Application of systematic evidence mapping to assess the impact of new research when updating health reference values: A case example using acrolein

Channa Keshavaa, J. Allen Davisa, John Staneka, Kristina A. Thayera, Audrey Galiziaa, Nagalakshmi Keshavaa, Jeff Giftb, Suryanarayana V. Vulimiric, George Woodalla, Carolyn Gigota, Kelly Garciaa, Andrew Greenhalgha, Brittany Schulza, Savannah Volkoffb, Krisa Camargoc, Amanda S. Persada,*

aaCenter for Public Health and Environmental Assessment, US EPA, NC, USA
bbPratt School of Engineering, Duke University, Durham, NC, USA
cVeterinary Integrative Biosciences and Geochemical Environmental Research Group, Texas A&M University, College Station, TX, USA

Abstract

Background: The environmental health community needs transparent, methodologically rigorous, and rapid approaches for updating human health risk assessments. These assessments often contain reference values for cancer and/or noncancer effects. Increasingly, the use of systematic review methods are preferred when developing these assessments. Systematic evidence maps are a type of analysis that has the potential to be very helpful in the update process, especially when combined with machine-learning software advances designed to expedite the process of conducting a review.

Objectives: To evaluate the applicability of evidence mapping to determine whether new evidence is likely to result in a change to an existing health reference value, using inhalation exposure to the air pollutant acrolein as a case example.
Methods: New literature published since the 2008 California Environmental Protection Agency’s Office of Environmental Health Hazard Assessment (OEHHA) Reference Exposure Level (REL) for acrolein was assessed. Systematic review methods were used to search the literature and screening included the use of machine-learning software. The Populations, Exposures, Comparators and Outcomes (PECO) criteria were kept broad to identify studies that characterized acute and chronic exposure and could be informative for hazard characterization. Studies that met the PECO criteria after full-text review were briefly summarized before their suitability for chronic point of departure (POD) derivation and calculation of a reference value was considered. Studies considered potentially suitable underwent a targeted evaluation to determine their suitability for use in dose–response analysis.

Results: Over 15,000 studies were identified from scientific databases. Both machine-learning and manual screening processes were used to identify 60 studies considered PECO-relevant after full-text review. Most of these PECO-relevant studies were short-term exposure animal studies (acute or less than 1 month of exposure) and considered less suitable for deriving a chronic reference value when compared to the subchronic study in rats used in the 2008 OEHHA assessment. Thirteen epidemiological studies were identified but had limitations in the exposure assessment that made them less suitable for dose–response compared to the subchronic rat study. Among the 13 studies, there were four controlled trial studies that have the potential to be informative for future acute reference value derivation. Thus, the 2008 subchronic rat study used by OEHHA appears to still be the most appropriate study for chronic reference value derivation. In addition, advances in dosimetric modeling for gases, including new evidence pertinent to acrolein, could be considered when updating existing acrolein toxicity values.

Conclusions: Evidence mapping is a very useful tool to assess the need for updating an assessment based on understanding the potential impact of new studies on revising an existing health reference value. In this case example, the focus was to identify studies suitable for chronic exposure dose–response analysis, while also identifying studies that may be important to consider for acute exposure scenarios, hazard identification, or for future research. This allows the evidence map to be a useful resource for a range of decision-making contexts. Specialized systematic review software increased the efficiency of the process in terms of human resources and time to conduct the analysis.

Keywords
Risk assessment; Systematic review; Hazard characterization; Hazardous air pollutant; Evidence map

1. Introduction
Systematic evidence maps (SEMs, also referred to as systematic maps or evidence maps) are gaining visibility in environmental health for their utility to inform decision-making and risk management priority setting and to serve as problem formulation tools to refine the focus of questions that get addressed in full systematic reviews (Wolffe et al., 2019). Although the definition of an evidence map has not always been clear or consistently used (Miake-Lye et al., 2016; Polisena et al., 2015; Khangura et al., 2014; Brugge et al., 2011), one definition that captures current sentiments is “A comprehensive summary of the characteristics and...
availability of evidence as it relates to broad issues of policy or management relevance. Systematic maps do not seek to synthesize evidence but instead to catalogue it, utilizing systematic search and selection strategies to produce searchable databases of studies along with detailed descriptive information” (Elsevier, 2017) (https://www.elsevier.com/journals/environment-international/0160-4120/guidance-notes). SEMs may include critical appraisal of studies, but there is no attempt to synthesize the evidence to answer an assessment question. Increasingly, journals are providing author guidance on publication of evidence maps, including environmental journals (CEE, 2019; EHP, 2019; EI, 2019).

In the context of priority setting, one application of an SEM could be to determine whether new evidence is likely to result in a change to an existing health reference value. This would be a very targeted application of an SEM, going beyond providing a broad survey of the available evidence to identifying a subset of studies that have the study design features considered most suitable for consideration in a regulatory risk assessment context. We explored this application using chronic inhalation exposure to acrolein as a case example. To take advantage of existing knowledge, we focused on the new literature published since the most recent assessment released by a Federal or state health agency that underwent a public comment period and external peer-review. We consider this an important application of SEM methodology because conducting chemical assessments is a very resource intensive process and they quickly become outdated based on year of release; thus, pragmatic approaches are needed to assess which assessments should be prioritized for updates. We consider this focus to fall within the “pre-analysis” domain of an SEM because although the process of identifying studies with study design features most suitable for chronic point of departure (POD) identification is critical, it is still several steps removed from actually identifying a POD and deriving a reference health value. In fact, these later steps are often the more contentious phases of conducting a chemical assessment as they require judgments to be made on the potential adversity of the findings, applicability to humans when animal studies are used as the basis for a POD, dose–response analysis, consideration of mechanistic information to inform judgements on biological plausibility and shape of the dose–response when extrapolating to dose levels below the observed range, and application of uncertainty factors for outcomes assumed to operate via a threshold mechanism. In addition to identifying studies that may be plausibly considered to derive a chronic reference value for acrolein, we also provide a broad characterization of the new evidence to identify health outcomes that may be important to consider for future research or for other decision-context, i.e., acute exposure.

Acrolein, an intermediate in the synthesis of acrylic acid and other organic chemicals, is used as a biocide, and may be released to the environment through the breakdown of certain pollutants in outdoor air. Fig. 1 provides a summary of acrolein emissions from the 2014 National Emissions Inventory (U.S. EPA, 2018b). Humans are exposed to acrolein primarily through structural and forest fires, gasoline and diesel exhaust, tobacco smoke, and partially combusted animal fats and vegetable oils (ToxNet Hazardous Substances Data Bank, 2017; U.S. EPA, 2012; IARC, 1995; Beauchamp et al., 1985). Seaman et al. (Seaman et al., 2007) reported that human inhalation exposure to acrolein is dominated by indoor air (3–40 times higher than concentrations measured in outdoor air). Occupational exposure to acrolein may occur through inhalation and dermal contact at workplaces where it is produced or used.
Additional information on chemical properties, environmental fate and transformation, and toxicokinetics are summarized elsewhere (U.S. EPA, 2019; TCEQ, 2016; OEHHA, 2008; Agency for Toxic Substances and Disease Registry (ATSDR), 2007; U.S. EPA, 2003).

The most recent chronic inhalation reference values based on lesions in respiratory epithelium in rats were developed in 2008 by the California Environmental Protection Agency’s (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA, 2008) and in 2016 by the Texas Commission on Environmental Quality (TCEQ, 2016) (Fig. 2) based on mild hyperplasia and lack of recovery of the respiratory epithelium. Both OEHHA and TCEQ relied on the same subchronic rodent study (Dorman et al., 2008) which identified a NOAEL of 0.2 ppm for nasal respiratory epithelium (RE) lesions and a NOAEL of 0.6 ppm for olfactory epithelium (OE) lesions. Assessments by the US Environmental Protection Agency’s (EPA) Integrated Risk Information System (IRIS) resulting in a chronic inhalation reference concentration (RfC) based on histological changes in nasal epithelium in rats from a subchronic rat inhalation study (Feron et al., 1978) and the Agency for Toxic Substances and Disease Registry (ATSDR) which established inhalation minimal risk levels (MRLs) for both acute (1–14 days) based on decrease in respiratory rate, nose and throat irritation and intermediate (15–365 days) durations based on nasal epithelial metaplasia in rats were last released in 2003 and 2007, respectively, prior to the publication of Dorman et al. (2008). Additional information on the available reference values on acrolein can be found in Appendix A.

