Coral ecotoxicological data evaluation for the environmental safety assessment of UV filters

Emily E. Burns[1] and Iain A. Davies[1]

[1]Personal Care Products Council, Washington DC, USA

Supporting Information

Table S1. Screening and data quality assessment for the Danovaro et al. (2008) study

Table S2. Screening and data quality assessment for the Downs et al. (2016) study

Table S3. Screening and data quality assessment for the Fel et al. (2019) study

Table S4. Screening and data quality assessment for the He et al. (2019a) study

Table S5. Screening and data quality assessment for the He et al. (2019b) study. Two scores are provided for certain data quality assessment questions. This is to differentiate between the adult assays and the larval settlement assay. The score for the adult assay always appears first. If only one score is presented, it applies to both the adult and larval assay

Table S6. Screening and data quality assessment for the McCoshum et al. (2016) study

Table S7. Screening and data quality assessment for the Stien et al. (2019) study

Table S8. Screening and data quality assessment for the Wijgerde et al. (2020) study

Table S1. Screening and data quality assessment for the Danovaro et al. (2008) study

Screening questions		Score	Comment
RQ1	Is the endpoint ecologically relevant?	Pass	The endpoints were zooxanthellae released (which is an appropriate way to quantify bleaching), bleaching initiation and bleaching rate. The bleaching rate is considered not ecologically relevant, as the speed in which bleaching occurred during is less relevant than the extent of bleaching during the experiment. Similarly, bleaching initiation is also not considered to be ecologically relevant.
RQ2	Test organism relevant to the compartment, test compound and/or assessment?	Pass	Wild adult *Acropora* sp. and *A. pulchra* were studied in the single UV filter exposures (whole sunscreen formulation tests beyond the scope of the assessment). UV filters studied include oxybenzone (BP-3), avobenzone (AVO), octinoxate (EHMC), octocrylene (OC), and octisalate (EHS). Other UV filters studied are not relevant as not permitted for use in the United States.
RQ3	Was a negative control and solvent control (if necessary) at least duplicated?	Fail	A negative control was included in both the Red Sea and Andaman Sea experiments. A solvent control was only included in the Red Sea experiment (*Acropora* sp.) but not the Andaman Sea experiment (*A. pulchra*).
RQ4	Are ≥ 4 treatment concentrations included (including control) or experiment specifically designed as limit test?	Pass	Reports exposure to four treatment concentrations; however, only data from a single treatment is reported for each UV filter experiment.
RQ5	Are endpoints based on measured concentrations if they deviate by ≥ 20% of the nominal concentration? If only nominal endpoints presented, is any analytical verification undertaken?	Fail	Test concentrations are not analytically verified at any point throughout the test.
Data quality assessment			
1	Biological endpoint stated and defined.	10	Biological endpoints are reported, zooxanthellae released was determined by counting zooxanthellae in tissue samples under a microscope and normalized for the coral nubbin area. Bleaching rate was determined by CMYK color shifts from photos and then compared to controls through time to give a 'bleaching rate' and 'bleaching initiation' time.
2	Are relevant validity criteria stated and met?	3	Expert judgement required to determine validity criteria. Is control mortality reported? Is the solvent control significantly different from the negative control? Are endpoint responses observed in controls? No bleaching was observed in any control, but 16% of zooxanthellae were released in the solvent control (0% in negative control). It is not reported if this was significant and difficult to interpret as negative control not provided for *A. pulchra*. No mortality reported in controls.
3	Is the test system defined and appropriate (flow-through, semi-static or static conditions)?	2	Static exposure conditions were applied and the length of the experiment exceeded 48 hours.
4	Is the test substance concentration maintained ± 20% throughout the exposure?	0	No analytical verification was undertaken to determine if the UV filters were stable over the duration of the test.

Table S1. Continued

Screening questions		Score	Comment
5	Is the test system appropriate for the test organism?	0	Exposed in 0.02 µm filtered seawater. It cannot be determined whether this filtering would also remove key nutrients/particulates to avoid unintended stress of the coral nubbins. Nubbins were placed in sealed plastic bags and placed back in the sea. The seawater may provide temperature consistency and some water movement; however, nubbins weren't secured and would likely hit each other, causing stress. The seawater was also not characterized or checked for background UV filters or other contaminants.
6	Biological effect stated?	0	No biological effect is stated (e.g., LOEC or NOEC).
Is a parallel reference toxicant study conducted? Evaluate based on scenario:			
7a	If required in relevant guideline.		
7b	If studying a wild organism.		
7c	If studying a non-standard organism/ non-standard endpoint.	0	A positive control was not included despite studying wild non-standard organisms in a non-standard test system.
8	Test substance specifically identified (e.g., chemical abstract service [CAS] number) and source reported?	4	CAS number is not provided in the text, but sufficient physicochemical information is provided. The source is reputable, Sigma Aldrich.
9	Test substance purity reported?	0	Purity is not provided and exposure concentrations are also not measured.
10	Is the experiment appropriately replicated and not pseudoreplicated?	4	Three replicate 'sets' reported. The methods indicate that there are three nubbins (possibly more) per experimental unit; therefore, the experiment is not pseudoreplicated.
11	Is a significant concentration-response relationship demonstrated?	0	A dose-response relationship is not presented in the study. The authors state that their results indicate effects were not dose dependent, but this could be referring to the sunscreen formulation exposures rather than the single compound exposures. Data is only presented for a single exposure concentration.
12	Was a suitable statistical method/model described to determine the toxicity?	0	No statistical testing reported for zooxanthellae released endpoint. An ANOVA based on CYMK color shift was calculated and these results were translated into a bleaching rate, which species this relates to is unclear (e.g., Table 3 in study text). Post hoc testing was not included.
13	Was significance level listed for NOEC/LOEC/MATC as 0.05 or less and for the NEC/LC/EC an estimate of variability reported?	0	Significance in the ANOVA was set to 0.05, but a statistical endpoint was not calculated resulting in no score.

Table S1. Continued

Screening questions		Score	Comment
14	Exposure duration stated? Is the duration appropriate considering the species, life stage and endpoint (acute or chronic)?	0	The exposure duration is not explicitly reported. The authors allude to 96 hours in the discussion, but it is not clear. It is also not clear when the zooxanthellae density endpoint was observed.
15	Is a suitable test concentration separation factor used?	3	An appropriate scaling factor was applied (e.g., < 10).
16	Do the test concentrations adequately bracket the biological endpoint?	0	This cannot be determined from the limited data presented as only results from a single treatment concentration are reported. Furthermore, a biological endpoint was not presented (e.g., EC50 or NOEC).
17	Are organisms appropriately acclimatized to test conditions?	0	Nubbins collected from the wild and not acclimatized to test conditions prior to exposure. They were cut from source colonies and placed in 0.02 µm filtered seawater. Therefore, no time given to heal fragments.
18	Are organisms well described? (e.g., length, mass, age, strain, sex, etc.)	3	Nubbins collected from wild colonies (3 – 6 cm diameter) in various regions. Each nubbin had a minimum of 300 polyps. For assessment, assumed wild colonies are genetically different.
19	Test vessels appropriate for the test substance?	0	The test vessels (polyethylene whirl-pack bags) are inappropriate as interactions with test vessel are possible (e.g., sorption). No validation was undertaken to determine whether this would impact exposures.
20	Are analytical methods described and appropriate QA/QC reported?	0	No analytical method provided.
Test medium parameters			
21a	Dissolved oxygen	0	None reported.
21b	Temperature	0	None reported.
21c	pH	0	None reported.
21d	Salinity/conductivity	0	None reported.
21f	Species specific – include if specific parameters needed.		
22	If used, is solvent in the appropriate range?	2	The solvent control is reported as 33 µL/L, which is acceptable for a static test.
23	Is the solvent suitable for the test species?	2	Polyethylene glycol is used as the carrier solvent, which is appropriate.

