Published in final edited form as:

J Expo Sci Environ Epidemiol. 2019 October ; 29(6): 842–851. doi:10.1038/s41370-018-0079-0.

Verifying locations of sources of historical environmental releases of dioxin-like compounds in the U.S.: implications for exposure assessment and epidemiologic inference

Rena R. Jones, PhD¹, Trang VoPham, PhD^{2,3}, Boitumelo Sevilla, BS, MPS⁴, Matthew Airola, MA⁴, Abigail Flory, MS⁴, Nicole C. Deziel, PhD⁵, John R. Nuckols, PhD⁶, Anjoeka Pronk, PhD⁷, Francine Laden, ScD^{2,3,8}, Mary H. Ward, PhD¹

¹Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health (NIH), Bethesda, Maryland, United States ²Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States ³Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States ⁴Westat Inc., Rockville, Maryland, United States ⁵Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, United States ⁶JRN Environmental Health Sciences, North Bethesda, Maryland, United States ⁷TNO, Netherlands Organization for Applied Scientific Research, The Hague, Netherlands ⁸Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States

Abstract

Polychlorinated dibenzo-*p*-dioxin and dibenzofuran (PCDD/F) emissions from industrial sources contaminate the surrounding environment. Proximity-based exposure surrogates assume accuracy in the location of PCDD/F sources, but locations are not often verified.

We manually reviewed locations (i.e., smokestack geo-coordinates) in a historical database of 4 478 PCDD/F-emitting facilities in 2009 and 2016. Given potential changes in imagery and other resources over this period, we re-reviewed a random sample of 5% of facilities (n=240) in 2016. Comparing the original and re-review of this sample, we evaluated agreement in verification (location confirmed or not) and distances between verified locations (verification error), overall and by facility type. Using the verified location from re-review as a gold standard, we estimated the accuracy of proximity-based exposure metrics and epidemiologic bias.

Overall agreement in verification was high (>84%), and verification errors were small (median=84 m) but varied by facility type. Accuracy of exposure classification (1 facility within 5 km) for a

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding author: Rena R. Jones, PhD, MS, Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Room 6E124, Rockville, MD 20850, (240) 276-7292, rena.jones@nih.gov.

Conflicts of interest: The authors declare they have no actual or potential competing financial interests.

hypothetical study population also varied by facility type (sensitivity: 69–96%; specificity: 95–98%). Odds ratios were attenuated 11–69%, with the largest bias for rare facility types.

We found good agreement between reviews of PCDD/F source locations, and that exposure prevalence and facility type may influence associations with exposures derived from this database. Our findings highlight the need to consider location error and other contextual factors when using proximity-based exposure metrics.

INTRODUCTION

Polychlorinated dibenzo-*p*-dioxins and dibenzofurans (PCDD/F) are persistent organic pollutants primarily emitted from industrial combustion facilities such as incinerators, smelters, kilns, and coal-fired power plants. These compounds are also by-products of the manufacture and use of polychlorinated biphenyls (PCBs), chlorine bleaching of paper and pulp, and chlorophenols and chlorophenoxy herbicide production.(1, 2) Reservoirs for these products (e.g., sewage sludge) are also a point source for dioxin and furan exposure, as they have potential for redistribution and circulation into the environment through various physical processes.(3) Ambient emissions from PCDD/F sources may also be transported and deposited away from their origin.(4, 5) General population exposures to PCDD/F are thought to primarily occur through dietary ingestion of meat, milk and other dairy products following environmental contamination and subsequent bioaccumulation in the food chain.(6) However, PCDD/F have been found in soil samples close to operating municipal solid waste incinerators (7–9) and in dust samples from homes near these and other industrial dioxin sources.(10, 11) Ingestion, inhalation, and dermal contact with dust and soil could therefore all contribute to PCDD/F exposures for individuals living in close proximity to sources. Moreover, as levels in food have declined over time,(6, 12) the relative contribution of non-dietary environmental exposures sources may have increased.

The International Agency for Research on Cancer has classified PCDD/F as human carcinogens based on epidemiologic studies of all cancers, as well as specific associations with lung, breast, and lymphohematopoietic malignancies.(2) Much of this evidence comes from studies in occupational settings or in populations exposed to high levels from accidental environmental releases. Ambient PCDD/F exposures have been infrequently studied in relation to cancer in the general population, and the epidemiologic literature reflects mixed evidence of associations with proximity-based or modeled exposures from point sources. Most positive associations are for non-Hodgkin lymphoma (NHL) and other lymphohematopoietic malignancies. An industrial accident and environmental release of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in Seveso, Italy was associated with an increased risk of NHL among residents living in contaminated zones surrounding a chemical plant.(13) In France, significantly elevated NHL incidence rates were found in a population living close to 13 municipal solid waste incinerators (MSWI).(3) In contrast, an ecologic study in Great Britain found that NHL incidence was not significantly increased within 3 km of 72 municipal solid waste incinerators.(14) A U.S. population-based case-control study found a positive association between NHL and serum levels of furans.(15) In this same study population, residential proximity <0.5 mi (0.8 km) from lumber and wood product

manufacturing facilities(16) and within 3 and 5 km of cement kilns(17) were also positively associated with NHL.