The primary goal of this evidence mapping was to review the human health pertinent literature published since 2008, the year OEHHA published a chronic Reference Exposure Level (REL). Although TCEQ updated its chronic Reference Value (ReV) more recently in 2016, the Cal EPA OEHHA assessment was selected for use as the base assessment because it is specifically referenced for use in the most recent U.S. EPA National Air Toxics Assessment (NATA) (U.S. EPA, 2018a) and underwent both public comment and external peer-review (whereas the TCEQ document did not undergo external peer-review). Although TCEQ also used the same study (Dorman et al., 2008) to derive their value, from a systematic review perspective, the public comment and external peer-review features are important in cases where a prior assessment that did not use systematic review methods is used as the anchoring point for a literature update because these steps reduce the chance that key studies would have been missed in the prior assessment. Our targeted goal in assessing the published literature since the OEHHA assessment was to identify studies potentially suitable for identifying a chronic exposure point of departure (POD) and calculation of a chronic reference value. At the same time, we wanted the evidence map to identify studies that may be important to consider for acute or short-term exposure scenarios, hazard identification, or to highlight as areas for future research. Finally, we were interested in assessing how quickly the evidence map could be conducted by use of specialized systematic review software, use of a relatively large assessment team to develop the SEM, and identifying opportunities for pragmatic implementation of systematic review methods that do not undermine credibility of the analysis, as suggested by the National Academy of Sciences (NASEM, 2017).
2. Methods

The systematic review methods used for screening, study evaluation, and data extraction used to conduct the evidence map are outlined in the Office of Research and Development (ORD) Staff Standard Operating Procedures for Developing IRIS Assessments. These methods have been reviewed previously by the National Academy of Sciences (NASEM, 2018) and used in other peer-reviewed systematic reviews (Yost et al., 2019; Radke et al., 2018). For this case example, a multi-disciplinary team of six experienced health scientists (CK, KAT, AG, NK, SV, ASP) conducted the majority of the systematic evidence map analysis. In addition, another four co-authors contributed significantly to defining the scope and design of the systematic evidence map and helped evaluate the results to inform follow-up dose–response analyses (AD, JS, JG, and GW). These 10 authors are full-time employees with EPA, each having a Master’s or PhD degree and each having at least 10 years of experience conducting chemical assessments on a broad range of chemicals. In addition, a number of graduate students or postbaccalaureate EPA student interns were recruited to help compile background information (BS for Fig. 2), screen studies (AGr, KC) and help with data extraction (AGr, SV, KC, KG).

2.1. Populations, exposures, comparators, and outcomes (PECO) criteria

A PECO is used to focus the research question(s), search terms, and inclusion/exclusion criteria in an evidence map or systematic review. The PECO for acrolein is presented in Table 1. In addition to the studies meeting the PECO criteria, studies containing potentially relevant supplemental material were tracked and categorized during the literature screening process as outlined in Table 2.

2.2. Literature search and screening strategies

2.2.1. Database searches—The literature search focused on studies published since Dorman et al. (2008), the subchronic rat study used in the 2008 Cal EPA OEHHA assessment to derive an acrolein chronic REL (OEHHA, 2008), under the assumption that any critical studies published before 2008 would have been considered in the public comment and external peer-review processes. The 2003 IRIS (U.S. EPA, 2003) and 2007 ATSDR (Agency for Toxic Substances and Disease Registry (ATSDR), 2007) assessments were not considered as useful reference materials because they were conducted before the publication of Dorman et al. (2008). The literature search focused only on the chemical name with no limitations on language, type of evidence (i.e., human, animal, in vitro, or in silico) or health outcomes. PubMed, Web of Science (WOS) and Toxline data-bases were searched for the date range of January 1, 2006 through December 2019 by an EPA information specialists (AC, CG, RJ) and stored in the Health and Environmental Research Online (HERO) database (https://hero.epa.gov/hero/). A search start year of 2006 was implemented to ensure identification of any studies published during the preparation, review and finalization of the OEHHA assessment. Full details of the search strategy for each database are presented in Supplemental Materials (Appendix B).

Because the number of studies retrieved during the initial literature search was large even after duplicate removal (> 15,000), the studies were imported into SWIFT Review software
(https://www.sciome.com/swift-review/; see also (Howard et al., 2016) to identify those that are most likely to be applicable to human health risk assessment. In brief, SWIFT Review has pre-set literature search filters developed by information specialists that can be applied to separate studies that are more likely to be useful for identifying human health content from those that likely do not (e.g., analytical methods). The filters function like a typical search strategy where studies are tagged as belonging to a certain filter if the terms in the filter literature search strategy appear in title, abstract, keyword or medical subject headings (MeSH) fields content. The SWIFT Review filters that were applied focused on lines of evidence (human, animal, in vitro) The details of the search strategies that underlie the filters are available at https://hawcprd.epa.gov/media/attachment/SWIFT-Review_Search_Strategies.pdf. Studies not retrieved using these filters were not considered further in screening. Studies that included one or more of the search terms in the title, abstract, keyword, or MeSH fields were exported as a RIS file for screening in another software application, SWIFT Active, as described below. Application of these filters reduced the number of studies for further consideration from 15,788 to 10,336.

2.2.2. Other resources consulted—In addition to the database searches, the following approaches were used to identify any studies that may have been missed based on the database search:

- Review the reference list from the Texas Commission on Environmental Quality (TCEQ) Development Support Document for Acrolein (last revised February 4, 2016).
- Search of the EPA ChemView database (https://chemview.epa.gov/chemview) to identify unpublished studies, information submitted to EPA under Toxic Substances Control Act (TSCA) Section 4 (chemical testing results), Section 8(d) (health and safety studies), Section 8(e) (substantial risk of injury to health or the environment notices), and FYI (voluntary documents).
- Search of the National Toxicology Program (NTP) database of study results and research projects (https://ntp.niehs.nih.gov/results/index.html).
- Search titles in the reference list of studies meeting the PECO criteria after full-text review for potentially relevant studies.
- A WoS citation mapping for studies that cited the references screened as meeting the PECO criteria after full-text review.

2.2.3. Screening process—For efficient screening, studies identified as potentially relevant in SWIFT Review were exported into SWIFT Active (https://www.sciome.com/swift-activescreener/), a specialized systematic review software application that uses “active” learning by which real time screening decisions are used to prioritize unscreened studies for relevance. Two screeners per study (ASP, CK, NK, KAT, AG, SVV, plus ICF screeners listed in the acknowledgements) conducted a title and abstract (TIAB) screen of the search results in SWIFT Active to identify (i.e., screen as “include”) study records that met the PECO eligibility criteria Table 1), were unclear on whether they met PECO eligibility criteria, or appeared to be supplemental material content (Table 2). The PECO
eligibility criteria in Table 1 and supplemental material categories in Table 2 were made available to screeners via the screening form in the software to facilitate screening. Chemical name, CAS number, and synonyms were keyword highlighted in the form. A pilot screening phase was conducted where screeners were asked to review 50 studies and the team met to discuss screening conflicts and clarify screening instructions. Subsequent screening conflicts are tracked within SWIFT Active and were resolved by discussion as needed. Studies that do not meet all elements in the PECO or did not appear to be supplemental material were excluded during SWIFT Active TIAB screening. Screeners were not asked to sub-tag on specific type of supplemental content within SWIFT Active, instead this was done as described below during subsequent TIAB screening to distinguish PECO relevant versus supplemental studies. Screening continued until SWIFT Active indicated that it was likely that at least 95% of the relevant studies were identified, a percent identification often used to evaluate the performance of machine learning applications and considered comparable to human error rates (Bannach-Brown, 2018; Howard et al., 2016; Cohen et al., 2006). Concern that a key PECO relevant study might have been missed were mitigated by manually consulting other resources as described above, which included conducting a forward search for studies that cited any of the studies considered relevant after full-text review. As another check for potential missed studies we consulted the TCEQ acrolein assessment released in 2016 (TCEQ, 2016), which although was not conducted using systematic review methods, quite likely did not miss any critical study pertinent to POD derivation given the high visibility of the TCEQ assessments.

The studies screened as “include” in SWIFT Active were then imported into DistillerSR software (https://www.evidencepartners.com/products/distillersr-systematic-review-software) for an additional round of title and abstract tagging and full-text review, also conducted by two independent reviewers per study (ASP, CK, NK, KAT, AG, SVV, plus ICF screeners listed in the acknowledgements). The DistillerSR screening form used a structured check all that apply format for categorizing supplemental material according to the content type presented in Table 2. The initial screening level distinctions between a study meeting the PECO criteria and a supplemental study are often made for practical reasons and the tagging structure in Table 2 is designed to ensure the supplemental studies are categorized for easy retrieval if needed while conducting an assessment, which we envision to occur subsequent to preparation of a systematic evidence map. In our case example, studies that meet the PECO criteria are those that have information on the chemical exposure of interest and a health outcome in experimental animal models or humans. As described below, the study design and results of these studies are summarized and undergo additional data extraction and individual level study evaluation if they have study design features that support consideration for POD purposes. In contrast, the impact of studies tagged as supporting material is often difficult to assess during the initial phase of the assessment. These studies may emerge as being critically important and need to be evaluated and summarized at the individual study level (e.g., cancer mode of action mechanistic or ADME studies), or be helpful to provide context (e.g., summarize current levels of exposure, provide hazard evidence from routes or durations of exposure not pertinent to the PECO), or not be cited at all in the assessment (e.g., individual studies that contribute to a well-established scientific conclusion). For the purpose of our systematic evidence map case
example, a detailed analysis of supplemental material content is considered to fall outside the primary goal of identifying studies published since Dorman et al. (2008) that could be used for POD purposes. However, further analysis of supplemental material content would be expected if the systematic evidence map was used as part of subsequent assessment that included hazard characterization and dose–response analysis.