Table S2. Screening and data quality assessment for the Downs et al. (2016) study

Screening questions		Score	Comment
RQ1	Is the endpoint ecologically relevant?	Pass	Mortality is an ecologically relevant endpoint. Only the 24 hour data is assessed (closer to standard test length than 8 hour data). The study also reported chlorophyll fluorescence as a measure of bleaching. The chlorophyll fluorescence method was cited by the authors in a previous study as a 'gross estimation' (of bleaching) as the method could not appropriately accommodate the geometry of larvae. Therefore, due to how this endpoint was quantified it is considered non-ecologically relevant. Other non-relevant endpoints were DNA damage and deformity.
RQ2	Test organism relevant to the compartment, test compound and/or assessment?	Pass	The species studied is *Stylophora pistillata* larvae (planulae) exposed to oxybenzone (BP-3). Multiple species are also included in a cell line experiment, but this is not a validated assay and *in vitro* studies are beyond the scope of this assessment and not discussed further.
RQ3	Was a negative control and solvent control (if necessary) at least duplicated?	Pass	A solvent and negative control were performed.
RQ4	Are ≥ 4 treatment concentrations included (including control) or experiment specifically designed as limit test?	Pass	Five treatment concentrations were included.
RQ5	Are endpoints based on measured concentrations if they deviate by ≥ 20% of the nominal concentration? If only nominal endpoints presented, is any analytical verification undertaken?	Fail	Test concentrations were not verified at any point throughout the test.
Data quality assessment			
1	Biological endpoint stated and defined.	10	The biological relevant endpoint considered in this assessment (mortality) is stated and clearly defined. The non-standard endpoints were also stated and defined; however, it is noted that the chlorophyll fluorescence quantification method may not be appropriate for coral planulae (Downs et al. 2015).
2	Are relevant validity criteria stated and met?	3	A modified OECD TG 236 (fish embryo test) is reported to be followed by the authors. The authors did not detail how the test was modified for coral planulae (they cited Downs et al. (2014) which also does not provide this detail). Based on OECD 236, the following criteria were assessed: Was the water temperature maintained? Was control survival greater than >90% at the end of the test (both negative and solvent control)? Is dissolved oxygen reported? Is a minimum of 30% mortality observed in a suitable positive control? Are endpoint responses observed in controls? All planulae survived and were not deformed in the controls. It is not reported if the water temperature is maintained. Survival in controls was greater than 90% at the end of the test. The dissolved oxygen concentration was not measured. Data for both controls (negative and solvent) are not reported to compare; however, the authors state that there was no significant difference between controls. A positive control was not included to assess.

Table S2. Continued

Screening questions		Score	Comment
3	Is the test system defined and appropriate (flow-through, semi-static or static conditions)?	3	Static exposure conditions were applied and this was not a coral reproductive assay; however, the length of the test was very short (24 hours) and therefore scored as a semi-static test.
4	Is the test substance concentration maintained ± 20% throughout the exposure?	0	Test concentrations were not measured throughout the test. Therefore, it is unknown if the test concentration was maintained throughout the test. Data from other laboratory studies suggest that significant losses of BP-3 could occur over 24 hours (e.g., He et al. 2019b).
5	Is the test system appropriate for the test organism?	0	The time from collection of wild organisms to test exposure is very short. Additionally, artificial seawater was used in the exposure which could stress planuae (no acclimation period included). Lighting not described, aeration not reported (dissolved oxygen not measured). Reported evaporation and necessary reconstitution of the test medium in the light exposure (as reported by Downs et al., 2014). Ideally laboratory-spawned or spawned under controlled conditions planulae would be used. Finally, the exposure volume is comparatively low to similar studies (i.e., 2.5 mL versus 10 mL reported by He et al. 2019).
6	Biological effect stated?	5	The LC50 is reported which is the appropriate endpoint.
Is a parallel reference toxicant study conducted? Evaluate based on scenario:			
7a	If required in relevant guideline.		
7b	If studying a wild organism.		
7c	If studying a non-standard organism/ non-standard endpoint.	0	A reference toxicant is required by the guideline cited. Additionally, wild non-standard organisms were used.
8	Test substance specifically identified (e.g., chemical abstract service [CAS] number) and source reported?	0	A CAS number is provided in the text; however, it corresponds to benzophenone-2 rather than BP-3. Therefore, it cannot be confirmed that the appropriate test substance was used. The reported source of the substance is reputable.
9	Test substance purity reported?	0	Purity is not provided and since catalogue number pertains to benzophenone-2, it is not possible to determine the purity of BP-3 or whether it was even used in the test.
10	Is the experiment appropriately replicated and not pseudoreplicated?	4	For controls: ten planulae per plate, with two replicates, which is aligned with OECD 236 for controls. Each treatment included four replicates with ten organisms each.
11	Is a significant concentration-response relationship demonstrated?	2	A significant dose-response is reported but demonstrated with fewer than five concentrations. For example in the 24 h test, the dose-response is based on four test concentrations as 100% mortality was reached in the highest 2 – 3 treatments.
12	Was a suitable statistical method/model described to determine the toxicity?	3	Statistical methods are reported. The PROBIT analysis followed is in OECD guidance. The data was checked for normality. A Kruskal-Wallis one-way ANOVA followed by Dunnett's method for post hoc testing between treatments was used to determine the NOEC.
13	Was significance level listed for NOEC/LOEC/MATC as 0.05 or less and for the NEC/LC/EC an estimate of variability reported?	0	An estimate of variability is not provided.

Table S2. Continued

Screening questions		Score	Comment
14	Exposure duration stated? Is the duration appropriate considering the species, life stage and endpoint (acute or chronic)?	3	The exposure duration was 24 hours under light and dark conditions. Endpoints were assessed at the end of each exposure. This an acute test.
15	Is a suitable test concentration separation factor used?	1	The maximum spacing factor was applied (i.e., 10). OECD 236 suggests a spacing factor of 2.2, why this was modified is not reported. A range-finding test would have been useful to reduce the spacing factor; however, the time-sensitive nature of organism collection may have prevented this.
16	Do the test concentrations adequately bracket the biological endpoint?	3	The LC50 was not calculated by extrapolation. The chlorophyll fluorescence LOEC was observed at the lowest concentration tested, which is of reduced reliability.
17	Are organisms appropriately acclimatized to test conditions?	0	Larvae were collected from wild spawning corals using planula traps. Planulae were retrieved from the wild at 6:00 am, sorted by 7:15 am, and 45 minutes later subject to the test treatment exposure. Therefore, there was limited time for acclimatization and could this could induce significant stress as these were wild organisms.
18	Are organisms well described? (e.g., length, mass, age, strain, sex, etc.)	3	Freshly spawned planulae, collected from wild colonies at least 25 cm in diameter within the Inter-University Institute of Marine Sciences designated research area in Eilat, Israel. It is assumed different colonies were sampled.
19	Test vessels appropriate for the test substance?	3	The planulae were exposed in PTFE-Teflon microplates. Based on the logKow of BP-3, glass exposure vessels would be appropriate, but use of PTFE has been suggested for toxicity testing of difficult substances.
20	Are analytical methods described and appropriate QA/QC reported?	0	No analytical method is used and therefore not reported.
Test medium parameters			
21a	Dissolved oxygen	0	
21b	Temperature	1	22˚C.
21c	pH	0	
21d	Salinity/conductivity	1	38‰.
21f	Species specific – include if specific parameters needed.		
22	If used, is solvent in the appropriate range?	2	The solvent concentration did not exceed 0.1 mL/L.
23	Is the solvent suitable for the test species?	1	DMSO was used as the solvent, this is a non-standard species where potential interactions with DMSO have not been investigated.

