FINAL PROGRESS REPORT

1 R03 CA83071-01
Agricultural Chemical Exposures & Childhood Cancer

J.R. Nuckols, Principal Investigator

8 MARCH 2002

Environmental Health Advanced Systems Laboratory
Department of Environmental Health
Colorado State University
Fort Collins, Colorado

TABLE OF CONTENTS

Acknowledgments

This project and report could not have been completed without the contribution of the following individuals:

- Dr. Peggy Reynolds, Epidemiologist and Co-Principal Investigator, California Department for Health Services, Oakland, California
- Mr. Ryan Miller, GIS Analyst, EHASL, Department of Environmental Health, Colorado State University
- Mr. Robert Gunier, Environmental Scientist, California Department for Health Services, Oakland, California
- Dr. Stephanie Weigel, Geographer, EHASL, Department of Environmental Health, Colorado State University
- Mr. Andrew Hertz, GIS Analyst, California Department for Health Services, Oakland, California
- Dr. Mary H. Ward, Epidemiologist and Consultant, Occupational Epidemiology Branch, National Cancer Institute
- Ms. Julie Von Behren, Epidemiologist, California Department for Health Services, Oakland, California

Background

In regions of intense agricultural production, pesticide-related risks have become an increasingly important topic of public health concern. Exposure to agricultural chemicals have been linked to disease outcomes such as cancer, immune system disorders, adverse reproductive outcomes, developmental disorders, and neurological disease (Keifer et al., 1997; Zahm et al., 1997). Location and type of crops can be important factors in determining pesticide type, and application rates and methods in such regions (Ward et al., 2000). For example, recent studies have demonstrated that levels of specific pesticides in residential house dust can be related to proximity of crop production where the same pesticide was applied (Simcox et al., 1995; Fenske et al. 2000). Proximity to chemical releases in the environment is commonly used as a surrogate for exposure in environmental epidemiological studies. If proximity can be combined with knowledge of the type of chemicals applied and their transport characteristics, better exposure metrics can be developed for epidemiological studies of agricultural chemicals (Ward et al., 2000).

The State of California has developed a Pesticide Use Reporting Database (CPUR), with an expressed objective of providing "a complete pesticide use data for evaluating possible human illness clusters in the future" (CDPR 1995). The reporting unit for the database is approximately 2.6 km^2 (1.0 mi^2) in area. While exposure metrics calculated at this level of spatial resolution may be useful in hypothesis generation, they could result in significant misclassification of exposure in an epidemiological study requiring exposure classification at the individual level (Bell et al., 2001). A further limitation of CPUR is that pesticide use is not linked to a specific spatial database that identifies the field where the pesticide application occurred.

Specific Aims

(1) To use remote sensing and geographic information system (GIS) technology to map crop locations in California in order to link pesticide use reported in CPUR to crops located near residences in the case-control study;

(2) To evaluate the exposure metric(s) from Specific Aim 1 in terms of: (a) integration with more advanced computer models for pesticide transport in the environment, (b) reduction in misclassification of exposure, and (c) transferability across the State of California; and

(3) To conduct a pilot, registry-based epidemiological study using the exposure metric(s).

Description of Data

We employed four datasets in our study, the California Pesticide Use Reporting Database (CPUR), the California Department of Water Resources crop map database (CDWR), the USGS National Land Cover Database, and a dataset of residence location from a childhood leukemia study conducted by the California Department for Health Services (Reynolds, et al., 2002). The CPUR and CDWR databases are accessible to the public [http://www.cdpr.ca.gov and http:// www.water.ca.gov].

The CPUR database system, established in the 1950's, contains tabular information on pesticide applications, agricultural and nonagricultural. Prior to 1990, reporting of agricultural chemical use in California was required only for a set of restricted use pesticides. Beginning in 1990, a full use reporting system was instituted, requiring applicators to report all chemicals used in agriculture (California Department of Pesticide Regulation, 2000). The data is compiled at the county level on an annual basis, and contains information on the type and amount of pesticide applied, the date and method of application, and the crop to which the chemical was applied. The reporting unit for the database is a Meridian-Township-Range-Section (MTRS) in the United States Public Land Survey System. An MTRS, referred to as a Section, is a fixed-boundary parcel of land approximately 2.6 square kilometers (1.0 mi^2) in area. The CPUR data used in our study was checked for errors and corrected using the method reported by Gunier (2001).