Studies may be tagged as supplemental material during either TIAB or full-text screening. Because a detailed analysis of supplemental material content falls outside the scope of this systematic evidence map, full-text retrieval was not conducted for studies tagged as supplemental at the TIAB level because of the cost and time involved in obtaining full-text. Typically, the number of studies tagged as supplemental exceeds the number considered PECO-relevant. As indicated above, full-text retrieval of supplemental content would be expected to occur if the systematic evidence is used as the basis for a subsequent assessment and it becomes clearer how the supplemental material would be used and therefore for which studies full-text retrieval is warranted.

This additional round of TIAB tagging was conducted in DistillerSR to separate PECO relevant studies from supplemental material studies, including tagging for the specific type of supplemental material content presented in Table 2. The reason for this secondary TIAB tagging in DistillerSR is that supplemental material studies, by definition, do not meet PECO criteria, thus unless they are marked as “included” in SWIFT Active the machine-learning algorithms will not prioritize them for consideration. This would increase the chances that important supplemental material content would be missed and decreases the utility of using a machine-learning software for chemical assessments. Although tagging the type of supplemental content can be done in either SWIFT Active and DistillerSR, we typically do this tagging in DistillerSR. This allows for the resolution of conflicts between reviewers at the level of how each supplemental material study was tagged, and the software workflow for this level of granularity in conflict resolution is easier in DistillerSR. For similar reasons, we conduct full-text screening in DistillerSR. In our experience, use of machine-learning applications such as SWIFT Active still reduces the overall screening level of effort even with an additional round of TIAB tagging to separate PECO relevant from supplemental studies.

Any studies not retrieved from the database search but identified from other resources (e.g., from the reference list of studies screened as PECO-relevant after full-text review) were imported into DistillerSR and screened for PECO eligibility beginning at the TIAB level.

2.3. Study design considerations for evaluating whether new studies may plausibly be useful for POD identification and calculation of a reference value

Studies considered PECO-relevant after full-text review were briefly summarized in a DistillerSR data extraction form. This is referred to as a literature inventory and is intended to provide a brief summary of study methods and results to help identify studies with the study design features that make them most suitable for identifying a chronic POD. The literature inventory as performed by one member of the evaluation team was checked by at least one other member (CK, KAT, AG, NK, SVV, SV, KC, ASP). For animal studies, the following information was captured: study type (e.g., acute, subchronic, developmental),
duration of exposure, route, species, strain, sex, concentration levels tested and units, health system and specific endpoints assessed, and a summary of findings at the health system level [null or no-observed-adverse-effect level/lowest-observed-adverse-effect level (NOAEL/LOAEL) with an indication of which specific endpoints were affected]. For epidemiologic studies, the following information was summarized: population type (e.g., general population-adult, occupational, pregnant women, infants and children, etc.), study type (e.g., cross-sectional, cohort, case-control), short free-text description of study population, sex, description of how exposure was assessed, health system and specific outcome assessed, and a summary of findings at the health system level (null or an indication of any associations found and a description of how the exposure was quantified in the analysis). These inputs were analyzed with respect to the considerations below to identify new studies that could plausibly be considered as suitable or more suitable than Dorman et al. (2008) for potential toxicity reference value derivation purposes. Studies prioritized from this process underwent study evaluation and full data extraction. Studies not prioritized from this process were summarized at the literature inventory level only and primarily used to identify areas of emerging concern.

• Animal studies with multiple dose groups that tested dose levels similar to or lower than previously studied. Studies that administered only a single dose level were considered less suitable, as Dorman et al. (2008) tested 5 dose levels.

• For human studies, studies that expressed exposure levels quantitatively were prioritized.

• For human studies, studies that used biomarker measurements in tissues or bodily fluids as the metric for exposure were only considered suitable for dose–response analysis if data or PBPK models are available to extrapolate between the reported biomarker measurements and the level of inhalation exposure.

• For either experimental animal or human studies, non-developmental studies with chronic or subchronic exposure durations were prioritized over non-developmental studies with short term (< 4 weeks) or acute exposure durations.

• For either experimental animal or human studies, those with overall study evaluation ratings of “high” or “medium” confidence are considered (see “Study Evaluation” below).

• For either experimental animal or human studies, the nature of the outcomes/ endpoints assessed as being interpretable with respect to potential adversity. Typically, apical or clinical measures, when available, are preferred over biochemical or mechanistic endpoints for dose–response analysis.

2.4. Study evaluation

Study evaluation was conducted for studies identified as potentially suitable for toxicity value derivation by two independent reviewers (AG, ASP, KAT, BS, CK, NK) using EPA’s version of Health Assessment Workspace Collaborative (HAWC, https://hawcrd.epa.gov/portal/), a free and open source web–based software application designed to manage and facilitate the process of conducting literature assessments. Given the known availability of
subchronic studies (Dorman et al., 2008; Feron et al., 1978) acute and short-term duration non-developmental studies (≤ 4 weeks) did not undergo study evaluation.

The general approach for evaluating epidemiology and animal toxicology is the same (Fig. 3) but the specifics of applying the approach differ. Key concerns were potential sources of bias (factors that could systematically affect the magnitude or direction of an effect in either direction) and insensitivity (factors that limit the ability of a study to detect a true effect). For epidemiology studies, the exposure measurement domain was evaluated first and potential use of the study for dose–response assessment was assessed before analyzing other domains. If the exposure domain raised issues regarding suitability of the study for dose–response, then the study was not further evaluated. This approach was taken because epidemiology studies considered to be ‘deficient’ or ‘critically deficient’ in exposure measurement would typically not be suitable for use in dose–response regardless of the ratings in other domain metrics. Moreover, it is possible that exposure characterization can be well-conducted with respect to internal validity but not easily interpreted in the context of quantifying inhalation exposure to acrolein for dose–response purposes, e.g., blood level or urinary metabolite levels. This can be considered a form of directness or applicability of the study to address the assessment topic rather than an issue of internal validity. Given that the focus of this review is to identify studies that could plausibly be used for dose–response, epidemiology studies were also considered in the context of applicability for dose–response before evaluation of other study domains is conducted. Consideration of potential applicability was informed by the availability of other data, either empirical or pharmacokinetic models. However, it should be emphasized that a study excluded from further consideration, based on the scenarios above, may be useful to address other types of assessment questions, such as those that focus on hazard characterization or to highlight emerging issues and key data gaps. For this reason, the findings of studies not considered suitable for POD derivation were briefly summarized in the literature inventory process described above.

During study evaluation, in each evaluation domain, at least two reviewers reached a consensus rating regarding the utility of the study for hazard identification, with categories of Good, Adequate, Deficient, Not Reported or Critically Deficient. No attempts were made to contact authors for information that was not reported in a study. Once the evaluation domains have been rated, the identified strengths and limitations will be considered to reach a study confidence rating of High, Medium, Low, or Uninformative for a specific health outcome. This will be based on the reviewer judgments across the evaluation domains and will include consideration of the likely impact the noted deficiencies in bias and sensitivity, or inadequate reporting, have on the results. The ratings, which reflect a consensus judgment between reviewers, are defined as follows:

- **High**: A well-conducted study with no notable deficiencies or concerns were identified; the potential for bias is unlikely or minimal, and the study used sensitive methodology. “High” confidence studies generally reflect judgments of good across all or most evaluation domains.

- **Medium**: A satisfactory (acceptable) study where deficiencies or concerns were noted, but the limitations are unlikely to be of a notable degree. Generally,
“medium” confidence studies will include adequate or good judgments across most domains, with the impact of any identified limitation not being judged as severe.

- **Low**: A substandard study where deficiencies or concerns were noted, and the potential for bias or inadequate sensitivity could have a significant impact on the study results or their interpretation. Typically, “low” confidence studies would have a deficient evaluation for one or more domains, although some “medium” confidence studies may have a deficient rating in domain(s) considered to have less influence on the magnitude or direction of effect estimates. Generally, low confidence results are given less weight compared to high or medium confidence results during evidence synthesis and integration and are generally not used as the primary sources of information for hazard identification or derivation of toxicity values unless they are the only studies available. Studies rated as “low” confidence only because of sensitivity concerns about bias towards the null require additional consideration during evidence synthesis.

- **Uninformative**: An unacceptable study where serious flaw(s) make the study results unusable for informing hazard identification. Studies with critically deficient judgments in any evaluation domain will almost always be classified as “uninformative” (see explanation above). Studies with multiple deficient judgments across domains may also be considered “uninformative.” Uninformative studies will not be considered further in the synthesis and integration of evidence for hazard identification or dose response but may be used to highlight possible research gaps.

The rationale for the classification, including a brief description of any identified strengths and/or limitations from the domains and their potential impact on the overall confidence determination is documented and retrievable in HAWC (https://hawcprd.epa.gov/assessment/100000047/).