Table S3. Screening and data quality assessment for the Fel et al. (2019) study

Screening questions		Score	Comment
RQ1	Is the endpoint ecologically relevant?	Fail	The endpoint studied was photosynthetic yield (Fv/Fm), which an indication of photoinhibition in the symbiont rather than the coral. This is a non-ecologically relevant endpoint.
RQ2	Test organism relevant to the compartment, test compound and/or assessment?	Pass	*Stylophora pistillata* was the test species. Various organic UV filters were studied, two of which are approved for use in the USA (octocrylene [OC] and avobenzone [AVO]) The assessment is limited to these two compounds as they are relevant to the problem formulation.
RQ3	Was a negative control and solvent control (if necessary) at least duplicated?	Pass	A solvent and negative control were included.
RQ4	Are there ≥ 4 treatment concentrations (including control) or experiment specifically designed as limit test?	Pass	There are four test concentrations (including control) for the AVO and OC exposures.
RQ5	Are endpoints based on measured concentrations if they deviate by ≥ 20% of the nominal concentration? If only nominal endpoints presented, is any analytical verification undertaken?	Pass	Test concentrations were analytically verified at various points throughout the test.
Data quality assessment			
1	Biological endpoint stated and defined.	10	The biological endpoint photosynthetic yield (dark-adapted maximum quantum yield of photosystem II in the symbiont) and how it was measured is reported.
2	Are relevant validity criteria stated and met?	5	Validity criteria not stated. Expert judgement is required to evaluate validity criteria. Is control mortality reported? Significant differences between solvent and negative control? Effects in controls? No significant difference between negative control and solvent controls. No mortality reported in controls. No significant effect reported in control Fv/Fm over the 35-day exposure period.
3	Is the test system defined and appropriate (flow-through, semi-static or static conditions)?	2	Semi-static exposure conditions with a 7-day renewal applied. Score reduced as renewal period exceeded 48 hours.
4	Is the test substance concentration maintained ± 20% throughout the exposure?	3	Measured concentrations are reported with an estimate of the 95% confidence interval. For the highest concentration, this interval is within 20% of mean, but at lower concentrations for both OC and AVO it is near or exceeds 20%. The frequency of sampling was not reported and differs by compound, reducing the score.
5	Is the test system appropriate for the test organism?	5	The temperature and light conditions are justified. Non-filtered seawater pumped from a 50 m depth was the exposure medium. The medium was constantly stirred with a pump. The suspended solids load is also not reported. There could be issues with the test medium stability, but this is assessed in question 21.
6	Biological effect stated?	3	The NOEC and LOEC are provided and therefore a MATC can be calculated from the data.

Table S3. Continued

Screening questions		Score	Comment
Is a parallel reference toxicant study conducted? Evaluate based on scenario:			
7a	If required in relevant guideline.		
7b	If studying a wild organism.		
7c	If studying a non-standard organism/ non-standard endpoint.	5	Diuron was included as a reference toxicant as it has been demonstrated to significantly reduce photosynthetic yield in previous coral experiments.
8	Test substance specifically identified (e.g., chemical abstract service [CAS] number) and source reported?	2	The CAS number is not provided; however, the correct structures are reported in Table 1 of the text. The source of the test compounds is not provided.
9	Test substance purity reported?	2	The purity of the substance is not reported; however, analytical monitoring was undertaken at the start of the test, reducing the importance of this parameter.
10	Is the experiment appropriately replicated and not pseudoreplicated?	2	The test setup in not clear in terms of how control aquaria were divided between negative and solvent controls and what constituted an experimental unit. Two tank replicates with three nubbins is inferred, but not explicitly defined. Certain compounds had four treatments rather than three, therefore the use of 6 tanks per treatment is unclear and therefore the score is reduced.
11	Is a significant concentration-response relationship demonstrated?	0	A dose-response cannot be determined for OC (only one treatment had a significant effect) and for AVO one treatment had a significant effect while the higher concentration resulted in 100% mortality and therefore a dose-response for photosynthetic yield cannot be established.
12	Was a suitable statistical method/model described to determine the toxicity?	3	Software used is reported. A *t* test was used to determine a significant difference. Only NOEC and LOEC derived.
13	Was significance level listed for NOEC/LOEC/MATC as 0.05 or less and for the NEC/LC/EC an estimate of variability reported?	3	The significance level was 0.05 for determination of the NOEC and LOEC.
14	Exposure duration stated? Is the duration appropriate considering the species, life stage and endpoint (acute or chronic)?	3	Exposure duration was 5 weeks, with a concentration renewal weekly. The endpoint, dark-adapted maximum quantum yield of PSII measured once a week. This is an appropriate length for a chronic test.
15	Is a suitable test concentration separation factor used?	1	A scaling factor of 10 and 5 used.
16	Do the test concentrations adequately bracket the biological endpoint?	3	Based on the design of the test, the bracketing is suitable for a LOEC (e.g., the lowest test concentration did not demonstrate a significant effect) and NOEC.
17	Are organisms appropriately acclimatized to test conditions?	3	After nubbins cut from colonies left to heal for 3 – 4 weeks when tissue was fully recovered. Kept in non-filtered sea water pumped from 50 m depth. Water from same source was used as the exposure medium.
18	Are organisms well described? (e.g., length, mass, age, strain, sex, etc.)	3	Adult fragments, 2 – 4 cm in size (total of 240 nubbins) from 10 genetically different colonies grown in the coral culture facility at the Centre Scientifique de Monaco.

Table S3. Continued.

Screening questions		Score	Comment
19	Test vessels appropriate for the test substance?	3	Glass aquaria used.
20	Are analytical methods described and appropriate QA/QC reported?	1	The analytical method is briefly described. No QA/QC, recovery, or LOQ.
Test medium parameters			
21a	Dissolved oxygen	0	
21b	Temperature	2	25˚C ± 0.5˚C
21c	pH	0	
21d	Salinity/conductivity	0	
21f	Species specific – include if specific parameters needed.		
22	If used, is solvent in the appropriate range?	2	Solvent used is in the acceptable range.
23	Is the solvent suitable for the test species?	2	Methanol is the solvent used.

Table S4. Screening and data quality assessment for the He et al. (2019a) study

Screening questions		Score	Comment
RQ1	Is the endpoint ecologically relevant?	Pass	Endpoints studied include mortality, bleaching, polyp retraction, and algal density. Mortality, bleaching (visual), and algal density are ecologically relevant endpoints. Polyp retraction is considered a non-ecologically relevant endpoint.
RQ2	Test organism relevant to the compartment, test compound and/or assessment?	Pass	Species studied include *S. caliendrum* and *P. damicornis* adults. Test compounds are octocrylene (OC) and octinoxate (EHMC). Also exposed as a mixture and sunscreen formula. The mixture and formula experiments are beyond the scope of this assessment.
RQ3	Was a negative control and solvent control (if necessary) at least duplicated?	Pass	A solvent and negative control were performed.
RQ4	Are ≥ 4 treatment concentrations included (including control) or experiment specifically designed as limit test?	Pass	Five treatment concentrations were used in the single chemical tests.
RQ5	Are endpoints based on measured concentrations if they deviate by ≥ 20% of the nominal concentration? If only nominal endpoints presented, is any analytical verification undertaken?	Fail	Test concentrations were analytically verified three times throughout the exposure: day zero, day one, and day seven. Concentrations deviate by greater than 20% of nominal by the end of the test. The significant drop in test concentrations resulted in different test treatment concentrations overlapping.
Data quality assessment			
1	Biological endpoint stated and defined.	10	All endpoints were stated and definitions referred to an earlier paper by the authors where the endpoints were sufficiently defined.
2	Are relevant validity criteria stated and met?	5	Validity criteria not stated. Expert judgement required to evaluate. Is control mortality reported? Significant differences between solvent and negative control? Effects in control? No mortality, bleaching or polyp retraction observed in controls. No significant differences reported between negative and solvent control.
3	Is the test system defined and appropriate (flow-through, semi-static or static conditions)?	2	Static test conditions were applied.
4	Is the test substance concentration maintained ± 20% throughout the exposure?	0	Based on the analytical data provided concentrations were not maintained at +/- 20% of the mean measured concentration throughout the test and strategies to reduce test concentration losses were not attempted.
5	Is the test system appropriate for the test organism?	3	Three nubbins suspended in glass bottles, volume of exposure medium not reported. Bottles kept in water bath to maintain temperature, aerated, and subject to 14:10 light:dark cycle and 200–300 μmol quanta m−2 s−1 irradiance. The test medium was not changed throughout the 7 day exposure period. There is no indication of flowing water, but aeration may have provided some agitation (score reduced).
6	Biological effect stated?	2	Only LOECs reported and not for all endpoints. For the rest of the endpoints no biological effect was observed.

