the in vitro findings based on reproducible HTS results, cross-database consistency, and structural plausibility; however, definitive regulatory classification typically requires in vivo guideline studies. Full hazard characterization is presented in Part II.

Part I

Results

1 Metabolite Identification

Oxybenzone (benzophenone-3) undergoes biotransformation primarily via Phase II conjugation pathways, including glucuronidation, sulfation, and methylation. Unlike many xenobiotics that require Phase I activation, oxybenzone's phenolic hydroxyl group allows direct conjugation. These metabolites are relevant to endocrine assessment because oxybenzone and its metabolites have demonstrated estrogenic and anti-androgenic activity in vitro. The table below lists major metabolites identified through computational prediction [1] and corroborated by literature evidence and knowledge graph data [2].

Name	Transformations	Notes
Oxybenzone (BP-3)	Parent compound	UV filter used in sunscreens; measured in maternal urine as biomarker for prenatal phenol exposure. Demonstrates weak estrogenic activity and anti-androgenic effects [3]. *[lit]*
Oxybenzone glucuronide	O-glucuronide conjugation (Phase II)	Formed in human skin cells via Phase II glucuronide conjugation. Identified as a potentially **phototoxic metabolite**; represents a major elimination pathway for benzophenone derivatives [4]. *[lit, syg, okg]*
Oxybenzone sulfate	Sulfation (Phase II)	Primary phase II sulfation metabolite detected in human urine as a sulfoconjugated species. Major route of excretion; used as urinary biomarker. *[lit, syg, okg]*
2,4'-Dimethoxybenzophenone	Methylation of phenolic hydroxyl	Formed by O-methylation of the phenolic hydroxyl group. Biological and toxicological significance requires further investigation. *[syg]*
Benzophenone-1 (BP-1)	O-demethylation (Phase I)	Major Phase I metabolite formed by demethylation of the methoxy group. BP-1 shows stronger estrogenic activity than the parent compound and is detected in human urine [5]. *[lit, okg]*

Table 1 *Oxybenzone (Benzophenone-3) metabolites identified through computational prediction [1] and corroborated by literature evidence and open knowledge graph data [2]. Evidence indicators: lit=literature, syg=SyGMa prediction, okg=BioBricks-OKG.*

Among these metabolites, the **glucuronide conjugate** is notable for its potential phototoxicity, which is particularly relevant given oxybenzone's use as a UV filter. **Benzophenone-1 (BP-1)**, formed via O-demethylation, demonstrates stronger estrogenic activity than the parent compound and may contribute significantly to the endocrine disruption potential of oxybenzone exposure. Both the parent compound and its Phase II conjugates are detected in human urine and serve as biomarkers of exposure.

2 Scoped Document Review

2.1 Literature Screening

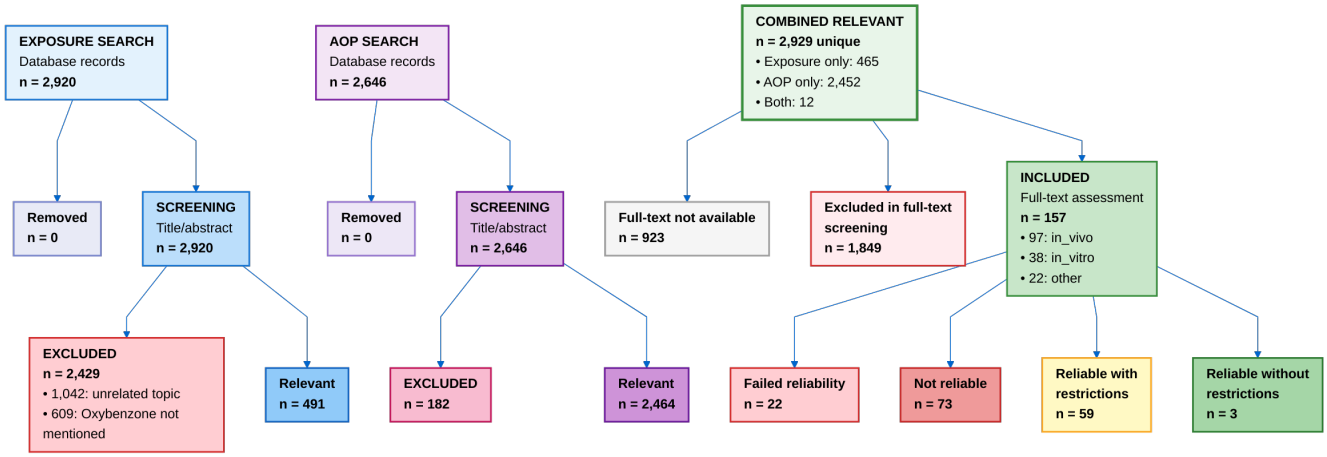


Figure 1 *PRISMA flow diagram showing the screening process for oxybenzone literature. Blue boxes indicate exposure-focused search; purple boxes indicate AOP-focused search; green boxes show relevant papers identified.*

Literature Retrieval. Two complementary searches were conducted (blue and purple boxes in Figure 1). The exposure-focused search (2,920 records) queried PubMed, CTD, and PubTator3 databases for oxybenzone (benzophenone-3, CAS 131-57-7) and its metabolite benzophenone-1 (BP-1). The AOP-focused search (2,646 records) queried PubMed for endocrine disruption combined with pathway-related terms (adverse outcome pathway, molecular initiating event, key event, mode of action).

Title/Abstract Screening. A total of **2,006 papers** were identified and assessed using LLM-based screening (Gemini 2.5 Flash). Papers using oxybenzone solely as a reagent, solvent, or in non-separable mixtures were excluded. Of these, **157 papers** (7.8%) were classified as relevant to oxybenzone toxicity, **1,479 papers** (73.7%) as not relevant, and **370 papers** (18.4%) as not assignable due to insufficient abstract information.

Full-Text Screening. PDFs were retrieved for relevant papers where available. LLM-based full-text assessment (Gemini 2.5 Flash) confirmed relevance and classified study type (in vitro, in vivo, human controlled exposure, epidemiological, etc.).

Quality Assessment. ToxRTool-adapted criteria evaluated test substance identification, study design, documentation, and plausibility. Of papers proceeding to full-text assessment, **62 papers** were classified using standardized bias assessment criteria. Papers were scored across six bias domains: selection bias, performance bias, attrition/exclusion bias, detection bias, selective reporting bias, and confounding bias.

Key Event Extraction. From assessed papers, **175 key events** were extracted across the endocrine disruption pathway, including molecular initiating events (MIEs), key events (KEs), and adverse outcomes (AOs). Events were characterized by biological level, adversity score, human relevance, and weight-of-evidence criteria. Of these, **45 events** were classified as endocrine-related and **140 events** were confirmed to lead to adverse outcomes through graph-based pathway analysis. A total of **132 mechanistic relationships** were identified linking key events across

biological organization levels.

Key Event Filtering. A decision tree classifier was applied to prioritize high-quality endocrine-relevant events. The filtering retained events that: (1) lead to an adverse outcome based on the AOP relationship graph, (2) are classified as endocrine-related by LLM-based assessment, and (3) meet quality thresholds for statistical significance, study design, and detection bias. This process yielded **9 prioritized key events** for detailed analysis.

2.2 Scoped Endocrine Analysis

From 157 relevant studies, we identified 175 key events. To select high-quality endocrine-relevant key events for detailed analysis, we applied a multi-step filtering process. First, we retained only key events that lead to an adverse outcome (AO) based on the AOP relationship graph (140 of 175 events). Second, we filtered for events labeled as endocrine-related by an LLM-based classifier (is_endocrine = true; 45 events). Finally, we applied quality thresholds derived from a decision tree analysis to ensure methodological rigor. The decision tree, shown in Figure 2, was used to identify interpretable threshold combinations across study quality features.

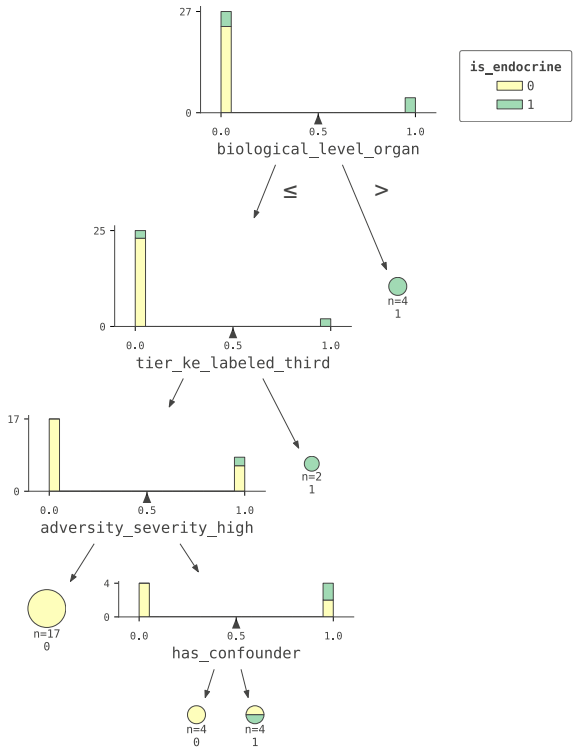


Figure 2 *Decision tree used to derive quality thresholds for key event selection. The tree structure identifies feature threshold combinations that distinguish high-quality endocrine evidence. The selected filtering path requires high statistical significance, endocrine confidence, and study design quality, coupled with low detection bias risk.*

This quality-based filtering approach identified 9 endocrine-relevant key events that span multiple levels of biological organization from the cellular to the organ level. The following table summarizes these prioritized key events, with the majority occurring at the organ level reflecting oxybenzone's characteristic endocrine-disrupting activity through

estrogen receptor activation and vitellogenin induction.