### 2.5. Full data extraction of study methods and results

The literature inventory described above was conducted in DistillerSR and visualized in Tableau software (www.tableau.com). Studies considered more suitable for dose–response analysis underwent a more detailed summarization of study methods and findings using HAWC. Data extraction in HAWC was performed by one member of the evaluation team and checked by at least one other member (KAT, BS, NK, AG, CK). No attempts were made to contact authors for information that was not reported in a study. Study evaluation and detailed summarization of study design and methods conducted in HAWC are available for download from EPA HAWC in Excel format at https://hawcprd.epa.gov/assessment/100000047/downloads/.

### 3. Results

#### 3.1. Literature screening results

The flow of studies is summarized in Fig. 4 and the interactive HAWC literature tree (Fig. 5). The database searches, covered January 1, 2006 through December 2019, yielded 15,788
unique records. As described earlier, these studies were imported into SWIFT Review and literature search filters for lines of evidence (human, animal, in vitro) was applied. Application of these filters reduced the number of studies for consideration to 10,336 which was then reduced to 9,656 after another round of duplicate removal within the SWIFT software. The studies were screened at the title and abstract level in Swift Active using predictive relevance. The use of predictive relevance resulted in manually screening a third of the references (3,073 studies) with 648 studies identified as potentially relevant (“included”) and 2,425 manually tagged as excluded. The Web of Science citation mapping search found 750 unique references. These references, along with 648 previously identified in SWIFT Active, underwent an additional round of title and abstract screening in DistillerSR, yielding 104 references being screened at the full text level in DistillerSR.

During full-text review, 60 studies were identified further consideration. At this point, there were a total of 446 studies tagged as supplemental, 412 from title and abstract screening and 34 from full text screening. A majority of these studies were mechanistic (289), with few falling into other categories. As the goal was to identify chronic/subchronic exposure studies from inhalation routes of exposure to acrolein, the supplemental references include non-inhalation/non oral route studies (64), mixture studies (4), and studies with no original data (28). It is noteworthy to mention that 10 non-English studies were tagged as supplemental because no translation services were sought, as these studies appeared to be of < 4 weeks in duration or focused on mechanistic endpoints that would have been more difficult to interpret for adversity (Appendix C) compared to the findings in Dorman et al. (2008). Of the 60 studies that were further considered, 55 studies, including 44 animal studies and 13 human studies, were summarized in the literature inventory based on full-text content (see interactive Tableau visual, Fig. 6). The studies on PBPK (3) or mechanistic outcomes (3) were not included in the literature inventory. After considering the duration of exposure, only 7 of the 37 studies (5 human studies and 2 animal toxicity studies, including Dorman et al. (2008) were considered for study evaluation and potential POD suitability. No additional PECO-relevant studies were reported in the EPA ChemView, the NTP database of study results and research projects, the 2016 TCEQ assessment reference list (the most recent scientific study cited was from 2009), or references cited in the 7 studies identified for further review. In addition, a targeted search using the “PBPK” filter in SWIFT Review identified two studies addressing dosimetric modeling of acrolein in the respiratory tract (Corley et al., 2012; Schroeter et al., 2008).

3.2. Identification of new studies suitable for POD derivation and calculation of a reference value

The studies meeting the PECO criteria after full-text review were briefly summarized in DistillerSR for study type, exposure duration, exposure concentrations tested, NOAEL/LOAEL, health outcome, and compared to Dorman et al. (2008) for suitability in deriving a chronic inhalation reference value. As a point of reference, Dorman et al. (2008) exposed male F344 rats to five exposure concentrations of acrolein, ranging from 0.02 to 1.8 ppm, for 13 weeks. Based on the results of this study, a NOAEL of 0.2 ppm for nasal respiratory epithelium (RE) lesions and a NOAEL of 0.6 ppm for olfactory epithelium (OE) lesions was established.
3.2.1. Animal studies—Most of the animal studies were of short-term duration (< 4 weeks) and many only tested a single concentration level, which was often a higher concentration compared to those tested in Dorman et al. (2008). These studies were not considered further for study evaluation, although the DistillerSR literature inventory data extraction is available in Supplemental Materials (see interactive Tableau visual). Ten non-English studies with an English abstract were identified, but they often did not report the exposure concentration level tested or details on the duration of treatment, although most appeared to be short-term (< 4 weeks) (Wei et al., 2015; Jia et al., 2013; Wei et al., 2012; Dan et al., 2010; Wang et al., 2010; Yuan et al., 2010; Liu et al., 2009; Guan et al., 2008; Wang et al., 2008; Xu et al., 2008) (Appendix C). Translation was not pursued for these studies because the abstracts indicated a primary focus on mechanistic aspects of acrolein-induced effects on respiratory, immune, and inflammation outcomes (see Appendix C for study summaries).

Only one subchronic study (30–90 days of exposure) was considered a plausible candidate (Conklin et al., 2017) for toxicity value derivation when compared to the Dorman et al. (2008) study and was advanced for study evaluation. In the study by Conklin et al. (Conklin et al., 2017) male C57BL/6J mice were treated with 0, 0.5, or 1 ppm acrolein for 12 weeks (6 hrs per day, 5 days per week). At 1 ppm, increased levels of urinary metabolite 3-hydroxypropyl mercapturic acid (3-HPMA), pulmonary acrolein-metabolizing enzymes, and Nrf2-regulated antioxidant proteins were observed. There was also suppression in circulating levels of endothelial progenitor cells (EPCs) and certain types of leukocytes.

Based on study evaluation, Dorman et al. (2008) was considered a high confidence study and Conklin et al. (Conklin et al., 2017) was considered to be low confidence due to issues of reporting quality and exposure characterization, as there was no mention of the source of acrolein, purity or other details of administration (Fig. 7). As an additional point of reference, the Feron et al. (Feron et al., 1978) study used to derive an RfC in the 2003 IRIS assessment was also evaluated and considered medium confidence. Thus, Conklin et al. (Conklin et al., 2017) was not considered to be more appropriate for dose–response analysis given its significant methodological limitations and the availability of higher confidence studies.

3.2.2. Human studies—A literature inventory summary of the thirteen human studies identified are provided in the Supplemental Materials (see interactive Tableau visual). Four of the human studies assessed acute exposure to acrolein through controlled exposure studies (Claeson et al., 2017; Claeson and Andersson, 2017; Claeson and Lind, 2016; Dwivedi et al., 2015). The remaining studies either assessed the levels of the metabolite 3-HPMA as a biomarker for the presence of acrolein (Dejarnett et al., 2014; Yoshida et al., 2012; Yuan et al., 2012), or had a study design with limited quantitative exposure assessment (Yasuo et al., 2019; Neghab et al., 2017; Feroe et al., 2016; Decastro, 2014; Dejarnett et al., 2014; Annesi-Maesano et al., 2012; Yoshida et al., 2012; Slaughter et al., 2004).

An initial study quality evaluation was performed on the five nonacute studies with specific emphasis on the exposure domain, in particular, the exposure measurement, timing/temporality, sensitivity, and applicability of the exposure measure for toxicity value.
Two of these studies (Decastro, 2014; Annesi-Maesano et al., 2012) measured air levels of acrolein and were considered most applicable for use in deriving an inhalation reference value. The other studies that relied on a urinary biomarker of exposure were considered less applicable since it is not clear how the metabolite level could be used to establish an inhalation value. All studies were considered deficient with respect to ability to assess temporality of exposure and the health outcome. Temporality is a scenario where the exposure is known (or can be inferred) to have preceded the health outcome or more specifically, must occur at a time period relevant to the etiology of the disease. Overall, the epidemiology studies were considered less suitable than Dorman et al. (2008) for dose response derivation given concerns regarding temporality and exposure misclassification in Decastro, 2014 and Annesi-Maesano et al., 2012. For this reason, full study evaluation was not conducted for the epidemiological studies.

Although not probed further in terms of utility for dose response for a chronic reference value derivation, the four acute studies do provide information that may be useful for the identification or confirmation of previously suggested LOAELs and NOAELs for acute effects. The acute effects observed were primarily sensory irritation of the eyes, nose and throat, with NOAELs ranging from 0.05 to 0.06 ppm and LOAELs from 0.01 to 0.15 ppm (Claeson and Lind, 2016; Dwivedi et al., 2015; (Claeson et al., 2017; Claeson and Andersson, 2017).

3.2.3. Conclusions on impact of new evidence for chronic POD identification

Based on the analysis above, it is concluded that Dorman et al. (2008) is likely to continue to represent the most suitable study for derivation of a chronic inhalation POD. In this study, male F344 rats (12 rats per group) were exposed whole body to air concentrations of 0, 0.02, 0.06, 0.2, 0.6, or 1.8 ppm acrolein for 6 h/day, 5 days/week, for 13 weeks. Respiratory tract histopathology was evaluated after 4, 14, 30 and 65 days of exposure, and at 60 days following the end of the 13-week exposure period. Histopathological evaluations were performed on several sections of the nasal respiratory and olfactory epithelium, and on the larynx, trachea, and lungs. A NOAEL of 0.2 ppm (0.46 mg/m$^3$) for nasal respiratory epithelium (RE) lesions and a NOAEL of 0.6 ppm (1.38 mg/m$^3$ for olfactory epithelium (OE) lesions was established. The study was considered overall high confidence (Fig. 7) and the findings consistent with portal-of-entry effects observed in studies of related reactive aldehydes and other studies in the acrolein database identifying nasal tissues as a target site. This includes the Feron et al. (1978) study used to derive an RfC in the 2003 IRIS assessment (U.S. EPA, 2003) (Fig. 9).