Table S4. Continued

Screening questions		Score	Comment
Is a parallel reference toxicant study conducted? Evaluate based on scenario:			
7a	If required in relevant guideline.		
7b	If studying a wild organism.		
7c	If studying a non-standard organism/new endpoint.	0	A reference toxicant is not included.
8	Test substance specifically identified (e.g., chemical abstract service [CAS] number) and source reported?	4	EHMC was obtained from Sigma-Aldrich, OC was obtained from US Pharmacopeia Reference Standards. Neither are sold as salts, therefore having the CAS number in addition to this information is not necessary.
9	Test substance purity reported?	4	Purity for EHMC (98%) and OC (99%) is reported.
10	Is the experiment appropriately replicated and not pseudoreplicated?	0	Each bottle contained three nubbins and to get six replicate measurements, two fragments from each of the three glass bottles were used. Endpoints could be recalculated without pseudoreplication, but the data is not provided to do this.
11	Is a significant concentration-response relationship demonstrated?	0	A significant dose-response relationship is not observed. The authors report LOECs as concentration-response models can't be fit to the data. A definitive test based on these results is needed to determine reliable statistical endpoints.
12	Was a suitable statistical method/model described to determine the toxicity?	3	Normality and equality of error variances checked using SPSS. Non-normal data, a Kruskal-Wallis and Mann—Whitney U two sample test for means comparisons. Normal data, one-way ANOVA (dose) and two-way ANOVA (day, dose, interaction) followed by Duncan's tests for differences among controls and treatments (α=0.05). Only calculating NOEC and LOEC.
13	Was significance level listed for NOEC/LOEC/MATC as 0.05 or less and for the NEC/LC/EC an estimate of variability reported?	3	Duncan's tests for differences among controls and treatments (α=0.05)
14	Exposure duration stated? Is the duration appropriate considering the species, life stage and endpoint (acute or chronic)?	1	The exposure duration was seven days, this is a suitable duration for an acute test according to Warne et al. (2018). Acute endpoints suitable for ERA not reported (e.g., EC/LC50).
15	Is a suitable test concentration separation factor used?	1	A scaling factor of ten is used which is considered the maximum reliable spacing.
16	Do the test concentrations adequately bracket the biological endpoint?	3	No LOECs were observed at the lowest concentration tested.
17	Are organisms appropriately acclimatized to test conditions?	3	Fragments were cut and left to heal for one month. Next, they were introduced to the test system and left to acclimatize for two days prior to initiating the toxicant exposures.

Table S4. Continued

Screening questions		Score	Comment
18	Are organisms well described? (e.g., length, mass, age, strain, sex, etc.)	2	Fragment size is 2 – 3 g. Coral colonies collected in Hobihu, Kenting National Park in Southern Taiwan. It is not clear whether analysis was undertaken to ensure that the coral colonies were genetically different. This could impact the results as there was a low replication and replicates may not have been genetically different (reduce score as this was not reported).
19	Test vessels appropriate for the test substance?	3	Test vessels were reported to be glass which is acceptable for the test substances.
20	Are analytical methods described and appropriate QA/QC reported?	3	The analytical method is sufficiently detailed in the Supplementary file. Recoveries and LOD are reported. An LOQ was not calculated. The analytical method was reported in a monitoring paper (Tsui et al. 2014). Test methods are adequately described.
Test medium parameters			
21a	Dissolved oxygen	2	6.6 ± 0.5 mg/L.
21b	Temperature	2	25.3˚C ± 0.8˚C held in controlled water bath.
21c	pH	2	7.7 ± 0.2.
21d	Salinity/conductivity	2	35.7 ± 0.6‰, reports adding double-distilled water to maintain.
21f	Species specific – include if specific parameters needed.		
22	If used, is solvent in the appropriate range?	2	The solvent concentration is 0.01 µL/L methanol, which is an acceptable concentration.
23	Is the solvent suitable for the test species?	2	Methanol is an OECD acceptable solvent.

Table S5. Screening and data quality assessment for the He et al. (2019b) study. Two scores are provided for certain data quality assessment questions. This is to differentiate between the adult assays and the larval settlement assay. The score for the adult assay always appears first. If only one score is presented, it applies to both the adult and larval assay

Screening questions		Score	Comment
RQ1	Is the endpoint ecologically relevant?	Pass	Adult ecologically relevant endpoints: mortality, bleaching, algal density. Adult non-ecologically relevant endpoints: polyp retraction. Larvae ecologically relevant endpoints: settlement, mortality, and bleaching. The assessment will be split where necessary to accommodate the two tests for the different life stages.
RQ2	Test organism relevant to the compartment, test compound and/or assessment?	Pass	Species studied include *S. caliendrum* and *P. damicornis* in larval and adult life stages. Exposure to benzophenones (BP1, BP3, BP4, BP8). BP1 results are not included in assessment as BP1 is not used as a UV filter in the United States or globally.
RQ3	Was a negative control and solvent control (if necessary) at least duplicated?	Pass	A solvent and negative control were included.
RQ4	Are ≥ 4 treatment concentrations included (including control) or experiment specifically designed as limit test?	Pass	Five treatment concentrations were included in the adult and larval tests.
RQ5	Are endpoints based on measured concentrations if they deviate by ≥ 20% of the nominal concentration? If only nominal endpoints presented, is any analytical verification undertaken?	Fail	Test concentrations were only analytically verified at the beginning and the end of the experiment. Furthermore, by the end of the 7 d and 14 d duration, test concentrations were analytically indistinguishable as multiple test concentrations fell below the limit of detection (LOD).
Data quality assessment			
1	Biological endpoint stated and defined.	10	All endpoints were stated and defined for both larvae and adults.
2	Are relevant validity criteria stated and met?	5	Validity criteria not stated. Expert judgement required: Is control mortality reported? Does the solvent control significantly different from negative control? Are endpoint responses observed in controls? Authors report that no control treatment showed significant impacts on coral larvae and nubbins and UV filters were not detected in controls. The solvent control results were not significantly different from negative control. No mortality reported in controls. Previously reported larval settlement experiments state that the % settlement in the control needs to be reported. The authors report this as 100%, but this is not consistent with data reported in Fig. S8 of the supplementary information where settlement rates range from 90-100%.
3	Is the test system defined and appropriate (flow-through, semi-static or static conditions)?	2	Static conditions applied in both the 7 day (adults) and 14 day tests (larvae).
4	Is the test substance concentration maintained ± 20% throughout the exposure?	0	Based on the analytical data provided concentrations were not maintained at +/- 20% of the mean measured concentration throughout the test and strategies to reduce test concentration losses were not attempted.