Exposure studies demonstrate that air emissions of PCDD/F from industrial point sources are dispersed, transported, and deposited into the surrounding environment,(4, 5, 7–9, 18–21) thus providing a rationale for the use of proximity-based exposure estimates in epidemiologic investigations. However, an important driver of the validity of such metrics is the accuracy of the source location, which may be challenging to determine for historical PCDD/F sources most relevant to studies of cancer. The U.S. Environmental Protection Agency (USEPA)'s database of sources of environmental releases of dioxin-like compounds in the United States includes the geographic location and emissions for the majority of PCDD/F sources in the U.S.(22) In 2009, about 75% of facility locations in this database were manually reviewed and/or corrected for linkage to a population-based case-control study of NHL in four National Cancer Institute (NCI) U.S. Surveillance, Epidemiology, and End Results (SEER) centers (hereafter the NCI-SEER NHL Study) and two subsequent exposure studies within this population;(10, 11, 17) the remaining facilities were reviewed in 2016 using the same protocol. This database is a potentially useful resource for studies of populations across the U.S., as it represents the only comprehensive assessment of PCDD/F sources appropriate for assessing general population exposure. However, subjectivity in the manual location review process and changes in the availability of spatial reference data over time could have introduced errors that have implications for exposure assessment and inference in epidemiologic studies. In the current study with a sample of facilities from the database, we used a single reviewer and additional resources to determine how consistently facility locations could be identified compared to the original reviews. We also evaluated the impact of differences in the assignment of facility locations on the accuracy of proximity-based exposure metrics and subsequent bias in epidemiologic measures of association.

MATERIALS AND METHODS

Dioxin database

The USEPA dioxin emissions database (obtained directly from USEPA environmental scientist David Cleverly in 2008) has been previously described.(17, 22) This national database of formation (i.e., non-reservoir) sources of PCDD/F contains the facility name, latitude and longitude (in decimal degrees) for the smokestack location, city, county, state, and estimated air emissions of PCDD/F (ng toxic equivalency quotient/year) in 1987 and 1995 from 10 facility types: secondary copper smelters, MSWI, ore sintering plants, medical waste incinerators (MWI), coal-fired power plants, sewage sludge incinerators (SSI), hazardous waste incinerators (HWI), industrial boilers, and cement kilns. These 10 types of facilities accounted for over 85% of air emissions of PCDD/F in the U.S. (22, 23). In 2009, we identified an additional N=842 hospitals that were presumed to have their own MWI prior to the mid-1990s (24) through an Environmental Systems Research Institute (ESRI) database of hospital locations. The full database (hereafter dioxin database) of N=4 478 facilities includes N=3 636 from USEPA and these additional MWI.

Original review and verification of facility locations in the dioxin database

As partially described previously,(17) we originally reviewed facility locations in a Geographic Information System (GIS) by comparing the coordinates to locations determined through publicly available web-based aerial photographs and satellite imagery and through ancillary information from internet searches. After converting all coordinates to a common projected coordinate system and datum (USA Contiguous Albers Equal Area Conic; ESRI WKID 102003), the stepwise process was as follows. First, we used current Google Earth® imagery to determine if the database coordinates were spatially concurrent with an observed facility location. A matching facility was defined as: 1) facility visually matched expectation (e.g., a MWI appears to be a hospital, or stacks were visible at other facility types, such as coal-fired power plants); 2) the facility name either matched the name in the original database or internet searches confirmed that the facility name had changed to the name available in Google Maps®. Second, if a facility's coordinates were inconsistent with Google Maps imagery, we conducted internet searches for references to the facility using a combination of facility name, city, county, and/or state. We then used historical imagery in Google Earth® to verify potential matches during a mid-1990s time frame. If the location was still not identified, a third and last step was to search the USEPA web site(25) for the facility name; if documents referencing the facility were available, they were examined for additional details. When facility locations were identified through any of these processes, they were considered "verified". If the facility location was determined to have different coordinates, the coordinates were corrected; where possible, points were placed at a smokestack. If multiple stacks were visible or when no stack was visible, points were placed at the center of the facility footprint that was often comprised of multiple buildings. If a facility location could not be confirmed using any of these procedures, it was considered "not verified".