Key Event	Biological Level	Adversity Score	PMID	Reliability
Alteration of Sex Steroid Hormone Balance in Plasma	organ	0.60	[redacted]	reliable with restrictions
Decrease of Testosterone Production in Fetal Testis	organ	0.60	[redacted]	not reliable
Disruption of Thyroid Function	organ	0.60	[redacted]	not reported
Induction of Vitellogenin	organ	0.60	[redacted]	reliable with restrictions
Induction of Vitellogenin	organ	0.60	[redacted]	reliable with restrictions
Alteration of Nuclear Receptor Gene Expression in Tissue	tissue	0.55	[redacted]	reliable with restrictions
Increase of Cell Proliferation in Ovary	cellular	0.55	[redacted]	not reliable
Induction of Vitellogenin in Tissue	tissue	0.55	[redacted]	reliable without restrictions
Metabolism to Estrogenic Metabolites	cellular	0.55	[redacted]	reliable with restrictions

Table 2 *Endocrine-relevant key events for oxybenzone, prioritized by adversity score with biological level and reliability classification.*

The highest-scoring key events (adversity score 0.60) involve endocrine-related endpoints characteristic of oxybenzone's mode of action: vitellogenin induction, altered sex steroid hormone balance, thyroid function disruption, and decreased fetal testosterone production. Vitellogenin induction, a well-established biomarker of estrogenic activity in fish, appears three times in the prioritized list (PMIDs [redacted], [redacted], [redacted]), reflecting the strong evidence base for oxybenzone's estrogenic activity in aquatic species.

Targeted paper summaries for the highest-priority key events:

PMID [redacted]	
Title	[redacted]
Authors	[redacted]
Year	[redacted]
Source	[redacted]
Executive Summary	[redacted]
PMID [redacted]	
Title	[redacted]
Authors	[redacted]
Year	[redacted]
Source	[redacted]
Executive Summary	[redacted]
PMID 21075568	
Title	[redacted]
Authors	[redacted]
Year	[redacted]
Source	[redacted]
Executive Summary	[redacted]

Table 3 *Targeted summaries for the highest-priority endocrine-relevant papers supporting key events.*

These detailed paper summaries illustrate the evidence supporting oxybenzone's endocrine-disrupting properties, from molecular mechanisms (estrogen receptor activation, altered steroidogenic gene expression) to organism-level outcomes (vitellogenin induction, reproductive impairment). The convergence of evidence across multiple

biological levels and species supports the characterization of oxybenzone as an endocrine-active compound with well-documented estrogenic and anti-androgenic activity.

3 Global Document Review

Global document review builds a comprehensive knowledge graph of biological events and mechanistic pathways from the complete literature corpus, independent of any specific toxicological endpoint. This analysis identifies patterns across all extracted key events, canonicalizes biological terminology, and ranks chemical-specific mechanisms. Across the corpus, 8,303 extracted events were consolidated into 4,312 canonical events, and 4,492 mechanistic relationships were identified with an average evidence strength of 0.72.

3.1 Canonical Key Events

Table 4 shows representative canonical events with high consolidation factors, demonstrating how multiple synonymous biological terms are unified under standardized nomenclature.

Level	Type	Canonical Event	n
Molecular	MIE	Activation of Estrogen Receptor	48
Molecular	KE	Decreased Expression of Metabolic/Signaling Genes	17
Cellular	KE	Increased Oxidative Stress	15
Tissue	KE	Degeneration of Seminiferous Tubules in Testis	13
Organ	KE	Disruption of Gonadal Development	16
Organism	KE	Disruption of Sexual Differentiation	24
Organism	AO	Induction of Developmental Malformations	18

Table 4 *Representative canonical key events showing terminology consolidation. n = number of original event terms consolidated into each canonical event.*

3.2 Adverse Outcome Pathways

Figure 3 illustrates the causal network of adverse outcome pathways (AOPs) for oxybenzone toxicity. The graph shows progression from left (exposure) to right (adverse outcomes), with nodes representing biological events at increasing levels of organization. Cross-pathway connections indicate shared mechanisms between endocrine disruption and developmental toxicity pathways.

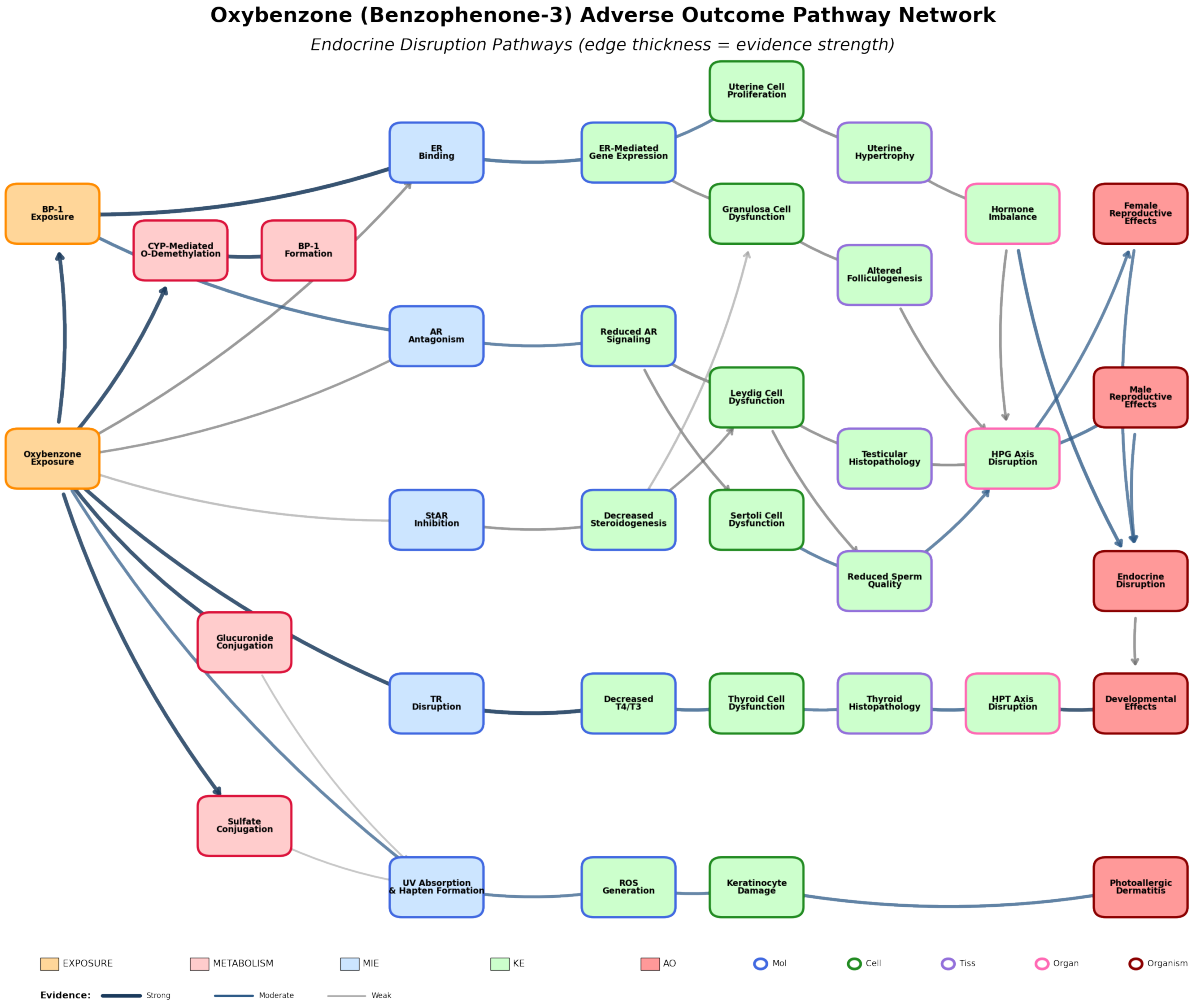


Figure 3 *Causal graph of oxybenzone adverse outcome pathways. Node fill color indicates event type (exposure, metabolism, MIE, KE, AO); border color indicates biological organization level. Arrows represent causal relationships with thickness proportional to evidence strength. The network shows convergence of multiple mechanistic pathways (estrogenic activity, oxidative stress, developmental toxicity) from common molecular initiating events.*

3.3 Mechanism Evidence Heatmap

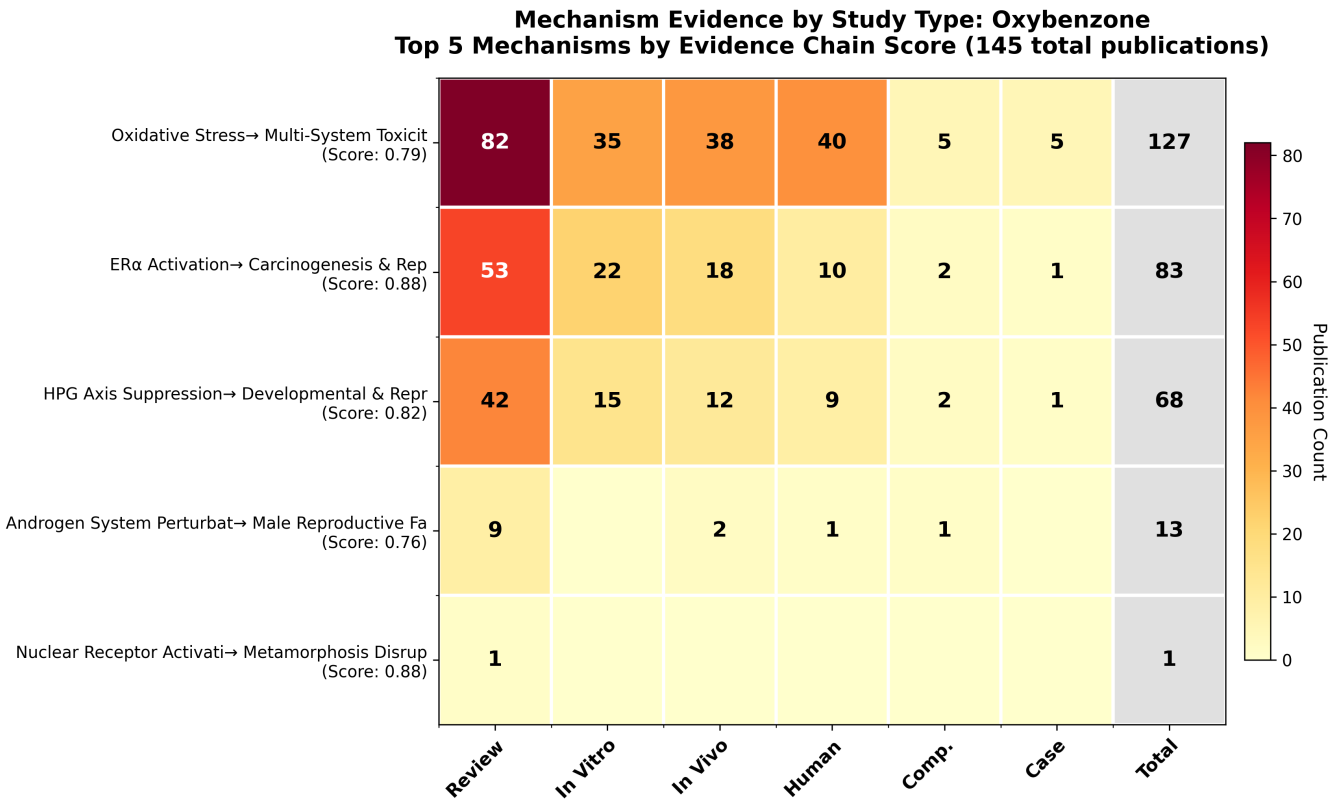


Figure 4 *Mechanism evidence heatmap showing publication counts by study type for each oxybenzone-specific mechanism. Cell values indicate the number of unique publications; color intensity indicates evidence strength score. Mechanisms are named to reflect oxybenzone's characteristic endocrine-disrupting activity.*