Table S5. Continued

Screening questions		Score	Comment
5	Is the test system appropriate for the test organism?	3	For adults, four nubbins suspended in glass beakers with 800 mL unfiltered seawater. Bottles kept in water bath to maintain temperature. 14:10 light:dark cycle and 200–300 µmol quanta m−2 s−1 irradiance and aeration using filtered air also provided. There does not appear to be agitated/flowing water and the exposure medium is not renewed over the 7 day exposure. The authors report reconstituting the test medium twice with double distilled water. This suggests it was challenging to maintain the test system and this may have affected the results. Larval tests were conducted in petri dishes with 10 mL of filtered seawater without aeration. Lower light intensity used than for adult tests (80 µmol quanta $s^{-1}m^{-2}$) because of rapid evaporation of seawater. The impact of the lower irradiance is unknown and is simply noted at this time. The exposure medium was not renewed for 14 days.
6	Biological effect stated?	2, 5	Adults: Only LOECs are reported for a few endpoints. For the rest of the endpoints no biological effect was observed. For larvae, an EC50 is reported for larval settlement, *S. caliendrum* exposure to BP-8 only (5 points).
Is a parallel reference toxicant study conducted? Evaluate based on scenario:			
7a	If required in relevant guideline.		
7b	If studying a wild organism.		
7c	If studying a non-standard organism/ non-standard endpoint.	0	A reference toxicant is not included.
8	Test substance specifically identified (e.g., chemical abstract service [CAS] number) and source reported?	4	BP-3 was obtained from US Pharmacopeia Reference Standards and BP-4 and BP-8 were obtained from Sigma-Aldrich. Neither are sold as salts, therefore having the CAS number in addition to this information is not necessary.
9	Test substance purity reported?	4	The purity for each of the test substances is reported, BP-3 (100%), BP-4 (97%), BP-8 (98%).
10	Is the experiment appropriately replicated and not pseudoreplicated?	0, 4	Adults: The individual nubbin is the experimental unit, but two nubbins per endpoint were exposed in the same bottle. Therefore, the adult experiment was pseudoreplicated and data is not reported to re-calculate endpoints (0 points awarded). Larvae: Each petri dish is the experimental unit (10 larvae), this experiment was appropriately replicated (4 petri dishes per treatment). The replication of the solvent control and negative control is unclear, but the authors state ‘blank controls’ which could be interpreted as more than one replicate (awarded 4 points).
11	Is a significant concentration-response relationship demonstrated?	0, 4	For adults,a dose-response relationship was not observed. Certain endpoints went from 0% to 100% response at the highest concentration indicating the spacing between treatments was too large, while for others no effect was observed at any concentration. This meant that for several endpoints (e.g., mortality, bleaching, zooxanthellae density) a LOEC resulted in 100% effect or a true NOEC could not be derived (no effect observed at highest test concentration, awarded zero points). Larvae settlement assays, a dose-response was observed and was significant (only for BP-8, awarded 4 points).

Table S5. Continued

Screening questions		Score	Comment
12	Was a suitable statistical method/model described to determine the toxicity?	3	For adults, normality and equality of error variances checked using SPSS. Non-normal data, a Kruskal-Wallis and Mann—Whitney U two sample test for means comparisons. Normal data, one-way ANOVA (dose) and two-way ANOVA (day, dose, interaction) followed by Duncan's tests for differences among controls and treatments (α=0.05), only calculated LOEC, NOEC. Larval settlement included enough treatments to calculate EC50, third-order polynomial equations of the best-fit curves using Graphpad Prism.
13	Was significance level listed for NOEC/LOEC/MATC as 0.05 or less and for the NEC/LC/EC an estimate of variability reported?	3, 0	For adults the significance level of the LOECs reported is 0.05 (awarded 3 points). For the larval settlement test, an estimate of variability is not provided (0 points awarded).
14	Exposure duration stated? Is the duration appropriate considering the species, life stage, endpoint and reported effect concentration (acute or chronic)?	1	The exposure duration for larvae was 14 days and for adults it was 7 days. According to Warne et al. (2018), larval settlement assay is chronic and the adult assays are acute. Therefore, in the larval test an NEC/EC10 should have also been reported and EC(LC)50 reported in the adult tests.
15	Is a suitable test concentration separation factor used?	1, 3	Adults: a scaling factor of 10 is used which is considered the maximum reliable spacing (1 point awarded). Larvae: Treatment spacing was $\leq$ 10 (3 points awarded).
16	Do the test concentrations adequately bracket the biological endpoint?	3	Adults: The LOEC was not observed at the lowest concentration tested. Larvae: The test concentrations adequately bracketed the EC50.
17	Are organisms appropriately acclimatized to test conditions?	3	Adult fragments were cut and left to heal for 1 month. Next, they were introduced to the test system and left to acclimatize for two days prior to initiating the toxicant exposures. Larvae were collected and washed once with pre-filtered seawater. An acclimation period is not included and the relevance or length of an acclimation period for this type of assay is unclear, particularly as the exposure medium was the same during the test and in the spawning tanks.
18	Are organisms well described? (e.g., length, mass, age, strain, sex, etc.)	2	Adults: Fragment size 2 – 3 g. Larvae: collected using a sieve (100 µm) during the period of lunar larval release from organisms maintained in the in the laboratory. Larvae collected before 9:00 am and washed with pre-filtered (0.2 µm) natural seawater prior to exposure. Coral colonies collected in Hobihu, Kenting National Park, Nanwan in Southern Taiwan. It is not clear whether analysis was undertaken to ensure that the coral colonies were genetically different. This could impact the results as there was a low replication in the adult experiment and replicates may not have been genetically different (reduce score as this was not reported).
19	Test vessels appropriate for the test substance?	3	Test vessels were glass (jars for adults, petri dished for larvae).
20	Are analytical methods described and appropriate QA/QC reported?	3	The analytical method is sufficiently detailed in the Supplementary file. Recoveries and LOD are reported. An LOQ was not calculated. The analytical method was reported in a previous paper Tsui et al. (2014).

Table S5. Continued

Screening questions		Score	Comment
Test medium parameters			
21a	Dissolved oxygen	2	6.5 ± 0.3 mg/L in adult test and 2.2 ± 0.1 mg/L in larval.
21b	Temperature	2	25.2 ˚C ± 0.8˚C in adult tests, 25.6 ± 0.3˚C in larval tests.
21c	pH	2	7.7 ± 0.1 in both adult and larval tests.
21d	Salinity/conductivity	2	35.9 ± 0.5‰ in both adult and larval tests.
21f	Species specific – include if specific parameters needed.		
22	If used, is solvent in the appropriate range?	0	In larvae tests, 0.2 mL/L methanol used as solvent. This exceeds the 0.1 mL/L acceptable solvent threshold. In adult tests, the solvent content is assumed to be comparable (same stock solution used) but this cannot be confirmed as it is not reported.
23	Is the solvent suitable for the test species?	2	Methanol is an OECD acceptable solvent.