The original reviews were conducted in two separate efforts in 2009 and 2016. In 2009, N=3450 facilities were reviewed as part of the NCI-SEER NHL Study that included cases and controls from four geographic areas (metropolitan areas of Detroit and Seattle, Los Angeles county, and Iowa).(17) In 2016, the remaining N=1 028 facilities were reviewed through the same procedures so that the dioxin database could be improved for additional epidemiologic analyses.

Re-review sample

The goal of the original review described above was to confirm the facility location, which could include multiple buildings and/or emissions stacks and therefore required reviewer judgment about where to assign the coordinates based on visual inspection of a map. Even when the location in the USEPA dioxin database was generally correct, the review process included refining the location of the stack where feasible. In the re-review, our goal was to determine how consistently we could identify facility locations compared to the original review, given these and other subjective decisions and possible changes in available reference information (e.g., maps, web sites) between 2009 and 2016.

From the N=4 478 facilities in the dioxin database, we sampled N=240 for re-review, including N=200 where the facility location had been verified and N=40 that were not

verified during the original review (Table 1). Because facilities had been reviewed at different time points and by different analysts, we randomly sampled the 200 previously verified facilities approximately proportionate to the number of facilities reviewed in 2009 and 2016 (75% and 25%, respectively; Table 1). We similarly selected 40 unverified facilities, oversampling MSWI (N=20) since they are the leading source of PCDD/F emissions across facility types in the dioxin database. Re-review followed the same general procedures as described above, but was a more labor-intensive effort that involved additional internet searches. These included queries of the facility name or type within the original city/state/county listed in the database, of municipal public works/engineering web sites in these locales (especially relevant for MSWI or sewage sludge facilities, which are managed at the local level), and the Energy Justice Network web site, which maintains detailed information on some of these facilities, especially coal-fired power plants.(26)

Comparisons between original and re-review of facilities

We used several metrics to compare the re-review of the 240 sampled facilities to the original review. We computed the percentages of facility locations verified exclusively in the original review or in the re-review sample, respectively. We also summarized locations that were verified or not verified in both efforts, and computed the crude percent agreement [sum of facilities mutually verified and not verified in original review and re-review / total].

Among the subset of facilities that were verified in both the original and re-review, we evaluated the verification error for all facilities overall and by facility type. Verification error was defined as the Euclidean distance (m) between the originally verified location and the re-reviewed location for each facility, and was calculated using ArcGIS (version 10.4.1; ESRI, Redlands, CA). Because locations for several facility types (cement kilns emitting hazardous waste, industrial boilers, iron ore sintering facilities, and secondary copper smelters) had been reviewed in 2009, and because the reviews were conducted by different analysts, we also separately evaluated verification errors for facilities originally reviewed in 2009 and 2016 (Supplemental Table 1). Median distances for reviews in both time periods were within a range typically observed for geocoding error, approximately <50–100m (Supplemental Table 2), therefore we present results comparing the re-review sample to data combined from both the 2009 and 2016 efforts as the “original review”.

Accuracy of proximity-based exposure metrics

To assess potential exposure misclassification, we simulated a spatially isotropic setting (i.e., assumed a uniform distribution of exposure in all directions) by randomly placing hypothetical participant residence locations within 5km (“exposed”) and between 5 and 10km (“unexposed”) of facility locations. To do this, we created 5km buffers around each facility verified in both the original review and re-review sample and calculated the proportion of the population of each Census block group within the buffer [block group population / total population in buffer] based on the 2010 Census (Figure 1). For block groups intersected by the buffer, we estimated the population proportional to the area within the buffer. We then generated exposed residences within the 5km buffer by randomly selecting block groups with replacement, with the probability of block group selection determined by the proportional population of each block group within the buffer. Selected

block groups received a randomly-placed point within the block group polygon. We likewise generated the unexposed residences, limiting the geographic range to areas between 5 to 10 km around the exposed buffer. The number of random points generated was selected to capture a range of exposure to 1 or more facilities. This number was determined based on the estimated prevalence (%) of residential proximity to facilities [number of participants living within 5km of at least one facility / total number of participants] observed in the Nurses' Health Study (NHS) and NHS II, two large prospective cohorts with participant residences distributed across the U.S. and enrolled in 1976 and 1989, respectively.(27) Our choice of a 5km exposure buffer was based on a recent exposure study in the NCI SEER study population that found a positive relationship between levels of PCDD/F in house dust and emissions from municipal and MWI within this distance.(11) We used block groups instead of smaller geography (blocks) to simulate hypothetical population residences within the 5km buffers for computational efficiency; however, there were a large number of block groups and population density between block groups and blocks was similar.