The Dorman et al. (2008) has several qualities that make it more suitable for chronic inhalation POD derivation compared to the Feron et al. (1978) study used in the 2003 IRIS assessment. Feron et al. (1978) supported the identification of NOAELs, LOAELs, and frank effect levels (FELs) for the three species tested, determination of the critical target site, and a comparison of sensitivity among the three species. However, several limitations in this study were noted, including 1) the incidence of nasal lesions for treated groups was not reported; 2) an exposure duration of 3 months was used, rather than lifetime; 3) histopathological examination was only done on only three sections of the nasal cavity while other studies had 6 sections; 4) there was a lack of characterization of the type of nasal lesions by sex; and 5)
only 6 rats/sex were exposed. Given these limitations, the study was considered medium confidence overall in terms of study quality (Fig. 7). In contrast to Feron et al. (1978) and other studies in the database, Dorman et al. (2008) included several study design attributes that better characterize the respiratory tract toxicity of acrolein including: increased number of exposure groups; increased number of lower exposure concentrations; interim histopathological evaluations; and increased tissue sectioning for histopathological evaluations (including 6 nasal sections, as well as sections from the larynx, trachea and lungs). This, along with a detailed consideration of the critical effect from Dorman et al. (2008) is provided in the Supplemental Materials (see Appendix D).

In terms of quantitative cancer value derivation, no new animal bioassays were identified, and thus no further assessment on carcinogenicity was performed. Thus, the OEHHA conclusion of inadequate direct evidence for carcinogenicity of acrolein in humans or experimental animals appears current. The same holds true for carcinogenicity conclusions reached by the National Toxicology Program (NTP) and the International Agency for Research on Cancer (IARC) (NTP, 2005; IARC, 2004), which were similar to that reached by OEHHA. With respect to other human health considerations, new studies suggest that more research on the long-term cardiovascular and metabolic effects of acrolein are warranted and additional data or models are needed in order to more fully utilize the existing epidemiological studies that focus on urinary biomonitoring for POD purposes. No reproduction of developmental toxicity studies was identified, so this remains a data gap.

Typically, for the derivation of health reference values based upon animal data, the calculated POD values are converted to human equivalent concentrations (HECs) using the appropriate default dosimetric adjustment factor (DAF). DAFs are ratios of animal and human physiologic parameters and are dependent on the nature of the contaminant (i.e., particle or gas) and the target site (i.e., respiratory tract or remote to the portal-of-entry [i.e., systemic]) (U.S. EPA, 2012, 2009, 1994). However, as outlined in (U.S. EPA, 2009, 2002, 1994), dosimetry models and chemical- and species-specific parameters represent more optimal approaches for dosimetry and interspecies extrapolation and, as such, should be considered as appropriate. For acrolein, advances in computational fluid dynamic (CFD) dosimetric modeling results (Corley et al., 2012; Schroeter et al., 2008) could be considered for interspecies extrapolation and calculation of the HEC. In brief, Corley et al. (2012) estimated a maximal flux (pg/cm\(^2\)-s) differential between rats and humans of 2.1 (1400 vs 660) in the anterior region of the nose at 0.6 ppm acrolein. Similarly, Schroeter et al. (2008) reported flux rates of 691 to 1045 (pg/cm\(^2\)-s) in the rat at the LOAEL of 0.6 ppm acrolein for respiratory epithelial effects. Schroeter et al. (2008) also estimated the highest average flux at the NOAEL of 0.6 ppm for olfactory epithelium effects to be 692 (pg/cm\(^2\)-s) in the rat. As a comparison, the 99th percentile/maximum flux was predicted to be approximately 400 to 470 (pg/cm\(^2\)-s) in the human. Comparative flux estimates in rats and humans at the NOAEL of 0.2 ppm for respiratory epithelial effects observed by Dorman et al., (2008) were not reported in either study. In addition, both studies reported similar nasal extraction efficiencies which were greater in the rat compared to the human. These modeling results indicate that where dosimetric comparisons can be made, flux estimates in the nasal regions for a given acrolein exposure concentration are greater in the rat than the human. However, because comparative flux estimates were not provided in rats and humans over a range of

*Environ Int. Author manuscript; available in PMC 2021 March 01.*
exposure concentrations or at the NOAEL of 0.2 ppm acrolein for nasal respiratory epithelium lesions, quantitative application of these results is limited. Therefore, a DAF of 1 for interspecies extrapolation could be considered appropriate. Additional details about the analysis of (Corley et al., 2012; Schroeter et al., 2008) are provided in the Supplemental Materials (see Appendix D).

4. Conclusions

We found systematic evidence mapping to be a useful tool to assess whether new evidence is likely to impact conclusions of an existing human health assessment, to identify emerging areas of human health concerns, and to understand the level of effort and specific staffing expertise that might be required to conduct an updated health assessment. In this case example with acrolein, we concluded that the study used for reference value derivation (Dorman et al. (2008)) in 2008 OEHHA (OEHHA, 2008) and in 2016 by TCEQ (TCEQ, 2016) is still likely the most appropriate study for derivation of a chronic inhalation reference value.

We took a number of pragmatic steps when developing this systematic evidence map. First, the scope of the literature search was designed to build from the 2008 OEHHA (OEHHA, 2008) assessment rather than the entire pool of literature. This approach reduces the level of effort needed to screen all available literature although it does rely on the assumption that “key” studies were not missed in prior assessments that were not conducted using systematic review methods. We believe this approach can be justified in cases where the prior assessments have undergone a rigorous peer-review process including public comment and external peer-review, as is the case for OEHHA and many other health assessments conducted by Federal agencies. For acrolein, we had the additional reassuring observation that the key study identified in the 2008 OEHHA (OEHHA, 2008) assessment was also identified as the key study in the 2016 TCEQ (TCEQ, 2016) assessment.

We found use of machine-learning screening applications greatly expedite the screening process. We used SWIFT Active for this case study, but other machine learning screening applications such as Sysrev (https://sysrev.com/), DistillerAI (https://www.evidencepartners.com/distiller-ai/) and others are rapidly proliferating as their use in systematic review is gaining acceptance. The Systematic Review Toolbox (http://systematicreviewtools.com/) is one resource for tracking available software applications. Overall, with use of a large screening team the process was relatively fast. A screening team of 17 people were able to complete the TIAB level screening in 18 days and full-text review in 5 days (including time for pdf retrieval). Per screener, the time to review studies at TIAB level ranged from 18 – 257 s per reference in SWIFT Active and 30 – 277 s per reference in DistillerSR. Full-text review in DistillerSR took longer, typically several minutes per study. The time to resolve conflicts is not tracked in these software applications but was estimated to take 3–8 min per conflict. Summarizing studies for the literature inventory was more time consuming and took 15 – 30 min per study.

Another pragmatic step we took was to triage study evaluation for epidemiology studies to focus first on the exposure domain and only consider other domains if the exposure domain
was considered ‘adequate’ or ‘good’ and the exposure metric could readily be used to derive an inhalation reference value. This approach greatly reduces the level of effort required by epidemiologists, which is important because analysis of epidemiology studies is one of the most time intensive phases of systematic review and epidemiology staff are often in short supply in many assessment groups. We felt this approach was reasonable for the main targeted goal of our systematic evidence map (to identify studies suitable for POD purposes and reference value derivation). However, it should be emphasized that studies with exposure assessment limitations may be useful to address other types of assessment questions, such as those that focus on hazard characterization or to highlight emerging issues and key data gaps. For this reason, all epidemiological studies were described at the “literature inventory” level in the acrolein systematic evidence map even if they were not considered suitable for identifying a POD.

One pragmatic step that we initially took turned out to be a mistake. In our initial screening, we used a single screener per record for TIAB screening in SWIFT Active in order to quickly process the larger number of references identified from the database search (+15,000). This raised concern during the journal peer-review process. In deciding how to best address this concern we did some initial analysis to determine the degree to which we might have missed studies by comparing the number of studies identified as PECO relevant from our WoS citation mapping search to those identified as included at TIAB level in SWIFT Active. We found out that several studies were missed when we had used a single screener. Thus, we completely re-did the screening when revising the manuscript. Our strong conclusion from this experience is that use of a single screener is not worth the detrimental impact on credibility of the review.