The CDWR database is composed of a set of digital GIS coverages[1] (ArcInfo, ESRI, Redlands, CA). Coverages are published at the county level, and are available for 38 counties in California for intermittent years between 1976 and 2000. Generally, an updated map for agricultural counties is published every 5-7 years. Agricultural field boundaries, delineated from aerial photography, are used as the reporting unit. In contrast to the reporting unit for CPUR, the minimum mapping unit of the CDWR is 0.81 hectares (0.003 mi^2). The CDWR database is attributed with crop classification data for each field based on 100% verification by field crews (once a year, usually in summer between July and September). The CDWR land use classification scheme contains 83 different land cover classes, including specific crop type. No accuracy statistics are reported for the CDWR database. However, the data is collected using a 100% ground verification procedure, and CDWR personnel report that the error should be minimal (Mr. Tom Hawkins, CDWR, personal communication, March, 2001). In our analysis, we assume an error of less than or equal to 5%.

The National Land Cover Database (NLCD) is derived from remotely sensed satellite imagery data and aerial photography. In most cases the database is developed from Landsat imagery, for California imagery between 1992-1994 was used. The dataset identifies relatively broad land-cover / land-use categories such as "orchards and vineyards" and "row crops". Currently the data is in production and pending completion for the entire United States. Status and more details on the database can be obtained from the Internet at http://landcover.usgs.gov/natllandcover.html .

The CDHS dataset was obtained from registry data for 308 cases of cancer for children less than 5 years old that were diagnosed between 1988 and 1997, and 607 controls from the vital statistics registry for the six county study region. We were able to map the location of residence at time of birth for 852 subjects (93%) in a GIS. Controls were

[1] A GIS coverage is defined as a spatial database possessing both spatial attributes and tabular data defining the spatial attributes. See [cite] for in-depth description of GIS technology.

randomly selected from state birth certificates, and matched to cases based on sex and date of birth. We used a subset of sixty-six (66) of these residences in the metric evaluation component of our study (Specific Aim 2). These residences were located in counties that had CDWR crop maps in a year between 1988 and 1994 (San Joaquin (1988), Kings (1991), and Fresno (1994)). We used a subset of 582 residences in the pilot epidemiology component of our study (Specific Aim 3). These data were based on a case-control study of children under 5 years of age diagnosed with cancer in California between 1988 and 1997. Two controls were randomly selected for each case. Maternal residence at birth for participants born between 1988 and 1994 in the six (6) county study area. was obtained from birth certificates and address geocoded.

Summary: Methods and Findings

Specific Aim 1: To use remote sensing and geographic information system (GIS) technology to map crop locations in California in order to link pesticide use reported in CPUR to crops located near residences in the case-control study.

1.1. Feasibility of Linking Spatial Crop Map Data to CPUR. Two GIS-based datasets of crops have been constructed for a 6 county study region in the Central Valley of California (Figure 1). One dataset, the National Land Cover Database (NLCD) is derived from remotely sensed satellite imagery data (a composite of 1992-1994 image data), and contains relatively broad land cover categories such as "orchards and vineyards" and "row crops". The advantage of this database is that coverage is for the entire State of California, making any exposure metric derived from this data potentially applicable at the statewide level. This is important when studying rare disease such as childhood cancer, since it minimizes the number of years needed to attain a sufficient number of cases to meet statistical power requirements. EROS data center is scheduled to publish another NLCD database in 2001. The other dataset, the California Department of Water Resource (CDWR) database has finer resolution, mapping specific crops (grapes, cotton, corn etc.) with a minimum mapping resolution of 0.003 mi^2. However, this data is only available for 1 or 2 years in each county in the Central Valley for the period of the RO3 study (1988-1996). CDWR attempts to publish revised land cover datasets for each county every 5-7 years.

We linked both of the mapped datasets to the California Pesticide Use Reporting Database (CPUR) using relational database and GIS technology. Examples of the results of this linkage for location of grape cultivation from the CDWR database, and of pesticide applications reported in San Joaquin County for 1988, are presented in Figure 2. In these examples we display CPUR data for all pesticides, but we have the capability to provide the same information for any specific pesticide. Also, the resolution of the examples is approximately 1 square mile for each cell, which is the reporting unit for CPUR . We have developed the capability to link and display CPUR data at the subsection section level. However, due to the reporting limitations inherent to CPUR, chemical data used at this level is based on all applications for the section. For example, an application rate of X lbs/acre of Chemical Y for a single vineyard in any one of the sections in Figure 2 would be based on the mean or median application rate calculated for all vineyards in the same section of land.