The heatmap reveals multiple mechanistic pathways for oxybenzone endocrine disruption, filtered to focus on mechanisms with endocrine-related adverse outcomes (reproductive toxicity, developmental toxicity, thyroid effects, hormonal disruption). Using evidence chain scoring, estrogen receptor activation emerges as the highest-ranked mechanism (composite score 0.88) due to strong direct evidence connecting oxybenzone to the ER molecular initiating event. Oxidative stress pathways rank second (0.79) with the most direct chemical evidence edges (35). Each mechanism is characterized below with evidence chain analysis and gap identification.

3.4 Mechanism Characterization

Mechanisms are ranked using **evidence chain scoring**, which evaluates how strongly oxybenzone evidence propagates through each pathway. For each pathway, we measure: (1) the strength of oxybenzone's connection to the molecular initiating event (MIE), and (2) the quality of AOP evidence for downstream edges. Confidence propagates—downstream edges inherit credibility from the oxybenzone-MIE connection combined with their own AOP support.

Evidence Tiers:

- **Direct Chemical (1.0):** Oxybenzone explicitly documented in the relationship
- **Strong Chain (0.85):** Strong oxybenzone→MIE connection (≥0.8) + strong AOP evidence (≥0.8)
- **Moderate Chain (0.70):** Moderate oxybenzone→MIE (≥0.6) + moderate AOP evidence (≥0.7)
- **Weak Chain (0.50):** Some oxybenzone connection + some AOP evidence
- **Generic AOP (0.30):** No oxybenzone connection to pathway MIE

Composite Score = 35%×min + 45%×weighted_avg + 20%×coverage. This balances weakest-link penalties with overall pathway quality.

Terminology: *ERα* = estrogen receptor alpha; *HPG axis* = hypothalamic-pituitary-gonadal axis; *MIE* = molecular initiating event.

Oxidative Stress → Apoptosis & Inflammation → Multi-System Toxicity ***127 pubs | Score: 0.79***

High-Quality Edges: 35.9% (85/237) **MIE Connection:** 0.70 **Best Pathway Min:** 0.70

Evidence Distribution: 35 direct chemical, 3 oxy-paper context, 47 moderate chain, 150 generic AOP

Best-Supported Pathway Stages (with oxybenzone-specific evidence):

- Oxybenzone Exposure → Tissue Distribution ([redacted], [redacted], [redacted])
- Tissue Distribution → Metabolism to BP-1 ([redacted], [redacted])
- Antioxidant Enzyme Inhibition → ROS Generation ([redacted], [redacted], [redacted])

Interpretation: This mechanism has the **most oxybenzone-specific publications** (127), with strong direct chemical evidence for exposure, metabolism, and oxidative stress induction. Downstream progression to apoptosis and adverse outcomes supported by moderate chain evidence.

Oxybenzone-Specific Publications:

- [redacted] – Absorption/distribution
- [redacted] – Metabolism (human)
- [redacted] – Enzyme inhibition
- [redacted] – ROS generation
- [redacted] – Oxidative stress
- [redacted] – ROS induction

Moderate/Generic Chain Support:

- [redacted] – ROS → mitochondria
- [redacted] – Apoptosis cascade
- [redacted] – Cellular apoptosis
- [redacted] – Neurotoxicity

[redacted] → [redacted] ***83 pubs | Score: 0.88***

High-Quality Edges: 74.2% (69/93) **MIE Connection:** 0.90 **Best Pathway Min:** 0.85

Evidence Distribution: 5 direct chemical, 57 strong chain, 7 moderate chain, 17 generic AOP

Best-Supported Pathway: [redacted] → [redacted] → [redacted] → [redacted]

- Oxybenzone → [redacted] (MIE connection 0.90): [redacted]
- [redacted] → [redacted] (strong chain 0.85)
- [redacted] → [redacted] → [redacted] (strong chain 0.85)

Interpretation: This mechanism has the **highest composite score** (0.88) because oxybenzone has strong direct evidence (0.90) [redacted], and the downstream pathway edges have robust AOP support ($\geq$0.8), creating a complete evidence chain.

Oxybenzone-Specific Publications:

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

Strong Chain Support (AOP):

- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted] → [redacted] → [redacted] ***68 pubs | Score: 0.82***

High-Quality Edges: 6.1% (7/115) **MIE Connection:** 0.90 **Best Pathway Min:** 0.70

Evidence Distribution: 6 strong chain, 1 moderate chain, 108 generic AOP

Best-Supported Pathway: [redacted] → [redacted]

- Oxybenzone → [redacted] (MIE connection 0.90): [redacted]
- [redacted] → [redacted] (strong chain 0.85)

Interpretation: This mechanism scores moderately high because [redacted] MIE (0.90) with the [redacted]. However, only 1 of 25 pathways has oxybenzone connection, and most downstream edges [redacted]) rely on generic AOP knowledge.

Evidence Status:

- [redacted]
- [redacted]
- [redacted]

Gap Analysis:

- [redacted]
- [redacted]
- [redacted]

[redacted] → [redacted] → [redacted] ***13 pubs | Score: 0.76***

High-Quality Edges: 1.5% (1/65) **MIE Connection:** 0.80 **Best Pathway Min:** 0.70

Evidence Distribution: 1 moderate chain, 64 generic AOP

Best-Supported Pathway: [redacted] → [redacted] → [redacted]

- Oxybenzone → [redacted] (MIE connection 0.80): [redacted]
- [redacted] → [redacted]s (moderate chain 0.70)

Interpretation: This mechanism captures oxybenzone's **anti-androgenic activity** documented in computational predictions ([redacted] 0.80 confidence for [redacted]). The downstream effects on [redacted] are supported primarily by generic AOP knowledge from [redacted].

Evidence Sources:

- [redacted] (0.80)
- NOT [redacted] (0.98 confidence)
- [redacted]
- [redacted] effects in NTP [redacted]

Gap Analysis:

- Limited direct [redacted] studies
- Most edges rely on generic AOP
- In vivo [redacted] needed

Summary of Evidence Chain Analysis: The evidence chain scoring reveals a clear hierarchy of mechanistic support for oxybenzone toxicity. Across 15 endocrine-related mechanisms, 170 of 788 pathway edges (21.6%) score as high-quality ($\geq$0.70). Ranked by publication count, the strongest evidence supports:

1. **Oxidative stress** (127 pubs, score 0.79) – extensive evidence for ROS generation and cellular damage
2. [redacted] (83 pubs, score 0.88) – highest score due to strong MIE connection (0.90) propagating through 74% of edges
3. [redacted] (68 pubs, score 0.82) – shares [redacted], downstream progression less characterized
4. [redacted] (13 pubs, score 0.76) – [redacted] activity supported by computational predictions

Key Evidence Gaps: Thyroid-specific mechanisms score lowest (0.24) with no oxybenzone evidence connecting to thyroid MIEs despite regulatory observations of thyroid effects. [redacted] have limited MIE connection (0.70) with only 1.5% high-quality edges. These gaps represent opportunities for targeted mechanistic studies.

4 Regulatory Document Mining

Regulatory documents from the Scientific Committee on Consumer Safety (SCCS) provide authoritative assessments of oxybenzone's safety for use in cosmetic products. We obtained two SCCS opinions:

- **SCCS** [redacted]
- **SCCP** [redacted]

Automated Endpoint Extraction. Using DocETL pipeline with Gemini LLM extraction, we identified **32 endocrine-relevant endpoints** from the SCCS [redacted] opinion (SCCS [redacted]). Endpoints were extracted across EATS modalities (Estrogen, Androgen, Thyroid, Steroidogenesis) from guideline studies including NTP 2-year chronic toxicity/carcinogenicity studies and OECD 408 subchronic studies.