We computed the sensitivity and specificity of proximity-based exposure metrics using the re-reviewed locations as the gold standard because they were reviewed by a single individual and with additional resources. Based on the verified facility locations from the original review, we first computed the proportion of exposed and unexposed residences, overall and by facility type; we repeated this process for re-reviewed facilities. We then computed the sensitivity (% of correctly classified exposed participants) and specificity (% of correctly classified unexposed participants) based on the re-reviewed facility locations. Lastly, we estimated the odd ratios (ORs) observed for the resulting sensitivities and specificities (OR_{observed}) in a hypothetical case-control analysis of NHL, assuming a 2-fold greater exposure odds among cases ($OR_{\text{true}}=2.0$). (28)

RESULTS

The dioxin database included 4 478 facilities in all 50 states (Figure 2). Of these, about half (53.0%) were MWI; coal-fired power plants (33.1%) were the next most common facility type (Table 1). The other facility types each represented less than 5% of the total, with hazardous waste-burning cement kilns and secondary copper smelters being the least common (0.5 and 0.1% of all facilities, respectively). In original review of the database, 92% of facility locations were verified (Table 1). The 240 re-reviewed facilities were distributed across 33 states (Figure 2). In the re-review sample, the originally verified proportion was lower (83%), which reflects our oversampling of previously unverified facilities, especially MSWI (Table 1). Some facilities originally reviewed in 2016 had previously been reviewed in 2009 but their locations were not verified, therefore a greater proportion of all facilities were unverified in the 2016 original review (1.4% in 2009 vs. 28.6% in 2016) (Supplemental Table 1).

Overall, 83.9% of facilities were verified in both efforts. In our re-review, we were able to verify locations for a greater percentage of facilities compared to the original review (12.3 versus 3.4% verified in re-review only and original review only, respectively; Table 2); this discordance was greatest for MSWI facilities (51.7 versus 3.5%). The crude percent agreement in verification was high overall (>84%) and varied by facility type, with slightly

better agreement for coal-fired power plants (89.2%) and lowest agreement for the MSWI locations (44.8%). A small number of facilities (7 of which were a MWI, 1 of which was a MSWI) were verified only in the original review.

Among the 192 facilities that were verified in both the original review and the re-review, the median verification error between facility locations was 84m (IQR 24–184m; Table 3). These distances varied by facility type, with the smallest errors observed between verified locations of coal-fired power plants (median=41m) and the greatest for hazardous waste incinerators (median=511m). Maximum distances were quite large, but for most facility types reflect a small number of discrepant observations. Further investigation into very large errors (e.g., >1000m, approximately the 90th percentile) indicated they were generally driven by an incorrect state, city, or county identification for the facility in the dioxin database (data not shown).

Overall, we observed high sensitivity (91.9%) and specificity (97.1%) in our classification of exposure to one or more facility within 5km of a hypothetical study population based on estimated exposure prevalence from the NHS and NHS II populations. This misclassification resulted in modest (14%) attenuation of a hypothetical OR of 2.0 (Table 4). Sensitivity and specificity were >95% for coal, HWI, and MWI and attenuation increased with decreasing exposure prevalence. The smallest bias was observed in association with proximity to MWI (prevalence=28%; OR_{observed}=1.89) and the greatest was for cement kilns (prevalence=2.0%; OR_{observed}=1.24). Exposure to MSWI (4%) and sewage sludge incinerators (4%) was rare; however, the higher specificity (>98.2%) for sewage sludge incinerators exposure compared to MSWI (95.7%) yielded less attenuation of the OR in association with this facility type.

DISCUSSION

GIS-based metrics are increasingly used as part of exposure assessment in environmental epidemiology. However, the value of these metrics may be diminished if there are errors in the location of pollutant sources used as the focal point for their estimation. In this examination of historical PCDD/F emitting source locations, we found overall good agreement and small verification errors between the original review and re-review of facilities, with variation by facility type. Given our findings in this small representative sample, a complete re-review of all facilities in the dioxin database is likely unnecessary to support its use in future epidemiologic studies. However, some facilities were more difficult to locate and our results indicate a potential for attenuation bias resulting from use of proximity-based exposure estimates when exposure prevalence is low.

The high overall percent agreement (>84%) in verification indicated that we could successfully replicate the original review of facility locations in the full dioxin database. The re-review also elucidated facility types that were consistently challenging to verify, including MWI, coal-fired power plants, and MSWI, the latter of which historically had the highest emissions of dioxins and furans (22). However, we did not consistently observe that these facility types were subject to greater exposure misclassification and attenuation bias in simulated epidemiologic analyses. This finding suggests that effect estimates derived from

proximity-based exposure metrics may be impacted more by the extent to which participants do or do not reside near certain facility types rather than location errors. We note that while MWI locations came from both the USEPA database as well as a resource identified from ESRI, we did not observe differences in verification percentages based on these two data sources (data not shown).