We tested the ability to reclassify CDWR, CPUR and NLCD into common land cover classification categories. This reclassification is necessary in order to be able to calculate a CPUR-based exposure metric integrated with crop location data. We were able to reclassify each of the databases into 13 categories of agricultural land use (Table 1). For exposure metrics where only CPUR and CDWR databases will be used there are 10 broad categories of agricultural land use that overlap, and approximately 50 specific crop types that overlap.

1.2 Spatial Accuracy of CPUR. Because any exposure metric derived from CPUR will be a function of proximity, we tested the spatial accuracy of CPUR data by comparing it by crop with high precision, 100% ground verified land-use data collected by the California Department of Water Resources (CDWR). CDWR identifies 83 specific crop types with a minimum mapping unit of 0.003 mi^2. The reporting unit for the CPUR database is Public Land Survey System Section, which is approximately 1.0 mi^2. We used a Geographic Information System (GIS), to conduct a comparative analysis of the location of pesticide application by crop, as reported in CPUR, with the location of the same crops in the CDWR database. We conducted the comparative analysis for ten crops (Table 2) with the greatest number of pesticide applications for two counties, San Joaquin (1988) and Kings (1991)

To assess accuracy of the CPUR data, we developed a GIS-based analytical procedure designed to calculate spatial congruence between the two datasets at the Section level. Congruence between the two data sets results when a specific crop reported in CPUR as having a pesticide application is located in a Section where the crop is reportedly grown according to CDWR. By applying this procedure we created a GIS coverage identifying sections *with congruence* and sections *without congruence* for each crop type and each county. From the results of this analysis, we synthesized the data into two categories: non-permanent crop species (i.e. row and field species) and semi-permanent crop species (i.e. orchard and vineyard species) and calculated summary statistics on the degree of agreement (congruence) between the CDWR and CPUR databases for each crop and category using the following equation:

$$\text{CONGRUENCE} = 1 - RE_k$$

Where,

$$RE_k = \left(\frac{C}{T}\right)$$

and,
RE = reporting error (the degree to which data reported in CPUR data was incongruent with that reported by CDWR

k = crop type (Table 2) or crop category (non- vs semi-permanent).
C = the number of sections with pesticide applications reported to crop k in CPUR, but crop k is not present in the section according to CDWR data (non-congruent Sections).
T = the total number of sections with pesticide applications reported to crop k in CPUR.

Values of RE range from 0.0 to 1.0, with 0.0 reflecting the greatest degree of congruence. We tested the congruence for specific crops and crop categories (permanent versus non-permanent) for statistical significance using a one sided binomial test (Johnson and Kotz, 1969) for two cases, one assuming 1% and the other 5% error in the CDWR database. We assumed the error in CDWR was distributed uniformly across each county, regardless of crop type.

We evaluated the general agreement between CPUR and CDWR in terms of general crop categories of non-permanent and semi-permanent crops using the kappa statistic (Lillesand and Kiefer; 1994). We calculated the kappa statistic for all sections in the county, whether or not pesticide application to one of our selected crops occurred.

The comparative analysis was performed at the Section extent (N = 3,856). We tested the resulting congruence estimates for statistical significance using a one-sided binomial test for two levels of assumed crop location error in the CDWR database (1% and 5% error).

The overall agreement between CPUR and CDWR crop location data in San Joaquin County in 1988 was relatively high, although location according to CPUR data was significantly different from that mapped by CDWR even when we assumed a 5% error in the CDWR data. Congruence for the four non-permanent crops analyzed, when aggregated, was 0.817 (CL = 0.800, 0.834) at the 95% confidence limit. Congruence for individual crops in this category ranged from 0.843 for alfalfa hay to 0.771 for tomatoes. Five semi-permanent crops were included in the analysis, and the aggregated congruence for these crops was 0.807 (CL = 0.785, 0.828) at the 95% confidence limit. Congruence for individual crops in this category ranged from 0.857 for grapes to 0.731 for cherries.