Table S6. Screening and data quality assessment for the McCoshum et al. (2016) study

Screening questions		Score	Comment
RQ1	Is the endpoint ecologically relevant?	Pass	The endpoint studied is growth which is ecologically relevant.
RQ2	Test organism relevant to the compartment, test compound and/or assessment?	Fail	Studied the effects of a sunscreen formulation on the pulse coral, *Xenia* sp. (soft coral). Note that a single colony is used for the study which means that inter-colony variation within the species is not captured in this study. The coral is exposed to a sunscreen formula that contained multiple UV filters and other ingredients (none of which were analytically characterized or concentrations confirmed). Therefore, effects observed cannot be attributed to any particular UV filter and the test substance is out of scope to the problem formulation.
RQ3	Was a negative control and solvent control (if necessary) at least duplicated?	Pass	A negative control is included. A solvent control is not necessary as solvent was not used (do not evaluate questions 22 and 23).
RQ4	Are ≥ 4 treatment concentrations included (including control) or experiment specifically designed as limit test?	Fail	A single test concentration is used and the test is not designed as a limit test. An effect was observed; therefore, a NOEC cannot be derived.
RQ5	Are endpoints based on measured concentrations if they deviate by ≥ 20% of the nominal concentration? If only nominal endpoints presented, is any analytical verification undertaken?	Fail	No analytical verification of the test concentration was undertaken. This is particularly important for this study which dosed using a sunscreen formulation and did not analytically confirm the UV filter concentrations. Exposure is based on the 'sunscreen' rather than concentrations of the individual sunscreen components.
Data quality assessment			
1	Biological endpoint stated and defined.	10	The endpoint (growth) is stated and and how it is quantified was sufficiently reported.
2	Are relevant validity criteria stated and met?	0	Expert judgement required to determine validity criteria. Is control mortality reported? Are endpoint responses observed in controls? As the endpoint is growth, it should be established what an acceptable level of growth is and whether is it achieved within the control. This was not investigated. Initial polyp numbers is also not reported (how growth was assessed). Control and treatment mortality are not reported. Validity criteria for this study would enhance reliability (e.g., how much growth in control should be expected, coefficient of variance of control growth).
3	Is the test system defined and appropriate (flow-through, semi-static or static conditions)?	2	Static-test conditions applied (72 hours exposed to sunscreen formulation followed by 28 days in aquaria without the test substance).
4	Is the test substance concentration maintained ± 20% throughout the exposure?	0	The exposure concentrations were not explicitly reported or verified.

Table S6. Continued

Screening questions		Score	Comment
5	Is the test system appropriate for the test organism?	5	After the 72 hour exposure, nubbins were placed in open plastic containers under a 14:10 h photoperiod. Aquariums were floated in a 26˚C water bath and aerated with pumps and air stones, dissolved oxygen was not measured. This aeration is expected to provide some agitation of flow for the coral. The pH, photoperiod, light intensity is reported. The artificial saltwater media is expected to contain the appropriate ions and nutrients for this species.
6	Biological effect stated?	0	No biological effect is stated in this experiment (e.g., NOEC/EC10).
Is a parallel reference toxicant study conducted? Evaluate based on scenario:			
7a	If required in relevant guideline.		
7b	If studying a wild organism.		
7c	If studying a non-standard organism/ non-standard endpoint.	0	A reference toxicant was not included.
8	Test substance specifically identified (e.g., chemical abstract service [CAS] number) and source reported?	4	The test results are reported based on exposure to a whole sunscreen formulation. UV filters are identified by INCI name on the product, Equate 50 SPF sunscreen.
9	Test substance purity reported?	0	The purity was not reported as this is a 'whole sunscreen' experiment. The % of each UV filter ingredient in the formulation is reported, but not calculated or analytically confirmed how much of each UV filter coral were exposed too.
10	Is the experiment appropriately replicated and not pseudoreplicated?	0	Reported number of replicates was 140. Pseudoreplication used (e.g., 4 'nubbins' per container, therefore *n* should be 35). Data not provided to re-calculate. It is unclear whether the initial 72 h exposure to sunscreen occurred all in the same aquarium, this would be inappropriate.
11	Is a significant concentration-response relationship demonstrated?	0	This is a single treatment study; therefore, establishing a dose-response cannot be achieved with the experimental design. An effect was observed and therefore no points are awarded.
12	Was a suitable statistical method/model described to determine the toxicity?	0	The t-test described is suitable as only one test variable and treatment concentration used (growth, determined as number of polyps for a single treatment); however, the toxicity was not reported (e.g., LOEC) and this is likely due to the inappropriate experimental design for deriving statistical endpoints for ERA.
13	Was significance level listed for NOEC/LOEC/MATC as 0.05 or less and for the NEC/LC/EC an estimate of variability reported?	0	The significance level is not reported.
14	Exposure duration stated? Is the duration appropriate considering the species, life stage, endpoint and reported effect concentration (acute or chronic)?	1	Coral nubbins were exposed to the sunscreen treatment for 72 h, then cleaned and placed in new containers for 28 d. Polyp numbers were counted every 3 - 4 days to estimate growth. The length of the growth period is consistent with a chronic test (28 days), but the exposure was acute (72 h).

Table S6. Continued

Screening questions		Score	Comment
15	Is a suitable test concentration separation factor used?	0	Only a single test concentration used.
16	Do the test concentrations adequately bracket the biological endpoint?	0	An effect was observed at the lowest concentration tested (only one concentration); therefore, not adequate bracketing.
17	Are organisms appropriately acclimatized to test conditions?	3	Coral nubbins detached from colony and placed in a new aquarium for 96 h (permits attachment to surface prior to exposure). It is unclear whether this is long enough to heal as this is a soft coral. For hard corals, closer to a month is appropriate.
18	Are organisms well described? (e.g., length, mass, age, strain, sex, etc.)	0	The size of coral nubbins is not reported or how many polyps each individual had at the start of the experiment. This is important information as growth (in terms of new polyps) is the studied endpoint. All nubbins were collected from the same colony.
19	Test vessels appropriate for the test substance?	3	The test aquarium was glass, which is appropriate.
20	Are analytical methods described and appropriate QA/QC reported?	0	No analytical method was reported.
Test medium parameters			
21a	Dissolved oxygen	0	
21b	Temperature	1	26˚C.
21c	pH	0	
21d	Salinity/conductivity	0	
21f	Species specific – include if specific parameters needed.		
22	If used, is solvent in the appropriate range?		No solvent.
23	Is the solvent suitable for the test species?		No solvent.

Table S7. Screening and data quality assessment for the Stien et al. (2019) study

Screening questions		Score	Comment
RQ1	Is the endpoint ecologically relevant?	Fail	Metabolomic profiling was the main effect studied and polyp retraction was a secondary endpoint. Both are non-ecologically relevant endpoints.
RQ2	Test organism relevant to the compartment, test compound and/or assessment?	Pass	Test compound is octocrylene (OC) and the test species is the scleractinian coral *Pocillopora damicornis*. Note that a single colony is used for the study, which means that inter-colony variation within the species is not captured in this study.
RQ3	Was a negative control and solvent control (if necessary) at least duplicated?	Fail	A solvent control was included, but a negative control was not included.
RQ4	Are ≥ 4 treatment concentrations included (including control) or experiment specifically designed as limit test?	Pass	Four test treatment concentrations were included.
RQ5	Are endpoints based on measured concentrations if they deviate by ≥ 20% of the nominal concentration? If only nominal endpoints presented, is any analytical verification undertaken?	Fail	Analytical verification was not undertaken to verify exposure concentrations.
Data quality assessment			
1	Biological endpoint stated and defined.	5	A description of how polyp retraction was identified and quantified was not included. An observation that 'most polyps were closed' reported. Metabolic profiling was adequately described.
2	Are relevant validity criteria stated and met?	0	This is not a standard test and no validity criteria are provided. Expert judgement required to determine validity criteria. Is mortality reported in the controls? Is there a significant difference between negative and solvent controls? Is an effect observed in the control? Control mortality is not reported. Without negative control cannot assess whether significant differences between negative and solvent control occurred, this is particularly important as DMSO was used as the solvent. For the metabolic profiling results are presented relative to the control so inclusion of a negative control would be useful.
3	Is the test system defined and appropriate (flow-through, semi-static or static conditions)?	3	Semi-static flow conditions were applied. The test medium was renewed every 24 hours.
4	Is the test substance concentration maintained ± 20% throughout the exposure?	0	No analytical monitoring to determine whether test concentrations were maintained was undertaken during the study.
5	Is the test system appropriate for the test organism?	3	Corals exposed in and maintained in artificial sea for over a year prior to the study. Lighting set to 250 µmol/m²/s and a 10:14 h photoperiod used. Beakers were aerated by gentle bubbling. It is unlikely that this oxygenation bubbling would have agitated the water sufficiently (similarly to He et al. 2019a,b) reducing the score.
6	Biological effect stated?	0	There is no biological effect stated (e.g., NOEC or LOEC).