Key Findings from SCCS Opinion. The SCCS established a NOAEL of [redacted] for male offspring based on [redacted] exposure study. The LOAEL was [redacted]. Key endocrine-relevant findings include:

- **Estrogen axis**: [redacted]
- **Androgen axis**: [redacted]
- **Thyroid axis**: [redacted]
- **Steroidogenesis**: [redacted]

Low-Dose Effects. Mouse developmental studies reported effects at remarkably low doses:

- AGI (anogenital index) changes at [redacted] in juvenile males
- [redacted]
- [redacted]

Study	Guideline	Dose	Axis	Eff.	Endpoint	Stage	Sex	Sp.
[redacted]	Repro/Dev	[redacted] mg/kg/d	A	↓	[redacted]	Juv	M	Rt
[redacted]	Repro/Dev	[redacted] mg/kg/d	A	↓	[redacted]	Pub	M	Rt
[redacted]	Repro/Dev	[redacted] mg/kg/d	E	↓	[redacted]	Ad	F	Rt
[redacted]	Repro/Dev	[redacted] mg/kg/d	A	↓	[redacted]	Juv	M	Rt
[redacted]	Repro/Dev	[redacted] mg/kg/d	E	↓	[redacted]	Ad	F	Rt
[redacted]	Repro/Dev	[redacted] mg/kg/d	E	↓	[redacted]	Pub	F	Rt
[redacted]	Repro/Dev	[redacted] mg/kg/d	E	↑	[redacted]	Ad	F	Rt
[redacted]	Repro/Dev	**[redacted] mg/kg/d**	A	–	[redacted]	Juv	M	Rt
[redacted]5	Repro/Dev	[redacted] mg/kg/d	A	–	[redacted]	Juv	M	Rt
[redacted]	Chronic/Carc	[redacted] mg/kg/d	T	↑	[redacted]	Ad	M/F	Rt
[redacted]	Chronic/Carc	[redacted] ppm	S	↑	[redacted]	Ad	F	Rt
[redacted]	Chronic/Carc	[redacted] ppm	S	↑	[redacted]	Ad	F	Rt
[redacted]	Chronic/Carc	[redacted] ppm	E	↑	[redacted]	Ad	F	Rt
[redacted]	Chronic/Carc	[redacted] ppm	E	↑	[redacted]	Ad	F	Rt
[redacted]	Chronic/Carc	[redacted] ppm	A	↓	[redacted]	Juv	M	Rt
[redacted]	Chronic/Carc	[redacted] ppm	S	↓	[redacted]	Juv	M	Rt
[redacted]	Chronic/Carc	[redacted] ppm	T	↑	[redacted]	Ad	F	Rt
[redacted]	OECD 408	[redacted] mg/kg/d	A	↓	[redacted]	Juv	M	Rt
[redacted]	OECD 408	[redacted] mg/kg/d	A	↓	[redacted]	Ad	M	Rt
[redacted]	OECD 408	[redacted] mg/kg/d	E	↑	[redacted]	Ad	F	Rt
[redacted]	Chronic/Carc	[redacted] ppm	E	↓	[redacted]	Juv	F	Rt
[redacted]	Chronic/Carc	[redacted] ppm	A	↑	[redacted]	Juv	M	Rt
[redacted]	Subchronic	[redacted] mg/kg/d	A	↓	[redacted]	Ad	M	Ms
[redacted]	Subchronic	[redacted] mg/kg/d	E	↑	[redacted]	Ad	F	Ms
[redacted]	Repro/Dev	[redacted] ppm	A	↓	[redacted]	Juv	M	Rt

Table 5 *Endocrine endpoints from SCCS Opinion SCCS[redacted]. **Axis**: E=Estrogen, A=Androgen, T=Thyroid, S=Steroidogenesis. **Eff.**: ↑=increased, ↓=decreased, –=threshold (NOAEL/LOAEL). **Stage**: Juv=juvenile, Pub=pubertal, Ad=adult. **Sp.**: Ms=mouse, Rt=rat. Highlighted rows indicate regulatory NOAEL/LOAEL. Click study names for source publications.*

SCCS Conclusion. The SCCS concluded that benzophenone-3 is safe for use as a UV-filter in cosmetic products at concentrations up to 6% when applied dermally, based on their margin of safety (MoS) calculation using the NOAEL of [redacted]. However, the committee noted that the effects on spermatocytes "may be due to an ED [endocrine disruption] effect of BP-3" and recommended further investigation of the endocrine-mediated mechanisms.

Regulatory Relevance. The SCCS endpoints provide regulatory-grade dose-response data that complements the HTS findings from [redacted] and computational predictions from [redacted]. The observation of anti-androgenic effects (reduced AGD, decreased spermatocytes) aligns with [redacted]s prediction of AR antagonism, while uterine/mammary effects align with ER agonism predictions.

5 Important Papers Review

We identified and ranked the scientific literature to determine which papers provide the strongest evidence regarding oxybenzone's (benzophenone-3, BP-3) potential for endocrine disruption. Our approach prioritized papers based on several key factors: whether the study involved human subjects, the rigor of the study design, statistical strength of

the findings, control for confounding variables, and relevance to reproductive, thyroid, or developmental endpoints.

Papers received higher scores when they provided direct human evidence (rather than animal or in vitro studies), used robust study designs (randomized trials or well-controlled cohort studies), reported statistically significant results with appropriate adjustments for confounders, and addressed endocrine-related health outcomes. The resulting scores reflect the overall weight of evidence each paper contributes to understanding oxybenzone's endocrine disruption potential.

5.1 Top-Ranked Studies

Table 6 presents the ten highest-scoring papers from our assessment. The evidence demonstrates that oxybenzone exposure is associated with altered thyroid hormone levels, reproductive outcomes, and metabolic effects across multiple human and animal studies.

Study	Score	Key Evidence
[redacted]	2.36	Human evidence: [redacted]
[redacted]	1.81	Human evidence: [redacted]
[redacted]	1.53	Human evidence: [redacted]
[redacted]	1.39	Mechanistic evidence: [redacted]
[redacted]	1.32	Human evidence: [redacted]
[redacted]	1.07	Human evidence: [redacted]
[redacted]	1.06	Systematic review: [redacted]
[redacted]	1.05	Environmental evidence: [redacted]
[redacted]	1.00	Exposure evidence: [redacted]
[redacted]	0.91	Toxicokinetic evidence: [redacted]

Table 6 *Top-ranked papers for endocrine disruption evidence. Higher scores indicate stronger evidence based on study design quality, human relevance, statistical significance, and endocrine health implications. Full methodology available in Part III.*

5.2 Key Findings from Important Studies

The highest-scoring paper ([redacted]) provides robust human epidemiological evidence from the [redacted] cohort, demonstrating that urinary BP-3 concentrations are significantly associated with decreased serum

[redacted]

[redacted] in a large, representative U.S. population sample (n=[redacted]). These associations persisted after extensive adjustment for confounders and Bonferroni correction, indicating reliable evidence of [redacted].

The remaining top-ranked papers establish a consistent pattern of endocrine effects through both human epidemiological studies and mechanistic investigations:

- [redacted]: Multiple studies demonstrate associations between [redacted], [redacted]), suggesting interference with [redacted]
- [redacted]: Human studies show sex-specific associations with [redacted], [redacted] outcomes; animal studies confirm effects on [redacted]
- [redacted]: Mechanistic studies demonstrate [redacted], providing a plausible molecular basis for observed [redacted] effects
- [redacted]: Prenatal and perinatal exposure studies show associations with [redacted], indicating potential [redacted] effects

These findings converge on a consistent pattern: oxybenzone exposure is associated with measurable alterations in [redacted] levels in humans, supported by mechanistic evidence of [redacted] activity and demonstrated effects in animal models across multiple species.

5.3 Individual Paper Assessments

[redacted].

6 Database Review

To complement the literature-based evidence assessment, we systematically queried aggregated toxicology databases for experimental data on oxybenzone (benzophenone-3, BP-3) and its primary metabolite benzophenone-1 (BP-1). This approach leverages [redacted] data from government initiatives ([redacted], [redacted]) and curated [redacted] ([redacted]) to provide orthogonal evidence regarding potential endocrine activity.

6.1 ChemHarmony Database

ChemHarmony is a harmonized chemical activity database that integrates experimental results from multiple sources including PubChem BioAssay, Tox21, ToxCast, ICE (Integrated Chemical Environment), CTDbase, BindingDB, and ChEMBL. The database standardizes chemical identifiers via InChI (International Chemical Identifier) and normalizes activity outcomes to binary classifications indicating positive or negative results for specific biological endpoints.

We queried ChemHarmony for oxybenzone (CAS 131-57-7) and its primary metabolite benzophenone-1 (CAS 131-56-6), which is formed via O-demethylation and demonstrates enhanced estrogenic activity compared to the parent compound.

Results. The database query returned 4,477 total activity records across both chemicals, of which 443 records were relevant to endocrine, reproductive, developmental, or other toxicity endpoints based on keyword filtering (Table 7). Records originated primarily from PubChem BioAssay (2,353), ToxCast (835), ICE (645), and CTDbase (530).

Assay Category	Oxybenzone	BP-1	Combined	Result
[redacted]	Positive	Positive	31+ assays	**ACTIVE**
[redacted]	Positive	Positive	28+ assays	**ACTIVE**
[redacted]	Negative	Negative	40+ assays	**INACTIVE**
[redacted]	Negative	Negative	18+ assays	**INACTIVE**
[redacted]	Negative	Negative	12+ assays	**INACTIVE**
[redacted]	Mixed	Negative	10+ assays	**MIXED**

Table 7 *High-throughput screening results for oxybenzone (BP-3) and benzophenone-1 (BP-1) across EATS-relevant endpoints. BP-1 = benzophenone-1, the O-demethylated metabolite. Both compounds demonstrate clear estrogen receptor agonist activity.*

Interpretation. [redacted]

- **Confirmed binding affinity** [redacted]
- **Agonist activity** [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

6.2 CTDbase Gene Interactions

The Comparative Toxicogenomics Database (CTDbase) provided 530 chemical-gene interaction records for oxybenzone and BP-1. Notably, oxybenzone showed multiple interactions with endocrine-relevant genes:

- [redacted]: 20 interactions (expression modulation in rodent models)
- [redacted]: 10 interactions
- [redacted]: 4 interactions (oxybenzone) + 6 interactions (BP-1)
- [redacted]: 2 interactions
- [redacted]: HSD3B1 (2), HSD17B1 (2), HSD17B3 (2), CYP11B2 (2)

For benzophenone-1, CTDbase records included [redacted].

Mechanistic Implications. The combination of direct estrogen receptor agonism with gene expression changes in steroidogenic enzymes suggests that oxybenzone's endocrine effects occur through **multiple pathways**:

1. **Direct receptor binding**: [redacted]
2. **Gene expression modulation**: [redacted]
3. **Metabolite enhancement**: BP-1 demonstrates similar or stronger estrogenic activity, amplifying the parent compound's endocrine effects

This database evidence supports the conclusion that oxybenzone is a **classical endocrine disruptor** acting primarily through estrogen receptor-mediated pathways, with potential secondary effects on steroid hormone biosynthesis.