We selected the location identified in re-review as our “gold standard”, under the assumption that a more recent review would have access to the most information to identify facility locations. However, manual review of geo-coordinates is a subjective process, and thus all reviews of facilities included some level of error that could not be accounted for in our comparisons. Reviewers’ ability to verify a facility location might be expected to vary by facility type because of differential availability of information. For instance, we observed in the most recent review that MWI were more readily identified on maps and web sites, including historical name changes, whereas MSWI were found less easily through these searches. A small proportion of facilities (3.4% of the total sample) were verified only in the original reviews; 7 out of 8 were MWI. We speculate this discrepancy could be due to changes in hospital names arising from consolidations (29) that were not well captured between reviews; however, we acknowledge differences in reviewer proficiency as another explanation. Increased regulatory scrutiny and community concern over certain types of facilities could influence the ability to locate certain types of facilities or individual facilities. We used the Energy Justice Network web site as a resource to locate some facilities, especially coal-fired power plants, but these data were not compiled for research or regulatory purposes. Other reasons for lack of agreement could be if a facility covers a large geographic area, contains multiple emissions stacks, or did not have a visible smokestack, all of which can make it difficult to consistently place a point. Poor agreement could also reflect differences in skills between reviewers, time spent searching, quality of the imagery or other resources used, or other unmeasured sources of variation.

Distance-decay or buffer-based metrics of PCDD/F exposure assume that residential proximity confers increased exposure opportunity through direct releases or migration of dioxins in the air or by contact with polluted soil. These metrics are assessed independent of a study participant’s disease status, thus the expected effect of exposure misclassification on epidemiologic associations is a bias towards the null. For exposures with low prevalence, attenuation bias due to poor specificity is a particular concern.(30) Our evaluation demonstrated overall high sensitivity and specificity (>90%) in proximity-based exposure classification based on having one or more facilities within 5 km. However, it also indicated that the impacts of location error on epidemiologic inference can be quite substantial when exposure prevalence is low. For instance, within the range of our simulated prevalences based on existing study populations, we observed that associations with proximity to facilities with <5% prevalence are impacted more severely than those for more common facility types. In contrast, less attenuation bias was apparent for the more prevalent facility types at similar sensitivities and specificities.

Others have shown that spatial errors can influence exposure misclassification and subsequent bias in epidemiologic studies that use geographic-derived exposures.(31, 32) These errors can reflect inaccuracies in locating both exposure sources and study

participants. For example, unaccounted for changes in residence over time has led to exposure misclassification and biased estimates of association with pregnancy outcomes. (32) Positional errors from geocoding are also potential sources of misclassification and poor inference.(33) Our findings indicate that the extent of potential bias in epidemiologic investigations using this dioxin database will depend on the geographic distribution of facilities in relation to that of the specific population under study. For example, if a study catchment area is characterized by a higher frequency of one of the more PCDD/F facility types (e.g., coal-fired power plants in the Northeast), the expected attenuation in effect estimates due to errors in facility location will be lower. Moreover, the consequences of spatial errors may also depend on the spatial gradient of the exposure of interest, such as if PCDD/F emissions from smokestacks decrease to background levels within short distances. These observations underscore the need for distance-based environmental exposure metrics to consider multiple sources of spatial error as well as important contextual factors, such as environmental fate and transport.

Our findings may also have implications for summary estimates that quantify exposure as a function of distance to PCDD/F emissions from multiple sources. For example, the average emission index (AEI) used by Deziel et al. (2017) and Pronk et al. (2013) in the NCI-SEER NHL Study was calculated as the sum of inverse distance squared-weighted emissions from all facilities within 5km (11, 13). We did not evaluate alternative exposure metrics in this study, as our primary objective was to ascertain reliability in locating a facility location. However, the expected impact of verification error on the AEI would be up- or down-weighting of emissions from certain facilities. Large verification errors in our evaluation are independent of the distance from the true facility location and a study participant's home, thus the direction of the bias in weighting emissions would be unpredictable.

One motivation to verify facility locations in the dioxin database is because of its potential value as an historical exposure assessment resource for epidemiologic studies of cancer. The intended end-user of this database includes "researchers who are interested in documented and time-specific dioxin source and emissions data".(22) It is inclusive of well-characterized emissions sources during several historical reference years, including the 1987 and 1995 data used in our analyses; data for facilities operating in 2000 are also described in a USEPA report.(22) The database is also potentially useful for studies of more contemporary PCDD/F exposures in the general population, as many of the facilities remain operative. We further demonstrated utility of the dioxin database by simulating the impacts of errors in facility location on exposure metrics in epidemiologic studies using exposure prevalence estimates from two U.S. cohorts. However, there are several other considerations for interpreting our findings as they relate to future use of the database. First, our simple exposure metrics have sources of uncertainty that could not be accounted for in this evaluation. It was not always possible to determine smokestacks from imagery; therefore, points were placed in the approximate center of an area based on reviewer judgement. The simulation also did not consider scenarios reflecting all patterns of exposures (e.g., to multiple facilities of different types) that may occur in the general population. TEQ estimates that are available for most facilities in the database would allow generation of summary exposure metrics that account for facility-specific emissions. Using smaller geography to simulate hypothetical population residences would have increased the level of geographic precision for the bias evaluation,