Table S7. Continued

Screening questions		Score	Comment
Is a parallel reference toxicant study conducted? Evaluate based on scenario:			
7a	If required in relevant guideline.		
7b	If studying a wild organism.		
7c	If studying a non-standard organism/ non-standard endpoint.	0	A reference toxicant was not included.
8	Test substance specifically identified (e.g., chemical abstract service [CAS] number) and source reported?	4	CAS number is not provided in the text, but sufficient physicochemical information is provided (e.g., high resolution *m/z* information). The source is Sigma Aldrich.
9	Test substance purity reported?	4	The purity is reported as 98%.
10	Is the experiment appropriately replicated and not pseudoreplicated?	0	Pseudoreplication was used as there were three nubbins per beaker and only 1 beaker per treatment.
11	Is a significant concentration-response relationship demonstrated?	0	A dose-response relationship was not reported for polyp retraction. A dose-response is reported to be observed in the volcano plots for the metabolomic analysis but significance between treatments is not reported, only compared to control.
12	Was a suitable statistical method/model described to determine the toxicity?	0	No statistical or quantitative method was applied for polyp retraction. Just reported that 'most polyps seemed to be closed' and therefore polyps appeared to be impacted. Volcano plots were used demonstrate which ions were up-regulated but a model to demonstrate a concentration-response not applied.
13	Was significance level listed for NOEC/LOEC/MATC as 0.05 or less and for the NEC/LC/EC an estimate of variability reported?	0	No significance level applied for polyp retraction. The significance level for volcano plots was 0.05, but a statistical endpoint was not reported (e.g., NOEC/LOEC).
14	Exposure duration stated? Is the duration appropriate considering the species, life stage, endpoint and reported effect concentration (acute or chronic)?	3	The exposure duration was seven days. Coral polyps were observed every 24 hours.
15	Is a suitable test concentration separation factor used?	3	The separation factor ranged from three to ten.
16	Do the test concentrations adequately bracket the biological endpoint?	0	A biological endpoint is not reported by the authors to bracket.
17	Are organisms appropriately acclimatized to test conditions?	3	Wild corals were collected in Oman and maintained in tanks at the Banyuls Oceanological Observatory for more than a year.
18	Are organisms well described? (e.g., length, mass, age, strain, sex, etc.)	3	Coral nubbins cut from branch tips of same mother colony (1-1.5 cm). Left to acclimatize for 1 month after cutting from mother colony.
19	Test vessels appropriate for the test substance?	3	Glass beakers used for test which is appropriate.
20	Are analytical methods described and appropriate QA/QC reported?	0	Quantitative analytical method for concentration verification not included. The analytical method for analyzing the up-regulation of ions is reported.

Table S7. Continued

Screening questions		Score	Comment
Test medium parameters			
21a	Dissolved oxygen	0	
21b	Temperature	1	Reported as 24 ˚C.
21c	pH	1	Reported as 8.
21d	Salinity/conductivity	1	Reported as 36 g/L.
21f	Species specific – include if specific parameters needed.		
22	If used, is solvent in the appropriate range?	0	The solvent concentrations exceed generally acceptable levels (e.g. 0.25% compared to 0.01%). Based on the information presented, DMSO concentrations were 0.5 mL/ per 200 mL or 2.5mL/L, which far exceeds acceptable levels
23	Is the solvent suitable for the test species?	1	DMSO used as the carrier solvent which reduces the score as safe use of this solvent for coral has not been demonstrated. This is further complicated by the lack of negative control and very high concentration.

Table S8. Screening and data quality assessment for the Wijgerde et al. (2020) study

Screening questions		Score	Comment
RQ1	Is the endpoint ecologically relevant?	Pass	The endpoints are mortality (reported as survival), growth rate, photosynthetic yield, zooxanthellae density and changes in the microbiome. Mortality and growth are ecologically relevant, and zooxanthellae density is an acceptable indication of bleaching. The ecological relevance of the sub-lethal effects photosynthetic yield and microbiome changes is unclear and therefore both are considered non-ecologically relevant.
RQ2	Test organism relevant to the compartment, test compound and/or assessment?	Pass	Studied reef-building corals, *Stylophora pistillata* and *Acropora tenuis*. Exposure to the organic UV filter, oxybenzone (BP-3). In addition to exposure to BP-3, the effects of a heatwave were also studied. This is beyond the scope of the current assessment and only the exposure of BP-3 to the two coral species at initial temperature is considered. The experiment was conducted on adults all from the same colony meaning that variation between colonies of the same species is not captured in this experiment.
RQ3	Was a negative control and solvent control (if necessary) at least duplicated?	Fail	A solvent and temperature control were included. A negative control was not included.
RQ4	Are ≥ 4 treatment concentrations included (including control) or experiment specifically designed as limit test?	Fail	The test consisted of a single treatment, exposure to 1 µg/L BP-3 (nominal) and the study is not designed as a limit test.
RQ5	Are endpoints based on measured concentrations if they deviate by ≥ 20% of the nominal concentration? If only nominal endpoints presented, is any analytical verification undertaken?	Pass	Analytical verification was undertaken. Samples were collected at 12 points throughout the exposure (roughly 3 - 4 days). Exposures were reported based on measured concentration as it deviated by > 20% of nominal.
Data quality assessment			
1	Biological endpoint stated and defined.	10	The endpoints are stated and clearly defined.
2	Are relevant validity criteria stated and met?	0	Expert judgement required. Is control mortality reported? Does the negative and solvent control mortality/growth differ? Is a significant effect reported in controls? For *A. tenuis*, 67% survival at 26˚C (solvent control), indicating >30% mortality rate in control for that species. This was phenomena was not observed for *S. pistillata*. Without a negative control cannot assess whether significant differences between negative and solvent control occurred.
3	Is the test system defined and appropriate (flow-through, semi-static or static conditions)?	5	The test system is an intermittent flow-through design. A header tank continuously supplies fresh test medium that has already been dosed with BP-3. The header tank is refreshed every 48 hours.
4	Is the test substance concentration maintained ± 20% throughout the exposure?	3	The authors report a BP-3 concentration of 0.05 ± 0.03 µg/L (s.e.m), a graph of the concentrations is also provided. The authors state that the 48-hour renewal was expected to be enough to maintain the concentration. Points are awarded for including a strategy to maintain test concentrations; however, concentrations varied by over 20% during the 40 d exposure.