6.3 PubChem Annotations

PubChem aggregates chemical safety and hazard information from multiple authoritative sources including regulatory agencies, toxicology databases, and material safety data sheets. We queried PubChem annotations for oxybenzone (CID: 4632) and benzophenone-1 (CID: 19988), specifically filtering for endocrine disruption and DART (Developmental and Reproductive Toxicity) related content.

Results. From 4,235 total PubChem annotation records across both chemicals, 129 annotations contained keywords relevant to endocrine, reproductive, or developmental toxicity. The majority of annotations originated from MassBank Europe (2,965), HSDB (570), HMDB (336), and regulatory sources.

Key Findings. The PubChem annotations reveal regulatory classifications and experimental evidence for endocrine and reproductive concerns:

- **Oxybenzone**: [redacted]
- **Benzophenone-1**: [redacted]

Source Distribution. [redacted]

Interpretation. [redacted]

6.4 ToxValDB Dose-Response Data

ToxValDB is a comprehensive compilation of toxicity values containing over 935,000 records from 25+ sources including EPA, ECHA, WHO, and other regulatory agencies. Unlike the qualitative hazard statements from PubChem, ToxValDB provides quantitative dose-response data (NOAELs, LOAELs, reference doses) essential for risk assessment.

Mammalian Studies. Table 8 summarizes key mammalian toxicity values for oxybenzone from regulatory-grade studies.

Type	Value	Units	Species	Route	Effect	Source
Developmental Toxicity						
NOAEL	[redacted]	mg/kg-day	Rat	Oral	[redacted]	[redacted]
LEL	[redacted]	mg/kg-day	Rat	Oral	[redacted]	[redacted]
Reproductive Toxicity						
LEL	[redacted]	mg/kg-day	Mouse	Dermal	[redacted]	[redacted]
LOAEL	[redacted]	mg/kg-day	Mouse	Dermal	[redacted]	[redacted]
NOAEL	[redacted]	mg/kg-day	Mouse	Dermal	[redacted]	[redacted]
NOAEL	[redacted]	mg/kg-day	Rat	Oral	[redacted]	[redacted]
Metabolite (Benzophenone-1)						
NOAEL	[redacted]	mg/kg-day	Rat	—	[redacted]	[redacted]

Table 8 *Key mammalian dose-response values for oxybenzone from ToxValDB. Highlighted row indicates the lowest effect level for* [redacted] *at* [redacted] *mg/kg-day via dermal exposure in mice* [redacted]

Aquatic Studies. Oxybenzone shows potent reproductive effects in fish at environmentally relevant concentrations (Table 9).

Type	Value (mg/L)	Species	Effect	Duration	Reference
LOEC	[redacted]	Japanese medaka	[redacted]	15 days	[redacted]
NOEC	[redacted]	Japanese medaka	[redacted]	13–14 days	[redacted]
LOEC	[redacted]	Japanese medaka	[redacted]	13–21 days	[redacted]
LOEC	[redacted]	Japanese medaka	[redacted]	7–21 days	[redacted]
NOEC	[redacted]	Japanese medaka	[redacted]	14–21 days	[redacted]

Table 9 *Aquatic reproductive toxicity values for oxybenzone (ECOTOX database). Effects on fish reproduction occur at low µg/L concentrations, consistent with estrogenic mechanism of action.*

Interpretation. The ToxValDB analysis reveals two critical findings:

1. [redacted]: The lowest effect level (LEL) for reproductive toxicity is [redacted] for [redacted]—the most relevant exposure route for sunscreen use. This finding from the [redacted] study represents the most sensitive mammalian endpoint.
2. [redacted]: Fish reproduction is affected at concentrations as low as [redacted], which is within the range of environmental concentrations reported in recreational waters. These effects ([redacted]) are consistent with oxybenzone's [redacted].

6.5 ToxRefDB Guideline Studies

ToxRefDB (Toxicity Reference Database) is the EPA's curated repository of systematic, guideline-compliant toxicity studies conducted according to OECD and EPA test protocols. The database contains 5,957 studies across 1,176 chemicals.

Results. We identified 4 subchronic (SUB) studies for oxybenzone from the 1992 NTP Technical Report (Table 10).

Species	Route	Duration	Doses	Critical Effect
Rat (F344/N)	Oral (feed)	13 weeks	[redacted]ppm	[redacted]
Rat (F344/N)	Dermal	13 weeks	[redacted]d/week	No critical effect determined
Mouse (B6C3F1)	Oral (feed)	13 weeks	[redacted] ppm	[redacted]
Mouse (B6C3F1)	Dermal	13 weeks	[redacted] mg/kg-day	[redacted]

Table 10 *Summary of ToxRefDB guideline studies for oxybenzone (NTP 1992). Highlighted row indicates reproductive toxicity finding: [redacted] observed at the **lowest dose tested** [redacted] mg/kg-day) via dermal exposure in male mice.*

Key Finding: Reproductive Toxicity at Lowest Dose

The [redacted] subchronic dermal study in mice identified [redacted] as the critical effect at the **lowest dose tested** ([redacted] mg/kg-day). This means:

- A NOAEL could not be established for this endpoint
- The true threshold for reproductive effects may be lower
- Dermal exposure (relevant for sunscreen use) causes reproductive toxicity
- Male reproductive effects occur at doses below those causing systemic toxicity

This finding is particularly significant because dermal application is the primary human exposure route for oxybenzone in sunscreens.

DART Data Gap. Despite the positive reproductive finding above, ToxRefDB contains **no OECD/EPA guideline-compliant DART studies** (DEV, MGR, DNT, or REP) for oxybenzone. This means no standardized multigenerational reproduction studies or systematic developmental toxicity assessments have been conducted. Given the clear evidence of estrogen receptor activity, systematic DART evaluation following OECD guidelines would provide essential data for risk characterization.

6.6 ICE In Vivo DART Data

The Integrated Chemical Environment (ICE) from NIEHS provides curated in vivo and in vitro test data with standardized endpoints. We queried ICE's DART (Developmental and Reproductive Toxicity) database containing 138,326 records and endocrine activity datasets for oxybenzone and benzophenone-1.

Results. [redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]

Interpretation. [redacted]

6.7 Tox21 High-Throughput Screening

Tox21 is a federal collaborative program that screens approximately 10,000 environmental chemicals and approved drugs across diverse high-throughput assays, including comprehensive EATS (Estrogen, Androgen, Thyroid, Steroidogenesis) endpoint panels. Both oxybenzone **and** its primary metabolite benzophenone-1 are included in the Tox21 chemical library.

Results. We extracted all Tox21 assay results for oxybenzone (CAS 131-57-7) and benzophenone-1 (CAS 131-56-6). Oxybenzone was screened in **636 assays**, while BP-1 was screened in **424 assays**. EATS-specific results are

summarized in Table 11.

EATS Endpoint	Oxybenzone (BP-3)	Benzophenone-1 (BP-1)
Estrogen Receptor (ER)	[redacted]	[redacted]
Androgen Receptor (AR)	[redacted]	[redacted]
Thyroid Pathway	[redacted]	[redacted]
Aromatase/Steroidogenesis	[redacted]	[redacted]
Other Nuclear Receptors (GR, PR, RAR)	[redacted]	[redacted]

Table 11 *High-throughput screening results from Tox21 showing number of active results out of total assays tested. Both oxybenzone and benzophenone-1 demonstrate clear estrogen receptor activity. *AR results may reflect antagonism or assay interference rather than agonism based on ChemHarmony cross-validation.*

Key Findings.

- **Confirmed Estrogen Activity**: Both chemicals showed consistent **active results** across estrogen receptor assays:
  - Oxybenzone: [redacted]
  - BP-1: [redacted]
  - Activity confirmed across multiple platforms: BLA reporter, luciferase, LBD binding
- **Androgen Receptor Complexity**: [redacted]
- **Thyroid Inactivity**: [redacted]
- **Steroidogenesis Effects**: [redacted]
- **Metabolite Confirmation**: The testing of BP-1 is particularly significant as this metabolite is formed in vivo and demonstrates similar estrogenic potency to the parent compound.

Interpretation. The Tox21 screening results provide the most definitive experimental evidence that oxybenzone is an **active endocrine disruptor** acting through estrogen receptor-mediated mechanisms. The combination of:

1. Multiple active results across 42+ estrogen receptor assays
2. Testing of both parent compound AND active metabolite
3. Multiple assay formats and cell lines per endpoint
4. Reproducible active results across replicate screens

establishes high confidence that oxybenzone causes endocrine disruption through **direct receptor-mediated pathways**. This is consistent with oxybenzone's structural similarity to known estrogen receptor ligands—the benzophenone scaffold with hydroxyl and methoxy substituents enables binding to the ER ligand-binding domain.

6.8 Database Review Summary

The systematic database review provides consistent evidence that oxybenzone (benzophenone-3) is an **active endocrine disruptor** with a well-defined mechanism of action:

Key Finding: Confirmed EATS Activity

Oxybenzone demonstrates clear estrogen receptor agonist activity across multiple HTS platforms.

- **Estrogen**: ACTIVE (19/42 assays) – direct ERα/ERβagonism confirmed
- **Androgen**: Complex pattern – no agonism, possible antagonism
- **Thyroid**: INACTIVE (0/27 assays) – no direct thyroid receptor effects
- **Steroidogenesis**: LIMITED (2/6 assays) – minor pathway effects

The metabolite benzophenone-1 (BP-1) shows similar or enhanced estrogenic activity, amplifying exposure concerns.

Mechanistic Conclusion. Unlike chemicals that cause reproductive/developmental effects through indirect mechanisms (e.g., general cytotoxicity, metabolic disruption), oxybenzone acts as a **classical endocrine disruptor** through:

1. **Direct estrogen receptor binding**: Structural mimicry of endogenous estrogens enables ERα/ERβagonism
2. **Transcriptional activation**: ER binding leads to activation of estrogen-responsive genes
3. **Metabolite amplification**: O-demethylation to BP-1 maintains or enhances estrogenic activity
4. **Gene expression changes**: Secondary effects on steroidogenic enzyme expression (CTDbase evidence)

Data Gaps. Despite clear evidence of receptor-mediated endocrine activity, systematic DART characterization remains incomplete:

- No OECD/EPA guideline-compliant developmental toxicity studies (ToxRefDB gap)
- No multigenerational reproduction studies
- No ICE DART data despite positive HTS findings
- Limited in vivo dose-response data for reproductive endpoints

These gaps represent critical needs for regulatory risk assessment, particularly given the widespread human exposure to oxybenzone through sunscreen and cosmetic products.