however, our results did not appreciably change in sensitivity analyses using blocks rather than block groups. We note that this database may not be suitable for developing air dispersion models that require smokestack height, since this information is not available in the USEPA database. Finally, despite our efforts to draw a representative set of facilities, the generalizability of our findings to the full database may be limited. Specifically, several uncommon facility types (including copper smelters, cement kilns, iron ore sintering plants, and industrial boilers) were not included in the re-review sample. Because facility types with the lowest exposure prevalence (<5%) in our sample yielded the highest estimated bias in risk associations, we suggest that manual review of locations for rare facility types is appropriate, as this is likely feasible even for studies covering large geographic areas. The uniqueness of these facilities in a given area might also make it easier to verify their location.

In this effort to demonstrate confidence in the locations of historical PCDD/F-emissions for future epidemiologic analyses, we found good agreement in the confirmation of facility locations compared to an original review. Our evaluation also suggested reasonable accuracy of exposure classification based on the originally verified location, with exceptions for some uncommon facility types. We recommend that studies incorporating proximity-based exposure metrics should consider facility type and exposure prevalence when interpreting epidemiologic associations. Additional research to verify locations of rare facility types is also warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This work was supported in part by the Intramural Research Program of the National Cancer Institute (NCI), NCI Training Program in Cancer Epidemiology (T32 CA009001), and Susan G. Komen for the Cure® (IIR13264020). Dr. Deziel was supported in part by the American Cancer Society (grant MRSG-15-147-01-CNE).

REFERENCES

1. International Agency for Research on Cancer (IARC). IARC Monographs on the evaluation of carcinogenic risks to humans. Polychlorinated dibenzo-para-dioxins and polychlorinated dibenzofurans 1997.
2. International Agency for Research on Cancer (IARC). IARC Monographs on the evaluation of carcinogenic risks to humans. 2,3,7,8-tetrachlorodibenzo-para-dioxin, 2,3,4,7,8-pentachlorodibenzofuran, and 3,3',4,4',5-pentachlorobiphenyl 2012.
3. Viel JF, Arveux P, Baverel J, Cahn JY. Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *Am J Epidemiology* 2000;152(1):13–9.
4. Floret N, Mauny F, Challier B, Arveux P, Cahn JY, Viel JF. Dioxin emissions from a solid waste incinerator and risk of non-Hodgkin lymphoma. *Epidemiology* 2003;14(4):392–8. [PubMed: 12843761]
5. Viel JF, Daniau C, Gorla S, Fabre P, de Crouy-Chanel P, Sauleau EA, et al. Risk for non Hodgkin's lymphoma in the vicinity of French municipal solid waste incinerators. *Environ Health* 2008;7:51. [PubMed: 18959776]