Although the main goal of this systematic evidence map was to identify studies published since the 2008 OEHHA assessment suitable for chronic POD identification, the scope of the screening and presentation of studies was wide enough to capture and catalog information on other areas of interest, including acute exposure health studies, mechanistic information, and studies in non-mammalian model systems. In closing, systematic evidence mapping, especially when conducted using specialized systematic review and visualization software tools, is an efficient and valuable analysis tool to aid problem formulation, scoping, and to determine when an updated assessment may be warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank the following individuals: Annamarie Cory (AC), Carolyn Gatling (CG), and Ryan Jones (RJ) for their assistance in conducting the literature search; Amina Wilkins and Margaret Pratt for DistillerSR support; Vicki Soto and Dahnish Shams for document production support; and Andrew Kraft and Brittany Jacobs for internal review. The authors would also like to thank the following ICF staff for assistance in screening studies and creating Tableau visualizations and HAWC trees: Carlye Austin, Robyn Blain, Katherine Duke, Alexandra Goldstone, Joanna Greig, Sophie Hearn, Audrey Ichida, Ellen Lee, Courtney Lemeris, Camryn Lieb, Jeanne Luh, Alicia Murphy, Johanna Rochester, Amanda Ross, Alessandria Schumacher, Jennifer Seed, Christopher Sibrizzi, Samantha Snow, Marie Socha, Nicole Vetter, Ashley Williams, and River Williams.
References

Agency for Toxic Substances and Disease Registry (ATSDR), 2007. Toxicological profile for acrolein (Update). (CIS/08/00763). Agency for Toxic Substances and Disease Registry (ATSDR).

Annesi-Maesano I, Hulin M, Lavaud F, Raherison C, Kopferschmitt C, de Blay F, Charpin DA, Denis C, 2012. Poor air quality in classrooms related to asthma and rhinitis in primary schoolchildren of the French 6 Cities Study. Thorax 67, 682–688. 10.1136/thoraxjnl-2011-200391. [PubMed: 22436169]

Bannach-Brown A, Przybyla P, Thomas J, Rice A,SC, Ananiadou S, Liao J, Macleod MR, 2018. The use of text-mining and machine learning algorithms in systematic reviews: reducing workload in preclinical biomedical sciences and reducing human screening error. Cold Spring Harbor Protocols 0,1–26. 10.1101/255760.

Beauchamp RO Jr, Andjelkovich DA, Kligerman AD, Morgan KT, Heck H.d'A., Feron VJ, 1985. A critical review of the literature on acrolein toxicity [Review]. Crit. Rev. Toxicol 14, 309–380. 10.3109/10408448509037461. [PubMed: 3902372]

Bragg P, Clavisi O, Turner T, Tavender E, Collie A, Gruen RL, 2011. The Global Evidence Mapping Initiative: scoping research in broad topic areas. BMC Med. Res. Methodol 11, 92. 10.1186/1471-2288-11-92. [PubMed: 21682870]

CEE (Collaboration for Environmental Evidence), 2019. Journal of Environmental Evidence. Available online at https://environmentalevidencejournal.biomedcentral.

Claeson AS, Andersson L, 2017. Symptoms from masked acrolein exposure suggest altered trigeminal reactivity in chemical intolerance. Neurotoxicology 60, 92–98.10.1016/j.neuro.2017.03.007. [PubMed: 28359837]

Claeson AS, Gouveia-Figueira S, Häggström J, Fowler CJ, Nording ML, 2017. Levels of oxylipins, endocannabinoids and related lipids in plasma before and after low-level exposure to acrolein in healthy individuals and individuals with chemical intolerance [Supplementary material]. Prostaglandins Leukot. Essent. Fatty Acids 121.

Claeson AS, Lind N, 2016. Human exposure to acrolein: Time-dependence and individual variation in eye irritation. Environ. Toxicol. Pharmacol 45, 20–27. 10.1016/j.etap.2016.05.011. [PubMed: 27235799]

Cohen AM, Hersh WR, Peterson K, Yen PY, 2006. Reducing workload in systematic review preparation using automated citation classification. J. Am. Med. Inform. Assoc 13, 206–219. 10.1197/jamia.M1929. [PubMed: 16357352]

Conklin DJ, Malovichko MV, Zeller I, Das TP, Krivokhizhina TV, Lynch BH, Lorkiewicz P, Agarwal A, Wickramasinghe N, Haberzetl P, Sithu SD, Shah J, O’Toole TE, Rai SN, Bhatnagar A, Srivastava S, 2017. Biomarkers of chronic acrolein inhalation exposure in mice: implications for tobacco product-induced toxicity. Toxicol. Sci 158, 263–274. 10.1093/toxsci/kfx095. [PubMed: 28482051]

Corley RA, Kabilan S, Kuprat AP, Carson JP, Minard KR, Jacob RE, Timchalk C, Glenny R, Pipavath S, Cox T, Wallis C, Larson RF, Fanucci MV, Postlethwait E, Einstein DR, 2012. Comparative computational modeling of air-flows and vapor dosimetry in the respiratory tracts of a rat, monkey, and human. Toxicol. Sci 128, 500–516. 10.1093/toxsci/kfs168. [PubMed: 22584687]

Dan QQ, Li Y, Zhang L, Zhao S, Wang SL, Yuan B, Zhang YH, 2010. Changes of proteomics in the injured lung of adult rats subjected to acrolein inhalation. Sichuan Daxue Xuebao (Yixue Ban) 41, 269–272. [PubMed: 20506650]

Decastro B, 2014. Acrolein and Asthma Attack Prevalence in a Representative Sample of the United States Adult Population 2000–2009. PLoS One 9, e96926. 10.1371/journal.pone.0096926. [PubMed: 24816802]

Dejarnett N, Conklin DJ, Riggs DW, Myers JA, O’Toole TE, Hamzeh I, Wagner S, Chugh A, Ramos KS, Srivastava S, Higdon D, Tollerud DJ, Defilippis A, Becher C, Wyatt B, Mccracken J, Abplanalp W, Rai SN, Ciszewski T, Xie Z, Yeager R, Prabhu SD, Bhatnagar A, 2014. Acrolein exposure is associated with increased cardiovascular disease risk. J. Am. Heart Assoc 3. 10.1161/JAHA.114.000934.
Dorman DC, Struve MF, Wong BA, Marshall MW, Gross EA, Willson GA, 2008. Respiratory tract responses in male rats following subchronic acrolein inhalation. Inhal. Toxicol 20, 205–216. 10.1080/08958370701864151. [PubMed: 18300043]

Dwivedi AM, Johanson G, Lorentzen JC, Palmberg L, Sjogren B, Ernstgard L, 2015. Acute effects of acrolein in human volunteers during controlled exposure. Inhal. Toxicol 27, 810–821. 10.3109/08958378.2015.1115567. [PubMed: 26635308]

EHP (Environmental Health Perspectives), 2019. Environmental Health Perspectives. Available online at https://ehp.niehs.nih.gov/ (accessed April 18, 2019).

EI (Environment International), 2019. Environment International. Available online at https://www.sciencedirect.com/journal/environment-international (accessed April 18, 2019).

Elsevier, 2017. Guidance notes for authors of systematic reviews, systematic maps and other related manuscripts. Available online at https://www.elsevier.com/journals/environment-international/0160-4120/guidance-notes.

Feroe AG, Attanasio R, Scinicariello F, 2016. Acrolein metabolites, diabetes and insulin resistance. Environ. Res 148, 1–6. 10.1016/j.envres.2016.03.015. [PubMed: 26991531]

Feron VJ, Kruysse A, Til HP, Immel HR, 1978. Repeated exposure to acrolein vapour: Subacute studies in hamsters, rats and rabbits. Toxicology 9, 47–57. 10.1016/0300-483X(78)90030-6. [PubMed: 653741]

Guan AY, Xu ZB, Wen FQ, Wang BD, Feng YL, 2008. Effect of gefinitib on airway mucus hypersecretion induced by acrolein in rats. Sichuan Daxue Xuebao (Yixue Ban) 39, 231–234. [PubMed: 18630690]

Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, Holmgren S, Pelch KE, Walker V, Rooney AA, Macleod M, Shah RR, Thayer K, 2016. SWIFT-Review: a text-mining workbench for systematic review. Syst. Rev 5, 87. 10.1186/s13643-016-0263-z. [PubMed: 27216467]

IARC (International Agency for Research on Cancer), 1995. Acrolein. In Dry cleaning, some chlorinated solvents and other industrial chemicals (pp. 337–372). Lyon, France. http://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Evaluation-Of-Carcinogenic-Risks-To-Humans/Dry-Cleaning-Some-Chlorinated-Solvents-And-Other-Industrial-Chemicals-1995.

IARC (International Agency for Research on Cancer). 2004. IARC Monographs on the Evaluation of the Carcinogenic Risk to Humans: Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs Volumes 1 to 82 (pp. 88). Lyon, France.

Jia WH, Yang P, Li J, Tian ZL, 2013. Effects of selective phosphodiesterase 4 inhibitor on expression of aquaporin 5 in airway mucus hypersecretion model of rats. Chung Hua Hsueh Tsa Chih 93, 619–622.