Table S8. Continued

Screening questions		Score	Comment
5	Is the test system appropriate for the test organism?	5	The 'flow-through' system provided water flow. Artificial seawater was used and nubbins kept in the artificial sweater for 4 weeks prior to exposure so unintended stress is not likely. Test system well characterized and justified.
6	Biological effect stated?	0	A biological effect is not stated (e.g., NOEC or LOEC).
Is a parallel reference toxicant study conducted? Evaluate based on scenario:			
7a	If required in relevant guideline.		
7b	If studying a wild organism.		
7c	If studying a non-standard organism/ non-standard endpoint.	0	A reference toxicant is not included.
8	Test substance specifically identified (e.g., chemical abstract service [CAS] number) and source reported?	4	The CAS number is not provided; however, *m/z* transitions from the MS/MS method are provided which are sufficient evidence the correct compound was studied. Source is reported as Sigma Aldrich.
9	Test substance purity reported?	2	The purity of the substance is not reported; however, analytical monitoring was undertaken at throughout the test, reducing the importance of this parameter.
10	Is the experiment appropriately replicated and not pseudoreplicated?	4	The *A. tenuis* were pooled (three per tank) for a total of five replicates per treatment (five tanks). Each aquarium was an experimental unit. There was only a single *S. pistillata* nubbin per tank. There were four total treatments. In total n = 20 and n = 60 nubbins of *S. pistillata* and *A. tenuis* were used respectively.
11	Is a significant concentration-response relationship demonstrated?	0	This is a single treatment test; therefore, establishing a dose-response was not part of the design; however, as an effect was observed this criterion is evaluated.
12	Was a suitable statistical method/model described to determine the toxicity?	0	The software used was reported and the statistics were described in sufficient detail. Normality and homogeneity of variance checked. A two-way factorial ANOVA was used to determine the effects of oxybenzone exposure on specific growth rate and zooxanthellae density; however, the toxicity was not reported (e.g., LOEC) and this is likely due to the inappropriate experimental design for deriving statistical endpoints for ERA.
13	Was significance level listed for NOEC/LOEC/MATC as 0.05 or less and for the NEC/LC/EC an estimate of variability reported?	0	The significance level is reported as 0.05, but a statistical endpoint (e.g., NOEC/LOEC) was not reported due to the design of the study, therefore this criterion is not fulfilled.
14	Exposure duration stated? Is the duration appropriate considering the species, life stage, endpoint and reported effect concentration (acute or chronic)?	3	The exposure duration was six weeks. Corals weighed on day 0, 16 and 41 to determine growthrate. This is a suitable duration for a chronic test.
15	Is a suitable test concentration separation factor used?	0	Only a single test concentration reported.
16	Do the test concentrations adequately bracket the biological endpoint?	0	Biological endpoint was not derived due to the test design (single test treatment).

Table S8. Continued

Screening questions		Score	Comment
17	Are organisms appropriately acclimatized to test conditions?	3	Fragments were cut from the parent colony, glued to PVC tiles and left to recover in a 550 L holding tank for four weeks.
18	Are organisms well described? (e.g., length, mass, age, strain, sex, etc.)	3	Adult fragments, 1 cm in length. All from a single parent colony.
19	Test vessels appropriate for the test substance?	0	Plastic test vessels and tubing was used. The authors also hypothesized this was a reason for significant losses.
20	Are analytical methods described and appropriate QA/QC reported?	3	The analytical method is described in sufficient detail. Recoveries, LOQ and LOD are provided.
Test medium parameters			
21a	Dissolved oxygen	0	
21b	Temperature	2	26˚C reported and maintained.
21c	pH	2	7.8 – 8.2 reported and maintained
21d	Salinity/conductivity	2	35 g/L reported and maintained.
21f	Species specific – include if specific parameters needed.		
22	If used, is solvent in the appropriate range?	2	Solvent is 0.001 % v/v, this is an acceptable level.
23	Is the solvent suitable for the test species?	1	DMSO was used and this reduces the score because a non-standard test organism was studied.

Literature Cited

Danovaro R, Bongiorni L, Corinaldesi C, Giovannelli D, Damiani E, Astolfi P, Greci L, Pusceddu A. 2008. Sunscreens cause coral bleaching by promoting viral infections. *Environ Health Perspect*. 116(4):441–447. doi:10.1289/ehp.10966.

Downs CA, Kramarsky-Winter E, Fauth JE, Segal R, Bronstein O, Jeger R, Lichtenfeld Y, Woodley CM, Pennington P, Kushmaro A, et al. 2014. Toxicological effects of the sunscreen UV filter, benzophenone-2, on planulae and in vitro cells of the coral, *Stylophora pistillata*. E*cotoxicology*. 23(2):175–191. doi:10.1007/s10646-013-1161-y.

Downs CA, Kramarsky-Winter E, Segal R, Fauth J, Knutson S, Bronstein O, Ciner FR, Jeger R, Lichtenfield Y, Woodley CM, et al. 2016. Toxicopathological effects of the sunscreen UV filter, oxybenzone (benzophenone-3), on coral planulae and cultured primary cells and Its environmental contamination in Hawaii and the U.S. Virgin Islands. *Arch Environ Contam Toxicol*. 70(2):265–288. doi:10.1007/s00244-015-0227-7.

Fel JP, Lacherez C, Bensetra A, Mezzache S, Béraud E, Léonard M, Allemand D, Ferrier-Pagès C. 2019. Photochemical response of the scleractinian coral *Stylophora pistillata* to some sunscreen ingredients. *Coral Reefs*. 38(1):109–122. doi:10.1007/s00338-018-01759-4.

Epstein N, Bak RPM, Rinkevich B. 2000. Toxicity of third generation dispersants and dispersed Egyptian crude oil on Red sea coral larvae. *Mar. Pollut. Bull*. 40(6): 497 – 503. https://doi.org/10.1016/S0025-326X(99)00232-5

He T, Tsui MMP, Tan CJ, Ma CY, Yiu SKF, Wang LH, Chen TH, Fan TY, Lam PKS, Murphy MB. 2019. Toxicological effects of two organic ultraviolet filters and a related commercial sunscreen product in adult corals. *Environ Pollut.* 245:462–471. doi:10.1016/j.envpol.2018.11.029.

He T, Tsui MMP, Tan CJ, Ng KY, Guo FW, Wang LH, Chen TH, Fan TY, Lam PKS, Murphy MB. 2019. Comparative toxicities of four benzophenone ultraviolet filters to two life stages of two coral species. *Sci Total Environ.* 651:2391–2399. doi:10.1016/j.scitotenv.2018.10.148.

McCoshum SM, Schlarb AM, Baum KA. 2016. Direct and indirect effects of sunscreen exposure for reef biota. *Hydrobiologia.* 776(1):139–146.

Stien D, Clergeaud F, Rodrigues AMS, Lebaron K, Pillot R, Romans P, Fagervold S, Lebaron P. 2019. Metabolomics reveal that octocrylene accumulates in *Pocillopora damicornis* tissues as fatty acid conjugates and triggers coral cell mitochondrial dysfunction. *Anal Chem*. 91(1):990–995. doi:10.1021/acs.analchem.8b04187.

Tsui MMP, Leung HW, Wai TC, Yamashita N, Taniyasu S, Liu W, Lam PKS, Murphy MB. 2014. Occurrence, distribution and ecological risk assessment of multiple classes of UV filters in surface waters from different countries. *Water Res*. 67:55–65. doi:10.1016/j.watres.2014.09.013.

Warne M, Batley G, van Dam R, Chapman J, Fox D, Hickey C, Stauber JL. 2018. Revised method for deriving Australian and New Zealand water quality guideline values for toxicants.

Prepared for the Revision of the Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Australian and New Zealand Governments and Australian State and Territory Governments, Canberra, Australia, 48 pp.

Wijgerde T, van Ballegooijen M, Nijland R, van der Loos L, Kwadijk C, Osinga R, Murk A, Slijkerman D. 2020. Adding insult to injury: Effects of chronic oxybenzone exposure and elevated temperature on two reef-building corals. *Sci Total Environ*. 733:139030. doi:10.1016/j.scitotenv.2020.139030.