6. Charnley G, Doull J. Human exposure to dioxins from food, 1999–2002. *Food Chem Toxicol* 2005;43(5):671–9. [PubMed: 15778006]
7. Domingo JL, Schuhmacher M, Muller L, Rivera J, Granero S, Llobet JM. Evaluating the environmental impact of an old municipal waste incinerator: PCDD/F levels in soil and vegetation samples. *J Hazard Mater* 2000;76(1):1–12. [PubMed: 10863010]
8. Floret N, Viel JF, Lucot E, Dudermel PM, Cahn JY, Badot PM, et al. Dispersion modeling as a dioxin exposure indicator in the vicinity of a municipal solid waste incinerator: a validation study. *Environ Sci Tech* 2006;40(7):2149–55.
9. Lorber M, Pinsky P, Gehring P, Braverman C, Winters D, Sovocool W. Relationships between dioxins in soil, air, ash, and emissions from a municipal solid waste incinerator emitting large amounts of dioxins. *Chemosphere* 1998;37(9–12):2173–97. [PubMed: 9828336]
10. Deziel NC, Nuckols JR, Colt JS, De Roos AJ, Pronk A, Gourley C, et al. Determinants of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans in house dust samples from four areas of the United States. *Sci Tot Evnrion* 2012;433:516–22.
11. Deziel NC, Nuckols JR, Jones RR, Graubard BI, De Roos AJ, Pronk A, et al. Comparison of industrial emissions and carpet dust concentrations of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans in a multi-center U.S. study. *Sci Tot Evnrion* 2017;580:1276–86.
12. Malisch R, Kotz A. Dioxins and PCBs in feed and food--review from European perspective. *Sci Tot Evnrion* 2014;491–492:2–10.
13. Pesatori AC, Consonni D, Rubagotti M, Grillo P, Bertazzi PA. Cancer incidence in the population exposed to dioxin after the “Seveso accident”: twenty years of follow-up. *Environ Health* 2009;8:39. [PubMed: 19754930]
14. Elliott P, Shaddick G, Kleinschmidt I, Jolley D, Walls P, Beresford J, et al. Cancer incidence near municipal solid waste incinerators in Great Britain. *Br J Canc* 1996;73(5):702–10.
15. De Roos AJ, Hartge P, Lubin JH, Colt JS, Davis S, Cerhan JR, et al. Persistent organochlorine chemicals in plasma and risk of non-Hodgkin’s lymphoma. *Canc Res* 2005;65(23):11214–26.
16. De Roos AJ, Davis S, Colt JS, Blair A, Airola M, Severson RK, et al. Residential proximity to industrial facilities and risk of non-Hodgkin lymphoma. *Environ Res* 2010;110(1):70–8. [PubMed: 19840879]
17. Pronk A, Nuckols JR, De Roos AJ, Airola M, Colt JS, Cerhan JR, et al. Residential proximity to industrial combustion facilities and risk of non-Hodgkin lymphoma: a case-control study. *Environ Health* 2013;12:20. [PubMed: 23433489]
18. Franzblau A, Zwica L, Knutson K, Chen Q, Lee SY, Hong B, et al. An investigation of homes with high concentrations of PCDDs, PCDFs, and/or dioxin-like PCBs in house dust. *J Occup Environ Hyg* 2009;6(3):188–99. [PubMed: 19152164]
19. Hinwood AL, Callan AC, Heyworth J, Rogic D, de Araujo J, Crough R, et al. Polychlorinated biphenyl (PCB) and dioxin concentrations in residential dust of pregnant women. *Environ Sci Process Impacts* 2014;16(12):2758–63. [PubMed: 25343442]
20. Richards G, Agranovski IE. Dioxin-like pcb emissions from cement kilns during the use of alternative fuels. *J Hazard Mater* 2017;323(Pt B):698–709. [PubMed: 27839925]
21. Tue NM, Suzuki G, Takahashi S, Kannan K, Takigami H, Tanabe S. Dioxin-related compounds in house dust from New York State: occurrence, in vitro toxic evaluation and implications for indoor exposure. *Environ Poll* 2013;181:75–80.
22. U.S. Environmental Protection Agency (USEPA). An inventory of sources and environmental releases of dioxin-like compounds in the United States for the years 1987, 1995, and 2000. [Internet]. National Center for Environmental Assessment. Office of Research and Development 2006 Available at: <https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=159286> (Accessed June 3, 2018)
23. Thomas VM, Spiro TG. Peer reviewed: the u.s. Dioxin inventory: are there missing sources? *Environ Sci Technol* 1996;30(2):82a–5a.
24. Pollution prevention and compliance assistance information for the healthcare industry: Incinerators: Healthcare Environmental Resource Center; 2015 Available from: <http://www.hercenter.org/facilitiesandgrounds/incinerators.cfm>.
25. U.S. Environmental Protection Agency (USEPA). Available from: www.epa.gov.

26. Energy Justice Network. Available from: www.energyjustice.net
27. Bao Y, Bertoia ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE, et al. Origin, Methods, and Evolution of the Three Nurses' Health Studies. *Am J Pub Health* 2016;106(9):1573–81. [PubMed: 27459450]
28. Blair A, Thomas K, Coble J, Sandler DP, Hines CJ, Lynch CF, et al. Impact of pesticide exposure misclassification on estimates of relative risks in the Agricultural Health Study. *Occup Environ Med* 2011;68(7):537–41. [PubMed: 21257983]
29. Cuellar AE, Gerler PJ. Trends in Hospital Consolidation: the Formation of Local Systems. *Health Affairs* 2003;22(6):77–87.
30. M. D, Stewart PA. Recommendations for reducing the effects of exposure misclassification on relative risk estimates. *Occup Hyg* 1996;3:169–76.
31. Healy MA, Gilliland JA. Quantifying the magnitude of environmental exposure misclassification when using imprecise address proxies in public health research. *Spat Spatiotemporal Epidemiol* 2012;3(1):55–67. [PubMed: 22469491]
32. Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. *Ann Epidemiol* 2017;27(1):1–9.2. Faure E, Danjou AM, Clavel-Chapelon F, Boutron-Ruault MC, Dossus L, Fervers B. Accuracy of two geocoding methods for geographic information system-based exposure assessment in epidemiological studies. *Environ Health*. 2017;16(1):15. [PubMed: 28081893]
33. Faure E, Danjou AM, Clavel-Chapelon F, Boutron-Ruault MC, Dossus L, Fervers B. Accuracy of two geocoding methods for geographic information system-based exposure assessment in epidemiological studies. *Environ Health* 2017;16: 15. [PubMed: 28235407]

Figure 1.