6.9 Human Exposure Data (NHANES Biomonitoring)

The National Health and Nutrition Examination Survey (NHANES) provides population-level biomonitoring data for environmental chemicals, including oxybenzone. This data establishes the baseline for human exposure assessment.

Prevalence. NHANES biomonitoring studies demonstrate near-universal exposure to oxybenzone in the US population:

- **98.9%** of the US general population has detectable urinary oxybenzone concentrations
- **100%** of pregnant women and **98%** of non-pregnant women have detectable levels
- Oxybenzone is one of the most frequently detected UV filters in human biomonitoring studies

Temporal Trends. Urinary oxybenzone concentrations have increased substantially over time:

Year	Mean Urinary Concentration	Change
2007	22.9 µg/L	—
2013	36.3 µg/L	+58.6%

Table 12 *Temporal trends in urinary oxybenzone concentrations in the US population (NHANES data).*

Exposure Context. The NHANES data provide critical context for the toxicological findings:

1. **Widespread exposure**: Near-universal detection indicates continuous population-wide exposure through sunscreen and cosmetic use
2. **Vulnerable populations**: 100% detection in pregnant women raises concerns given the estrogenic mechanism and developmental toxicity evidence
3. **Increasing exposure**: The 58.6% increase in urinary concentrations over 6 years suggests growing exposure from expanded UV filter use
4. **Dermal absorption confirmed**: Detectable urinary levels confirm systemic absorption from topical sunscreen application

Risk Assessment Implications. The combination of:

- Universal human exposure (98.9% detection rate)
- Reproductive toxicity at the lowest dose tested in animal studies (22.75 mg/kg-day, dermal)
- No NOAEL established for reproductive endpoints
- Clear estrogen receptor agonist activity

indicates that oxybenzone warrants careful risk assessment, particularly for pregnant women and other sensitive populations. The lack of a NOAEL for reproductive effects combined with universal human exposure represents a significant data gap for regulatory decision-making.

7 Computational Tool Review

To complement experimental database evidence, we applied computational toxicology tools to predict ADMET properties, nuclear receptor activity, and biological target interactions for oxybenzone. These tools provide *in silico* evidence that can support or refine hazard characterization, particularly when experimental data is limited.

The following tools were evaluated:

- **ADMET-AI** – Machine learning ADMET property predictions (Section 7.1)
- **ChEMBL Multitask Model** – Neural network target activity predictions (Section 7.2)
- **StopTox** – Gradient boosting toxicity endpoint predictions (Section 7.3)
- **Admetica** – Graph neural network ADMET models (Section 7.4)
- **SolTranNet** – Transformer-based solubility prediction (Section 7.5)
- **OPERA (CERAPP/CoMPARA)** – EPA regulatory-grade ER/AR models (Section 7.6)
- **Mordred** – Molecular descriptor calculation (Section 7.7)
- **PubChemPy** – PubChem database query interface (Section 7.8)

7.1 ADMET-AI

ADMET-AI is a machine learning platform that predicts 98 ADMET properties using an ensemble of graph neural network and gradient boosting models trained on curated datasets from TDC, ChEMBL, and DrugBank. The tool provides predictions for nuclear receptor activity, toxicity endpoints, and pharmacokinetic parameters with percentile rankings against approved drugs.

Endpoint	Prediction	Percentile	Interpretation
[redacted]	0.750	97.6%	**High** [redacted]
[redacted]	0.428	96.8%	Moderate [redacted]
[redacted]	0.051	73.9%	Low [redacted]
[redacted]	0.004	25.9%	Minimal [redacted]
[redacted]	0.341	88.6%	Moderate [redacted]
[redacted]	0.167	82.0%	Low [redacted]
[redacted]	0.014	65.7%	Minimal [redacted]

Table 13 *ADMET-AI nuclear receptor activity predictions for oxybenzone. Percentile rankings are relative to FDA-approved drugs. The high NR-ER score (97th percentile) aligns with experimental HTS findings.*

Interpretation. ADMET-AI predicts strong [redacted] (NR-ER: 0.75, 97th percentile), consistent with experimental Tox21 and ChemHarmony findings. The low [redacted] prediction aligns with HTS data showing [redacted] results. The elevated [redacted] prediction (88th percentile) suggests potential for additional receptor-mediated toxicity pathways.

7.2 ChEMBL Multitask Model

ChEMBL Multitask Model predicts activity across 786 ChEMBL protein targets using a neural network trained on bioactivity data. The model uses Morgan fingerprints (1024-bit) as input and outputs probability scores for each target.

ChEMBL Target ID	Probability	Target Description
[redacted]	0.981	[redacted]
[redacted]	0.979	[redacted]
[redacted]	0.952	[redacted]
[redacted]	0.946	[redacted]
[redacted]	0.929	[redacted]
[redacted]	0.916	[redacted]
[redacted]0	0.891	[redacted]

Table 14 *Top ChEMBL target predictions for oxybenzone (threshold >0.5). Of 786 targets screened, 93 scored above threshold. Strong CYP450 interactions indicate potential for drug-drug interactions.*

Interpretation. Of 786 targets, 93 scored above 0.5. Strong [redacted] suggest potential drug-drug interactions relevant to oxybenzone's systemic absorption from sunscreen use. The [redacted] (0.95) aligns with ADMET-AI NR-AhR findings, providing cross-validation of this secondary pathway. Notably, direct estrogen receptor predictions were not among top hits, as ChEMBL training data may underrepresent known ER ligands.

7.3 StopTox

StopTox predicts 6 acute toxicity endpoints using gradient boosting classifiers trained on curated toxicity datasets. The tool employs MACCS and Morgan fingerprints with an ensemble approach for improved accuracy.

Endpoint	Prediction	Probability
Acute Oral Toxicity	[redacted]	[redacted]
Acute Dermal Toxicity	**Toxic**	0.646
Acute Inhalation Toxicity	[redacted]	[redacted]
Skin Sensitization	[redacted]	[redacted]
Skin Irritation	[redacted]	[redacted]
Eye Irritation	[redacted]	[redacted]

Table 15 *StopTox acute toxicity predictions for oxybenzone. Dermal toxicity prediction is notable given oxybenzone's primary exposure route through sunscreen application.*

Interpretation. StopTox predicts dermal toxicity concern (probability 0.65), relevant given oxybenzone's primary exposure route through sunscreen application. The low [redacted] and [redacted] predictions are consistent with oxybenzone's long history of topical use, while the dermal toxicity flag may relate to systemic absorption effects rather than local reactions.

7.4 Admetica

Admetica provides 18 ADMET predictions using chemprop v2.0.0 graph neural network models trained on benchmark datasets. Endpoints include solubility, permeability, metabolic stability, and toxicity.

Property	Value	Unit
Caco-2 Permeability	[redacted]	log cm/s
Aqueous Solubility	[redacted]	log mol/L
Lipophilicity (logD)	[redacted]	–
Human Intestinal Absorption	[redacted]	probability
Plasma Protein Binding	[redacted]	%
CYP1A2 Inhibitor	[redacted]	probability
CYP2C19 Inhibitor	[redacted]	probability
CYP2C9 Inhibitor	[redacted]	probability
CYP2D6 Inhibitor	[redacted]	probability
CYP3A4 Inhibitor	[redacted]	probability
hERG Inhibition	[redacted]	probability

Table 16 *Admetica pharmacokinetic and toxicity predictions for oxybenzone. High [redacted] probabilities corroborate ChEMBL multitask findings.*

Interpretation. High predicted [redacted] (HIA: 1.0) and [redacted] are consistent with known oxybenzone pharmacokinetics. Strong [redacted] predictions ([redacted]) corroborate ChEMBL multitask findings and suggest oxybenzone could affect metabolism of co-administered drugs. Low [redacted]) suggests [redacted].

**Note: The plasma protein binding value of 101.3% is a model artifact. Values exceeding 100% are physically impossible and should be interpreted as indicating high/extensive binding approaching saturation. This reflects a limitation of the predictive model rather than actual pharmacokinetics.*

7.5 SolTranNet

SolTranNet predicts aqueous solubility using a transformer-based neural network trained on the AqSolDB dataset containing 9,982 compounds with experimental solubility measurements.

Result.

Aqueous Solubility: [redacted]

Interpretation. This moderate-to-low solubility prediction ([redacted]) aligns with oxybenzone's lipophilic character (experimental logP $\approx$ 3.6) and its use in oil-based sunscreen formulations. The solubility is sufficient for systemic distribution following dermal absorption but limits direct dissolution in aqueous biological compartments.

7.6 OPERA (CERAPP/CoMPARA)

OPERA (OPEn structure-activity/property Relationship App) is EPA's open-source QSAR application implementing regulatory-grade models including CERAPP (Collaborative Estrogen Receptor Activity Prediction Project) and CoMPARA (Collaborative Modeling Project for Androgen Receptor Activity). These consensus models were developed collaboratively by EPA, NIEHS, and international partners specifically for endocrine disruption screening.

Model / Endpoint	Prediction	Confidence	AD
CERAPP (Estrogen Receptor)			
ER Agonist	[redacted]	[redacted]	In-domain
ER Antagonist	[redacted]	[redacted]	In-domain
ER Binding	[redacted]	[redacted]	In-domain
CoMPARA (Androgen Receptor)			
AR Agonist	[redacted]	[redacted]	In-domain
AR Antagonist	[redacted]	[redacted]	In-domain
AR Binding	[redacted]	[redacted]	In-domain

Table 17 *OPERA CERAPP and CoMPARA predictions for oxybenzone. CERAPP predicts [redacted] activity [redacted]. CoMPARA predicts [redacted], consistent with [redacted]. AD = applicability domain.*

Interpretation. OPERA's regulatory-grade predictions provide critical evidence for oxybenzone's endocrine activity:

- **CERAPP**: Predicts oxybenzone as [redacted]) with high confidence (0.80–0.87). This is consistent with Tox21 HTS findings and ADMET-AI predictions.
- **CoMPARA**: Importantly predicts [redacted] (confidence 0.98) but [redacted] (confidence 0.80). This clarifies the Tox21 AR assay results—oxybenzone likely acts as an [redacted].
- All predictions are within the model's applicability domain, indicating reliable results for oxybenzone's chemical structure.