Khangura S, Polisena J, Clifford TJ, Farrah K, Kamel C, 2014. Rapid review: an emerging approach to evidence synthesis in health technology assessment. Int. J. Technol. Assess. Health Care 30, 20–27. 10.1017/S0266462313000664. [PubMed: 24451157]

Liu WJ, Li WC, Xu ZB, Feng YL, 2009. Effect of peroxisome proliferators activated receptor gamma and its ligand on airway mucus hypersecretion in rats. Zhonghua Jie He He Hu Xi Za Zhi 32, 282–286. [PubMed: 19576043]

Miake-Lye IM, Hempel S, Shanman R, Shekelle PG, 2016. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products [Review]. Syst. Rev 5, 28. 10.1186/s13643-016-0204-x. [PubMed: 26864942]

NASEM (National Academies of Sciences, Engineering, and Medicine), 2017. Application of systematic review methods in an overall strategy for evaluating low-dose toxicity from endocrine active chemicals. Washington, D.C.: The National Academies Press. 10.17226/24758.

NASEM (National Academies of Sciences, Engineering, and Medicine), 2018. Progress toward transforming the Integrated Risk Information System (IRIS) Program. A 2018 evaluation (2018). Washington, D.C.: The National Academies Press, 10.17226/25086.

Neghab M, Delikhoo M, Norouzian Baghani A, Hassanzadeh J, 2017. Exposure to cooking fumes and acute reversible decrement in lung functional capacity. Int. J. Occup. Environ. Med 8, 207–216. 10.15171/ijoem.2017.1100. [PubMed: 28970595]
NTP (National Toxicology Program), 2005. 11th Report on carcinogens. Research Triangle Park, NC. http://ntp-server.niehs.nih.gov/index.cfm?objectid=32BA9724-F1F6-975E-7FCE50709CB4C932.

OEHHHA (California Office of Environmental Health Hazard Assessment), 2008. Acute, 8-hour and chronic toxicity summary - acrolein (pp. 42–67). Sacramento, CA: Office of Environmental Health Hazard Assessment, California EPA.

Polisena J, Garrity C, Kamel C, Stevens A, Abou-Setta AM, 2015. Rapid review programs to support health care and policy decision making: a descriptive analysis of processes and methods. Syst. Rev 4, 26. 10.1186/s13643-015-0022-6. [PubMed: 25874967]

Radke EG, Braun JM, Meeker JD, Cooper GS, 2018. Phthalate exposure and male reproductive outcomes: A systematic review of the human epidemiological evidence [Review]. Environ. Int 121, 764–793. 10.1016/j.envint.2018.07.029. [PubMed: 30336412]

Schroeter JD, Kimbell JS, Gross EA, Willson GA, Dormon DC, Tan YM, CH III 2008. Application of physiological computational fluid dynamics models to predict interspecies nasal dosimetry of inhaled acrolein. Inhal. Toxicol 20, 227–243. 10.1080/08958370701864235. [PubMed: 18300045]

Seaman VY, Bennett DH, Cahill TM, 2007. Origin, occurrence, and source emission rate of acrolein in residential indoor air. Environ. Sci. Technol 41, 6940–6946. 10.1021/es0707299. [PubMed: 17993132]

Slaughter JC, Koenig JQ, Reinhardt TE, 2004. Association between lung function and exposure to smoke among firefighters at prescribed burns. J. Occup. Environ. Hyg 1, 45–49. 10.1080/15459620490264490. [PubMed: 15202156]

TCEQ (Texas Commission on Environmental Quality), 2016. Development support document - Acrolein CASRN: 107-02-8 Revised. Austin, TX. https://www.tceq.texas.gov/assets/public/implementation/tox/dsd/final/acrolein.pdf.

ToxNet Hazardous Substances Data Bank, 2017. HSDB: Acrolein. Bethesda, MD: National Institute of Health, U.S. National Library of Medicine. Retrieved from https://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@DOCNO+177.

U.S. EPA (U.S. Environmental Protection Agency), 1994. Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry [EPA Report]. (EPA/600/8-90/066F), Research Triangle Park, NC: U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=71993&CFID=51174829&CFTOKEN=25006317.

U.S. EPA (U.S. Environmental Protection Agency), 2002. A review of the reference dose and reference concentration processes (pp. 1–192). (EPA/630/P-02/002F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://www.epa.gov/osa/review-reference-dose-and-reference-concentration-processes.

U.S. EPA (U.S. Environmental Protection Agency), 2003. Toxicological review of acrolein [EPA Report]. (EPA/635/R-03/003). Washington, DC. https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm? substance_nbr=364.

U.S. EPA (U.S. Environmental Protection Agency). (2009). Graphical arrays of chemical-specific health effect reference values for inhalation exposures [EPA Report]. (EPA/600/R-09/061). Research Triangle Park, NC. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=211003.

U.S. EPA (U.S. Environmental Protection Agency), 2012. Advances in inhalation gas dosimetry for derivation of a reference concentration (RfC) and use in risk assessment, pp. 1–140. (EPA/600/R-12/044). Washington, DC. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=244650&CFTID=50524762&CFTOKEN=17139189.

U.S. EPA (U.S. Environmental Protection Agency), 2018a. 2014 National Air Toxics Assessment (NATA). Available online at https://www.epa.gov/national-air-toxics-assessment (accessed August 27, 2018).

U.S. EPA (U.S. Environmental Protection Agency), 2018b. 2014 National Emissions Inventory (NEI) data (Version 2). Washington, DC. Retrieved from https://www.epa.gov/air-emissions-inventories/2014-national-emissions-inventory-nei-data.

U.S. EPA (U.S. Environmental Protection Agency). (2019). U.S. EPA Chemistry Dashboard: Acrolein. Available online at https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5020023.
Wang SL, Yuan B, Dan QQ, Yang XY, Meng BL, Zhang YH. 2010. Effects of enalapril on IL-1 beta, IL-6 expression in rat lung exposure to acrolein. Sichuan Daxue Xuebao (Yixue Ban) 41 (1003–1007), 1038.

Wang X, Li JQ, Chen L, WangHX, Sun BB, Wang T, Xu D, Wen FQ. 2008. Effects of phosphodiesterase-4 inhibitor on the mucin secretion in airway stimulated with acrolein: Experiment with rats. Chung Hua Hsueh Tsa Chih 88, 2988–2993.

Wei M, Tu L, Liang Y, Li J, Gong Y, Zhang Y, Yang L. 2015. Changes of CD4(4(+)) Foxp3+ regulatory T cells and CD4(4)(+)IL-17+T cells in acrolein exposure rats. Zhonghua Laodong Weisheng Zhiye Zazhi 33, 652–657. [PubMed: 26832697]

Wei M, Tu L, Liang YH, Liu J, Zhang JH, Gong YJ. 2012. Effect of acrolein exposure on the percentage of CD4CD25 regulatory T cells and expression of transcription factor Foxp3 in asthmatic rats. Zhonghua Yufang Yixue Zazhi 46, 736–739. [PubMed: 23157870]

Wolfife TAM, Whaley P, Halsall C, Rooney AA, Walker VR. 2019. Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. Environ. Int 130, 104871. 10.1016/j.envint.2019.05.065. [PubMed: 31254867]

Xu ZB, Feng YL, Xiao J, Tang YJ, Qiu T. 2008. Interventions of enalapril on airway inflammation in rat models induced by acrolein. Sichuan Daxue Xuebao (Yixue Ban) 39, 583–587. [PubMed: 18798499]

Yasuo M, Droma Y, Kitaguchi Y, Ito M, Imamura H, Kawakubo M, Hanaoka M. 2019. The relationship between acrolein and oxidative stress in COPD: in systemic plasma and in local lung tissue. Int. J. Chronic Obstruct. Pulmonary Disease (Online) 14, 1527–1537. 10.2147/COPD.S208633.

Yoshida M, Mikami T, Higashi K, Saiki R, Mizoi M, Fukuda K, Nakamura T, Ishii I, Nishimura K, Toida T, Tomitori H, Kashiwagi K, Igarashi K. 2012. Inverse correlation between stroke and urinary 3-hydroxypropyl mercapturic acid, an acrolein-glutathione metabolite. Clin. Chim. Acta 413, 753–759. 10.1016/j.cca.2012.01.020. [PubMed: 22293277]

Yost EE, Euling SY, Weaver JA, Beverly BEJ, Keshava N, Mudipalli A, Arzuaga X, Blessinger T, Dishaw L, Hotchkiss A, Makris SL. 2019. Hazards of diisobutyl phthalate (DIBP) exposure: A systematic review of animal toxicology studies [Review]. Environ. Int 125, 579–594. 10.1016/j.envint.2018.09.038. [PubMed: 30591249]

Yuan B, Dan QQ, Zhao S, Wang SL, Meng BL, Zhang YH, 2010. Effects of acrolein exposure on the expression of Muc5ac in the airway of rats. Sichuan Daxue Xuebao (Yixue Ban) 41, 994–997. [PubMed: 21265101]