Example of census block groups and residences within a 5km buffer (red circle) and adjacent 5–10 km unexposed area (blue circle) around a facility location (star).

Figure 2.
Geographic distribution of dioxin- and furan-emitting facilities in the U.S.

Table 1. Verification status from the original review of dioxin- and furan-emitting facilities, overall and by facility type.

Facility Type	Full database ¹			Sample for Re-review ^{1,2}		
	Total	Verified	Not verified	Total	Verified	Not Verified
	N	N (%)	N (%)	N	N (%)	N (%)
All facilities	4 478	4 136 (92.4)	342 (7.6)	240	200 (83.3)	40 (16.7)
Cement kiln, hazardous waste	22	22 (100)	0 (0)	2	2 (100)	0 (0)
Cement kiln, non-hazardous waste	89	89 (100)	0 (0)	7	7 (100)	0 (0)
Coal-fired power plant	1 483	1 423 (96.0)	60 (4.0)	83	74 (89.2)	9 (10.8)
Hazardous waste incinerator	112	111 (99.1)	1 (0.9)	8	8 (100)	0 (0)
Industrial boiler	34	34 (100)	0 (0)	-	-	-
Iron ore sintering	10	10 (100)	0 (0)	-	-	-
Medical waste incinerator ²	2 373	2 146 (90.4)	227 (9.6)	105	94 (89.5)	11 (10.5)
Municipal solid waste incinerator	207	157 (75.8)	50 (24.2)	29	9 (31.0)	20 (69.0)
Secondary copper smelter	3	3 (100)	0 (0)	-	-	-
Sewage sludge incinerator	145	141 (97.2)	4 (2.8)	6	6 (100)	0

¹ 2009 and 2016 verifications combined.

² Includes 842 facilities identified from an ESRI database.

- indicates facility type not sampled.

Table 2.

Agreement in verification status for facilities in the re-review sample, overall and by facility type (N=240).

Facility Type	Verified in Original Review ¹ Only		Verified in Re-review Only		Verified in Both		Not Verified in Both		Percent Agreement	
	N (%)		N (%)		N (%)		N (%)			%
All facilities ²	8 (3.4)		30 (12.3)		192 (83.9)		10 (2.9)			84.2
Cement kiln, hazardous + non-hazardous waste	0		0		9 (100.0)		0			100.0
Coal-fired power plant	0		9 (10.8)		74 (89.2)		0			89.2
Hazardous waste incinerator	0		0		8 (100.0)		0			100.0
Medical waste incinerator	7 (6.7)		6 (5.7)		87 (83.0)		5 (4.8)			87.6
Municipal solid waste incinerator	1 (3.5)		15 (51.7)		8 (27.6)		5 (17.2)			44.8
Sewage sludge incinerator	0		0		6 (100.0)		0			100.0

¹ 2009 and 2016 reviews combined.

² Percentages for all facilities are weighted by the sampling proportions for each facility type.

Verification error (distance in meters) between locations of facilities verified in both the original review¹ and re-review sample (N=192), overall and by facility type.

Table 3.

Facility Type	Distance (m)							
	N	Min	Q25	Mean	Median	Q75	P95	Max
All facilities	192	0	24	3171	84	184	5372	357423
Cement kiln, hazardous + non-hazardous waste	9	33	115	1334	448	1208	5249	5249
Coal-fired power plant	74	0	19	5497	41	92	5372	357423
Hazardous waste incinerator	8	45	159	457	511	581	1112	1112
Medical waste incinerator	87	0	20	1623	101	189	1670	81077
Municipal solid waste incinerator	8	5	12	258	66	386	1135	1135
Sewage sludge incinerator	6	5	70	6454	120	15391	23020	23020

¹ 2009 and 2016 reviews combined.

Table 4.

Sensitivity and specificity of proximity metrics (yes/no 1 facility within 5km) and attenuation in an odds ratio (OR) of 2.0, overall and by facility type, for hypothetical study participants.

Facility Type	Prevalence within 5km (%) ¹	Sensitivity	Specificity	OR _{observed} ²
All facilities	8.0	91.9	97.1	1.72
Cement kiln, hazardous + non-hazardous waste	2.0	69.0	95.8	1.24
Coal-fired power plant	7.0	95.5	98.3	1.80
Hazardous waste incinerator	1.0	95.2	97.9	1.31
Medical waste incinerator	28.0	95.0	97.2	1.89
Municipal solid waste incinerator	4.0	90.3	95.7	1.46
Sewage sludge incinerator	4.0	70.6	98.2	1.60

¹Based on facilities verified in both the original review (2009 and 2016 combined) and re-review sample (N=192); Gold standard is the verified location from re-review.

²Where OR_{true}=2.0.