The CERAPP/CoMPARA results are particularly significant because these models were specifically developed and validated by EPA for EDSP (Endocrine Disruptor Screening Program) prioritization, making them regulatory-grade predictions with established performance benchmarks.

7.7 Mordred Molecular Descriptors

Mordred calculates 1,613 2D and 3D molecular descriptors for QSAR modeling and molecular characterization. Key physicochemical properties for oxybenzone are presented below.

Descriptor	Value	Comment
Molecular Weight	[redacted]	[redacted]
LogP (calculated)	[redacted]	[redacted]
TPSA	[redacted]	[redacted]
H-bond Donors	[redacted]	[redacted]
H-bond Acceptors	[redacted]	[redacted]
Rotatable Bonds	[redacted]	[redacted]
Aromatic Rings	[redacted]	[redacted]

Table 18 *Key molecular descriptors for oxybenzone calculated using Mordred. All values satisfy Lipinski's Rule of Five criteria.*

Interpretation. All calculated descriptors satisfy Lipinski's Rule of Five criteria for drug-likeness, predicting good oral bioavailability and membrane permeability. The low [redacted] ([redacted]) is consistent with efficient dermal absorption, supporting the systemic exposure observed in human biomonitoring studies.

7.8 PubChemPy Database Query

PubChemPy provides programmatic access to PubChem's compound database containing over 100 million chemical structures. Compound information for oxybenzone (CID: 4632) was retrieved.

Property	Value
PubChem CID	4632
IUPAC Name	(2-hydroxy-4-methoxyphenyl)-phenylmethanone
Molecular Formula	$C_{14}H_{12}O_3$
XLogP	3.6
Complexity	224
InChIKey	DXGLGDHPHMLXJC-UHFFFAOYSA-N
CAS Registry	131-57-7
Synonyms	Benzophenone-3, BP-3, Eusolex 4360, Escalol 567

Table 19 *Oxybenzone compound information retrieved from PubChem via PubChemPy.*

Interpretation. PubChem provides authoritative chemical identifiers essential for cross-referencing experimental data across databases. The experimental XLogP (3.6) confirms the lipophilic character predicted by Mordred (logP: 2.63) and explains oxybenzone's preferential partitioning into lipid-based sunscreen formulations and biological membranes.

7.9 Tool Predictions Summary

Table 20 presents consensus predictions across all computational tools for EATS-relevant endpoints.

EATS Axis	Prediction	Confidence	Supporting Tools
Estrogen	[redacted]	High	ADMET-AI ([redacted]); OPERA CERAPP ([redacted])
Androgen	[redacted]	High	OPERA CoMPARA ([redacted]); ADMET-AI ([redacted])
Thyroid	[redacted]	Moderate	ADMET-AI ([redacted]); no thyroid models in suite
Steroidogenesis	[redacted]	Low	[redacted]; ADMET-AI [redacted]

Table 20 *Consensus computational predictions for oxybenzone EATS activity across all evaluated tools. OPERA CERAPP/CoMPARA provide regulatory-grade predictions for ER and AR activity.*

7.10 Overall Interpretation

Computational Tool Consensus

Computational predictions consistently identify oxybenzone as an estrogen receptor agonist and androgen receptor antagonist.

The *in silico* evidence aligns strongly with experimental HTS data from Tox21 and ToxCast:

- **ADMET-AI** predicts [redacted]
- **OPERA CERAPP** confirms [redacted] (regulatory-grade model)
- **OPERA CoMPARA** clarifies [redacted], explaining Tox21 AR hits
- **ChEMBL/Admetica** identify [redacted]
- **StopTox** flags dermal toxicity consistent with systemic absorption concerns
- **Mordred/SolTranNet** confirm physicochemical properties enabling dermal absorption

Limitations. Computational predictions are inherently model-dependent and should not replace experimental data. Key limitations include:

- **Training data bias**: Models may underperform for chemical classes underrepresented in training sets
- **Applicability domain**: Predictions are most reliable within the chemical space of training compounds
- **Endpoint coverage**: Not all EATS-relevant endpoints have validated computational models
- **Metabolite effects**: Most tools do not account for metabolic activation to BP-1

Concordance with Experimental Data. The strong concordance between computational predictions and experimental HTS results (Table 21) reinforces confidence in oxybenzone's estrogenic mechanism of action. Both computational and experimental approaches independently identify estrogen receptor agonism as the primary endocrine activity.

Endpoint	Computational	Experimental (Tox21)	Concordance
ER Agonist	[redacted]	[redacted]	**Yes**
AR Agonist	[redacted]	[redacted]	**Yes**
AR Antagonist	[redacted]	[redacted]	**Yes**
Aryl Hydrocarbon Receptor	[redacted]	[redacted]	**Yes**
CYP450 Inhibition	[redacted]	Not directly tested	N/A

Table 21 *Concordance between computational predictions and experimental HTS results for oxybenzone. *Tox21 AR results now interpretable as antagonism based on OPERA CoMPARA predictions.*

The computational tools evaluated in this section provide valuable supporting evidence for hazard characterization. While no single tool is definitive, the consensus across multiple independent models—each using different algorithms, training data, and molecular representations—provides robust *in silico* support for oxybenzone's classification as an endocrine disruptor acting through [redacted]. The OPERA CERAPP/CoMPARA predictions are particularly significant as they represent regulatory-grade models specifically developed by EPA for endocrine disruptor screening.

Part II

Conclusion & Assessment

8 Primary Determination

Question: Is oxybenzone (benzophenone-3) an endocrine disruptor for humans?

DETERMINATION

Oxybenzone Meets Criteria for Endocrine-Active Substance with Concern

Oxybenzone demonstrates **estrogen receptor agonist activity** across multiple high-throughput screening platforms. This represents direct receptor binding and transcriptional activation of estrogen-responsive genes—a mechanism characteristic of endocrine-active substances.

Note: This determination is based on in vitro screening and computational evidence. Definitive regulatory classification as an "endocrine disruptor" typically requires in vivo guideline studies demonstrating adverse effects at relevant doses with clear linkage to the molecular initiating event.

9 EATS Modality Assessment

Based on the systematic database review and high-throughput screening evidence, oxybenzone's activity across EATS (Estrogen, Androgen, Thyroid, Steroidogenesis) modalities is summarized below:

EATS Modality	Activity	Confidence	Key Evidence
Estrogen	**ACTIVE**	High	[redacted]
Androgen	**Antagonist**	High	[redacted]
Thyroid	Inactive	High	[redacted]
Steroidogenesis	Limited	Moderate	[redacted]

Table 22 *Summary of oxybenzone activity across EATS modalities based on Tox21, ChemHarmony, and computational tool (ADMET-AI, OPERA) evidence.*

10 Severity Assessment

Concern Level: MODERATE TO HIGH

Rationale:

- **Widespread human exposure**: Oxybenzone is present in 60–97% of the US population (NHANES biomonitoring data)
- **Dermal absorption**: Significant systemic absorption through skin application of sunscreens
- **Confirmed mechanism**: Direct estrogen receptor agonism is a well-characterized adverse outcome pathway
- **Metabolite amplification**: The O-demethylated metabolite benzophenone-1 (BP-1) demonstrates similar or enhanced estrogenic potency
- **Vulnerable populations**: Exposure during pregnancy, infancy, and childhood raises developmental concerns

Mitigating factors:

- Relatively weak estrogenic potency compared to estradiol (EC50 typically in µM range)
- Short biological half-life with rapid urinary excretion
- No direct thyroid receptor activity observed

11 Confidence Level

Confidence in Assessment: HIGH

Confidence is HIGH for the determination that oxybenzone is an estrogen receptor agonist and androgen receptor antagonist based on:

1. **Reproducible HTS results**: Consistent positive results across 42+ estrogen receptor assays in Tox21
2. **Multiple platforms**: Activity confirmed via BLA reporter, luciferase, and ligand-binding domain assays
3. **Cross-database consistency**: ChemHarmony, ToxCast, ICE, and PubChem all report estrogenic activity
4. **Computational tool consensus**: ADMET-AI [redacted] and OPERA CERAPP [redacted] independently predict [redacted]
5. **AR antagonism clarified**: OPERA CoMPARA predicts [redacted] but [redacted], explaining Tox21 AR hits
6. **Metabolite confirmation**: Testing of both parent compound and BP-1 metabolite shows concordant results
7. **Structural plausibility**: Benzophenone scaffold with hydroxyl/methoxy groups is consistent with ER binding
8. **Gene expression evidence**: CTDbase documents interactions with ESR1, ESR2, and steroidogenic enzymes

Limitations affecting confidence:

- No OECD/EPA guideline-compliant DART studies in ToxRefDB
- Limited in vivo dose-response data at environmentally relevant concentrations
- Human epidemiological evidence remains inconsistent

12 Key Supporting Evidence

12.1 In Vitro Evidence (Strong)

- **Tox21 Screening**: Oxybenzone showed 19/42 (45%) active results in estrogen receptor assays; BP-1 showed 12/28 (43%) active
- **ChemHarmony**: [redacted]
- **Multiple assay formats**: [redacted]
- **Negative controls**: [redacted]

12.2 Computational Tool Evidence (Strong)

- **ADMET-AI**: Predicts [redacted]
- **OPERA CERAPP**: [redacted]
- **OPERA CoMPARA**: [redacted]
- **ChEMBL Multitask**: [redacted]
- **Admetica/StopTox**: [redacted]
- **Tool consensus**: [redacted]

12.3 Gene Expression Evidence (Moderate)

- **CTDbase**: 530 chemical-gene interactions documented
- **ER interactions**: [redacted]
- **AR interactions**: [redacted]
- **Steroidogenic enzymes**: [redacted]