Yuan JM, Gao YT, Wang R, Chen M, Carmella SG, Hecht SS. 2012. Urinary levels of volatile organic carcinogen and toxicant biomarkers in relation to lung cancer development in smokers. Carcinogenesis 33, 804–809. 10.1093/carcin/bgs026. [PubMed: 22298640]
Fig. 1.
Acrolein emissions (in tons) reported by sector from the 2014 National Emissions Inventory (NEI). Mobile sources include a wide variety of vehicles, engines, and equipment that generate air pollution and that move, or can be moved, from place to place; examples include cars, trucks, buses, earth-moving equipment, lawn and garden power tools, ships, railroad locomotives, and airplanes.
Fig. 2.
Comparison of acrolein inhalation reference values. Line segments indicate relevant durations for individual reference values. Categories for the reference values based on their intended purpose are shown in the legend – red for Emergency Response, gold for Occupational, and green for values applicable to the General Public. Abbreviations: ACGIH = American Conference of Governmental Industrial Hygienists; AEGL = Acute Exposure Guideline Level; ATSDR MRL = Agency for Toxic Substances and Disease Registry Minimal Risk Level; CA-REL = California Environmental Protection Agency Reference Exposure Level; ERPG = Emergency Response Planning Guideline; IRIS RfC = Integrated Risk Information System Reference Concentration; NIOSH IDLH = National Institute for Occupational Safety and Health Immediately Dangerous to Life or Health Value; NIOSH REL (TWA) = NIOSH Recommended Exposure Limit (Time Weighted Average); NIOSH STEL = NIOSH Short-Term Exposure Limit; OSHA PEL (TWA) = Occupational Safety and Health Administration Permissible Exposure Limit (Time Weighted Average); TCEQ ReV = Texas Commission on Environmental Quality Reference Value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. 3.
Study evaluation approach for experimental animal and epidemiology studies.
Fig. 4.
Study flow selection diagram.
Fig. 5. 
**HAWC Literature Inventory Tree.** Available in interactive format at [https://hawcprd.epa.gov/lit/assessment/10000047/references/visualization/](https://hawcprd.epa.gov/lit/assessment/10000047/references/visualization/).
**Fig. 6.**
Summary of studies by evidence type, study design and health systems assessed. Click https://public.tableau.com/profile/literature.inventory#!/vizhome/AcroleinEvidenceMapVisualizations/ReadMe to view the interactive version studies.
Fig. 7.
Animal study evaluation. Click to see more detailed rationales for the ratings are available at https://hawcprdr.epa.gov/summary/visual/10000103/.
Fig. 8.
Epidemiology study evaluation. Click to see more detailed rationales for the ratings are available at https://hawcprd.epa.gov/summary/visual/100500038/.
Fig. 9.
Effects of acrolein on respiratory tract histopathology (interactive figure available at https://hawcprd.epa.gov/summary/data-pivot/assessment/100000047/acrolein-respiratory-effects/).
Table 1
Populations, Exposures, Comparators, and Outcomes (PECO) Criteria.

| P | Human: Any population and lifestage (occupational or general population, including children and other sensitive populations). Animal: Nonhuman mammalian animal species (whole organism) of any lifestage (including preconception, in utero, lactation, peripubertal, and adult stages). Surgically or genetically-modified mammalian models were screened as PECO relevant. |
| E | Human: Direct exposure to acrolein via inhalation. People can also be exposed to acrolein from cigarette smoking, but these studies are not considered to meet PECO criteria unless the specific inhalation level of acrolein exposure is presented and quantified. Animal: Exposure to acrolein via inhalation. Studies involving exposure to mixtures will be included only if they include an arm with exposure to acrolein alone. |
| C | Human: A comparison or reference population exposed to lower levels (or no exposure/exposure below detection limits) of acrolein or exposed to acrolein for shorter periods of time. However, worker surveillance studies are considered to meet PECO criteria even if no referent group is presented. Case reports describing findings in 1–3 people in non-occupational or occupational settings will be tracked as “potentially relevant supplemental material”. Epidemiological or controlled-exposure studies were both included. Animal: A concurrent control group exposed to vehicle-only treatment and/or untreated control (control could be a baseline measurement, e.g., acute toxicity studies of mortality). |
| O | All health outcomes (both cancer and noncancer). |

Classical Pharmacokinetic (PK) or Dosimetry Model Studies: Classical PK or dosimetry modeling usually divides the body into just one or two compartments, which are not specified by physiology, where movement of a chemical into, between, and out of the compartments is quantified empirically by fitting model parameters to ADME data. This category is for papers that provide detailed descriptions of PK models, that are not a PBPK model.

- The data are typically the concentration time-course in blood or plasma after oral and/or intravenous exposure, but other exposure routes can be described.
- A classical PK model may be elaborated from the basic structure applied in standard PK software, for example to include dermal or inhalation exposure, or growth of body mass over time, but otherwise does not use specific tissue volumes or blood flow rates as model parameters.
- Such models can be used for extrapolation like PBPK models, though such use may be more limited.

Note: ADME studies often report classical PK parameters, such as bioavailability (fraction of an oral dose absorbed), volume of distribution, clearance rate, and/or half-life or half-lives. If a paper only provides such results in tables with minimal description of the underlying model or software (i.e., uses standard PK software without elaboration), including “non-compartmental analysis”, it should only be listed as a supplemental material ADME study.

Physiologically-based Pharmacokinetic (PBPK) or Mechanistic Dosimetry Model Studies: PBPK models represent the body as various compartments (e.g., liver, lung, slowly perfused tissue, richly perfused tissue) in order to quantify the movement of chemicals or particles into and out of the body (compartments) by defined routes of exposure, metabolism and elimination, and thereby estimate concentrations in blood or target tissues.

- Usually specific to humans or defined animal species; often a single model structure is calibrated for multiple species.
- Some mechanistic dosimetry models might not be compartmental PBPK models but predict dose to the body or specific regions or tissues based on mechanistic data, such as ventilation rate and airway geometry.
- A defining characteristic is that key parameters are determined from a substance’s physicochemical parameters (e.g., particle size and distribution, octanol–water partition coefficient) and physiological parameters (e.g., ventilation rate, tissue volumes); i.e., data that are independent of in-vivo ADME data which are otherwise used to estimate model parameters.
- Chemical-specific information on metabolism (e.g., Vmax, Km) or other molecular processes (e.g., protein binding) may be obtained by fitting the model to in vivo ADME data or determined from in vitro experiments and extrapolated to in vivo predictions.

They allow extrapolation between species, routes of exposure, or exposure durations and levels; i.e., they don’t just quantify ADME for specific experiments to which they’ve been fitted.
| Category | Evidence |
|----------|----------|
| In vitro, ex vivo, or in silico "mechanistic" studies | In vitro, ex vivo, or in silico studies reporting measurements related to a health outcome that inform the biological or chemical events associated with phenotypic effects, in both mammalian and non-mammalian model systems. *There is a vast mechanistic literature on acrolein, much of which focuses on its endogenous production through oxidative reactions and involvement in oxidative stress responses. For this reason, we focused our tagging of supplemental mechanistic studies on those studies that directly assessed the effects on exogenous treatment of acrolein in mechanistic model systems. |
| Non-mammalian model systems | Studies in non-mammalian model systems, e.g., fish, birds, C. elegans |
| Toxicokinetic (ADME) | Toxicokinetic (ADME) studies are primarily controlled experiments, where defined exposures usually occur by intravenous, oral, inhalation, or dermal routes, and the concentration of particles, a chemical, or its metabolites in blood or serum, other body tissues, or excreta are then measured. • These data are used to estimate the amount absorbed (A), distributed to different organs (D), metabolized (M), and/or excreted/eliminated (E) through urine, breathe, feces. • The most informative studies involve measurements over time such that the initial increase and subsequent concentration decline is observed, preferably at multiple exposure levels. • Data collected from multiple tissues or excreta at a single time-point also inform distribution. • ADME data can also be collected from human subjects who have had environmental or workplace exposures that are not quantified or fully defined. However, to be useful such data must involve either repeated measurements over a time-period when exposure is known (e.g., is zero because previous exposure ended) *or* time- and subject-matched tissue or excreta concentrations (e.g., plasma and urine, or maternal and cord blood). • ADME data, especially metabolism and tissue partition coefficient information, can be generated using in vitro model systems. Although in vitro data may not be as definitive as in vivo data, these studies should also be tracked as ADME. For large evidence bases it may be appropriate to separately track the in vitro ADME studies. |
| Non-inhalation routes of exposure | Experimental studies utilizing a non-inhalation route of administration. This categorization generally does not apply to epidemiological studies where the exposure source may be unclear. Such studies are tracked as PECO relevant when inhalation exposure is plausible. |
| Exposure characteristics (no health outcome assessment) | Exposure characteristic studies include data that are unrelated to toxicological endpoints, but which provide information on exposure sources or measurement properties of the environmental agent (e.g., demonstrate a biomarker of exposure). |
| Mixture studies | Mixture studies do not meet the PECO criteria because they do not contain an exposure or treatment group assessing only the chemical of interest. However, they may still be useful for hazard characterization and identifying data gaps. |
| Case studies | Case reports describing health outcomes after exposure will be tracked as "potentially relevant supplemental material" when the number of subjects is ≤3. |
| Records with no original data | Records that do not contain original data, such as other agency assessments, informative scientific literature reviews, editorials or commentaries. |
| Conference abstracts | Records that do not contain sufficient documentation to support study evaluation and data extraction. |