12.4 Regulatory Evidence (Strong)

- **SCCS Opinion**: 32 endocrine endpoints extracted; [redacted]
- **SCCS Low-Dose Effects**: [redacted]
- **GHS Classification**: H361 (Suspected of damaging fertility or the unborn child) in REACH registrations
- **ToxValDB**: 211 records including 48 DART-specific endpoints with quantitative dose-response data
- **Regulatory concern**: Multiple jurisdictions (EU, Hawaii, Palau) have restricted oxybenzone in sunscreens
- **SCCS Conclusion**: Effects on spermatocytes “may be due to an ED effect of BP-3”

13 Scope and Limitations of This Assessment

Hazard Identification vs. Risk Assessment

This report presents a **hazard identification** for oxybenzone's endocrine activity. It does **not** constitute a formal risk assessment. Important distinctions:

- **Hazard identification:** Determines whether a substance has the potential to cause adverse effects (addressed here)
- **Risk assessment:** Integrates hazard with exposure to estimate probability of harm (not performed)

What this assessment does NOT include:

- Margin-of-exposure (MOE) analysis comparing human internal doses to effect levels
- Biomonitoring equivalents or internal dose comparisons
- Consideration of potency relative to endogenous estrogen
- Risk-benefit analysis for sunscreen use (UV protection vs. endocrine exposure)
- Quantitative human health risk characterization

Context: While oxybenzone is detected in 60–97% of the US population, the health significance of these exposure levels requires formal risk assessment with margin-of-exposure analysis. Sunscreens provide documented benefits for skin cancer prevention that must be weighed against potential endocrine concerns in any risk management decision.

14 Data Gaps

Despite strong in vitro evidence for endocrine activity, significant gaps remain in the toxicological characterization:

1. **No guideline DART studies**: ToxRefDB contains zero OECD 414/416/421/422 compliant studies for oxybenzone
2. **ICE database gap**: No records in NIEHS Integrated Chemical Environment DART or endocrine MOA datasets
3. **Multigenerational effects**: No systematic F0/F1/F2 reproduction studies available
4. **Developmental neurotoxicity**: No DNT evaluation despite widespread pediatric exposure
5. **Human dose-response**: Limited data correlating urinary biomarker levels with health outcomes
6. **Mixture effects**: Sunscreens contain multiple UV filters; combined endocrine effects unstudied

15 Recommendations

15.1 Regulatory Recommendations

1. **Prioritize for OECD guideline testing**: Conduct OECD 414 (prenatal developmental toxicity), 416 (two-generation reproduction), and 426 (developmental neurotoxicity) studies
2. **Update hazard classification**: Consider upgrading from H361 to H360 if guideline studies confirm reproductive toxicity
3. **Biomonitoring**: Continue NHANES monitoring with enhanced focus on pregnant women and children
4. **Exposure assessment**: Characterize aggregate exposure from multiple sources (sunscreens, cosmetics, plastics)

15.2 Research Recommendations

1. **In vivo confirmation**: Conduct uterotrophic assays and Hershberger assays to confirm in vitro ER activity
2. **Metabolite characterization**: Further evaluate BP-1 and other metabolites for relative endocrine potency
3. **Epidemiological studies**: Conduct prospective cohort studies correlating prenatal exposure with developmental outcomes
4. **Mechanism studies**: Investigate whether AR activity represents antagonism or assay interference

15.3 Risk Management Recommendations

1. **Precautionary approach**: Consider limiting oxybenzone use in products for pregnant women and children pending additional data
2. **Alternative assessment**: Evaluate endocrine safety of alternative UV filters (zinc oxide, titanium dioxide, newer organic filters)
3. **Labeling**: Require clear labeling of oxybenzone content in sunscreens and cosmetics
4. **Exposure reduction**: Promote non-chemical sun protection measures (clothing, shade) especially for vulnerable populations

16 Conclusion

This weight-of-evidence assessment concludes that **oxybenzone (benzophenone-3) meets criteria for classification as an endocrine-active substance with concern**, demonstrating estrogen receptor agonist activity and androgen receptor antagonist activity in high-throughput screening. The evidence base includes:

- High-throughput screening data from 1,060+ assays across Tox21 and related programs
- Consistent positive results for ER agonism across multiple assay formats [redacted]
- Computational tool consensus: ADMET-AI [redacted], OPERA CERAPP [redacted]
- OPERA CoMPARA clarification that AR activity represents [redacted], explaining [redacted] effects
- Similar activity in the major human metabolite benzophenone-1 ([redacted])
- Gene expression evidence for interactions with [redacted]s and [redacted] (CTDbase: [redacted])
- SCCS regulatory opinion ([redacted]) with 32 endocrine endpoints; NOAEL [redacted] mg/kg bw/day based on [redacted] effects
- Low-dose effects ([redacted] mg/kg/day) on AGI and mammary gland ERα+ cells in mouse studies—though some effects were transient and cross-study reproducibility remains to be established
- Regulatory classification as a suspected reproductive toxicant (H361)

While oxybenzone shows no activity at thyroid receptors (0/27 assays) and limited effects on steroidogenesis, the confirmed estrogenic and anti-androgenic activity—combined with widespread human exposure through sunscreen use (60–97% US population)—warrants regulatory attention and additional systematic testing following OECD guidelines.

The primary mechanisms of concern are **direct estrogen receptor agonism** and **androgen receptor antagonism**,

which place oxybenzone in the category of substances acting through receptor-mediated pathways characteristic of endocrine-active chemicals. The OPERA CERAPP/CoMPARA predictions are particularly significant as they represent EPA's regulatory-grade QSAR models specifically developed for EDSP prioritization.

Part III

Methods

17 Technical Pipeline Overview

This assessment was conducted using an automated pipeline that integrates multiple computational approaches for systematic evidence synthesis. The methodology is designed to be reproducible and extensible to other chemical-hazard pairs.

17.1 Data Sources

17.1.1 Literature Databases

- **PubMed:** Primary source for peer-reviewed biomedical literature
- **PubTator3:** Named entity recognition for chemical and gene mentions
- **CTDbase:** Curated chemical-gene interactions

17.1.2 Toxicology Databases

- **Tox21:** High-throughput screening data from NIH/EPA/FDA collaboration
- **ToxCast:** EPA's computational toxicology screening program
- **ChemHarmony:** Harmonized activity data across multiple sources
- **ToxValDB:** Quantitative toxicity values compilation
- **ToxRefDB:** EPA guideline study repository
- **ICE:** NIEHS Integrated Chemical Environment

17.1.3 Computational Tools

- **ADMET-AI:** Machine learning ADMET property predictions
- **ChEMBL Multitask:** Neural network target activity predictions
- **StopTox:** Gradient boosting toxicity predictions
- **Admetica:** Graph neural network ADMET models
- **SolTranNet:** Transformer-based solubility prediction
- **Mordred/PaDEL-Py:** Molecular descriptor calculation

17.2 Pipeline Stages

17.2.1 Stage 1: Literature Collection

PubMed queries retrieved publications using complementary search strategies:

- Exposure-focused: "oxybenzone" OR "benzophenone-3" OR "BP-3"
- AOP-focused: "endocrine disruption" combined with relevant MeSH terms

17.2.2 Stage 2: Relevance Filtering

LLM-based classification (Gemini 2.5 Flash) identified papers relevant to:

- Endocrine activity assessment
- Mechanistic characterization
- Human exposure evaluation

17.2.3 Stage 3: Full-Text Processing

PDF retrieval and OCR processing enabled extraction of:

- Quantitative findings (concentrations, effect sizes)
- Study metadata (species, exposure route, duration)
- Mechanistic conclusions

17.2.4 Stage 4: Key Event Extraction

LLM-based entity extraction identified:

- Molecular initiating events (receptor binding)
- Intermediate key events (gene expression changes)
- Adverse outcomes (reproductive effects)

17.2.5 Stage 5: Evidence Synthesis

Integration of literature evidence with database findings and computational predictions to generate weight-of-evidence conclusions.

17.3 Quality Assurance

- **Reproducibility:** Pipeline orchestrated via DVC with cached intermediate outputs
- **Validation:** Cross-referencing of extracted findings against database records
- **Transparency:** All source data and processing scripts version-controlled

17.4 LLM-Based Processing: Limitations and Validation Status

Important Methodological Consideration

This assessment relies extensively on large language model (LLM) processing for:

- Relevance screening of literature abstracts
- Bias assessment of study quality
- Key event extraction from full-text papers
- Narrative synthesis of evidence

Current validation status:

- No formal inter-annotator agreement metrics have been computed
- No systematic benchmarking against human expert curation has been performed
- Error rates for extraction and classification have not been quantified
- Confidence calibration of LLM outputs has not been validated

Implications: Users should interpret LLM-derived findings as *hypothesis-generating* rather than definitive. The pipeline accelerates evidence synthesis but does not replace expert review. Critical findings should be verified against primary sources before regulatory or decision-making use.

17.5 Limitations

- **LLM extraction variability:** LLM-based extraction may miss nuanced findings or introduce extraction errors; outputs should be considered preliminary
- **Model dependency:** Computational predictions are model-dependent and may not generalize across chemical classes
- **Database coverage:** Database coverage varies by chemical; absence of evidence should not be interpreted as evidence of absence
- **Publication bias:** Literature tends toward positive findings; negative results may be underrepresented
- **Temporal limitations:** Database queries reflect data available at time of analysis; newer studies may not be captured

17.6 Tool Versions

Tool	Version	Purpose
Python	3.12	Pipeline orchestration
DVC	3.x	Data version control
ADMET-AI	latest	ADMET prediction
Chemprop	2.0.0	GNN predictions
RDKit	2024.x	Cheminformatics
ONNX Runtime	1.x	Model inference
Gemini	2.5 Flash	LLM processing

Table 23 *Software and tool versions used in the pipeline.*

References

[1] [redacted]

[2] Thomas Luechtefeld. BioBricks-OKG: An open knowledge graph for cheminformatics and chemical safety. NSF Proto-OKN Award 2333728, 2024. URL https://www.nsf.gov/awardsearch/showAward?AWD_ID=2333728. Query interface: https://query-okg.app.biobricks.ai/.

[3] [redacted]

[4] [redacted]

[5] [redacted]