For Kings County in 1991, congruence for five non-permanent crops analyzed, when aggregated was 0.929 (CL = 0.916, 0.940) at the 95% confidence limit. Congruence for individual crops in this category ranged from 0.951 for cotton to 0.795 for tomatoes. The difference between location of all of our study crops according to CPUR versus CDWR was statistically significant at the 95% confidence limit when we assumed a uniformly distributed 5% error in the CDWR database, with the exception of corn, cotton, and walnuts. The difference for those crops was significant at the 95% confidence limit only when we assumed a 1% error in the CDWR database. Five semi-permanent crops were included in the analysis of the Kings County data for 1991, and the aggregate congruence for these crops was 0.874 (CL = 0.831, 0.909) at the 95% confidence limit. Congruence for individual crops in this category ranged from 0.947 for walnuts to 0.778 for almonds.

In general, the overall agreement between CPUR and CDWR crop location was significantly higher in Kings County for 1991 than for San Joaquin County in 1988 for both non-permanent (congruence = 0.929 vs 0.817, respectively) and semi-permanent crop categories (congruence = 0.874 vs 0.807, respectively). This trend holds when analyzing specific crops in both counties. Non-permanent crops that overlapped included alfalfa hay, corn, and tomatoes, and the average congruence for these was 0.882 for Kings County in 1991 versus 0.809 for San Joaquin County in 1988. Individual congruence for these crops was higher for Kings County as well, although that for tomatoes was not significantly higher. Semi-permanent crops that overlapped included almonds, grapes, and walnuts, and the average congruence for these crops was 0.867 for Kings County in 1991 versus 0.821 for San Joaquin County in 1988. Agreement between the databases for these crops individually was higher for Kings County as well, but was only statistically significantly higher in the case of walnuts.

Using the Kappa statistic (Lillesand and Kiefer, 1994), we determined that the degree of agreement between CPUR and CDWR trended towards actual as opposed to chance agreement for non-permanent crops, with Kappa = 0.78 for San Joaquin County and 0.75 for Kings County. A less strong trend in this direction was determined for semi-permanent crops, with Kappa = 0.63 for both counties. Kappa values can range from 0.0 to 1.0, with 1.0 indicating increased actual agreement.

1.3 Spatial Accuracy of NLCD data. We also tested the classification accuracy for the NLCD database using CDWR as the "gold standard". We conducted this study for years when CDWR data was collected (1991 and 1996 in Kings County, and 1988 and 1996 in San Joaquin County). The average and range of accuracy for the four years were: orchard and vineyard (68.0, 55.8 to 76.0 %), grain and hay crops (20.8, 6.0 to 42.2 %), and row and field crops (65.4, 43.7 to 85.7%). Based on these results we have concluded that NLCD is suitable for identifying broad agricultural crop categories, except grain and hay crops, and is feasible for use in epidemiological investigations. However, a CPUR-based exposure metric using CDWR data would be more accurate.

1.4 Temporal Accuracy of Spatial Crop Map Data. Epidemiological studies of cancer require exposure estimates over the period of latency between exposure and clinical onset of the disease. In the case of childhood cancer, this exposure period is generally considered to be the period between conception and diagnosis. Land cover data with spatial detail sufficient to calculate these metrics are not available for the entire State of California in every year. Therefore, in order to develop a CPUR-based exposure metric integrated with crop location data; there is a need to determine if significant change (temporal variation) in acreage occurred within land cover categories for years without crop data, such as CDWR data. We analyzed temporal variation in CDWR data for two periods between dates of availability in Kings County (1991 and 1996) and San Joaquin County (1988 and 1996). We determined variation for both specific crop classification as well as more broad classification categories such as "orchards". The results of the analysis found that acreage of cultivated land for annual crops changed very little (5-10%) in the two counties when classified in broad thematic categories such as "row and field crops" or "grain and hay crops". However, there was a large degree of change within a category when examining specific crop types. Acreage of more "permanent" crops (orchard, vineyard) changed from 5-20% in the two counties. Based on these analyses, we conclude that change in land cover between the years of available data was not significant when broad categories were used. Coupled with the results of the accuracy assessment, we conclude that location of these categories is fairly consistent over the period of our study (1988-1996). As a result, the location of broad agricultural land cover categories can be used to define sub-sectional crop location for each year of the study period, even for years when CDWR data is not available for validation.

1.5 Conclusions.

- It is feasible to link spatial crop map data to the California Pesticide Use Reporting Database.
- Spatial differences in accuracy of the CPUR dataset in predicting location of crops exist at both the aggregate and individual crop level.
- Land use data with relatively crude crop classification, such as those in the National Land Cover Database were less accurate in our study than CPUR data, when predicting crop location by Section.
- The "gold standard" for crop location used by our study, CDWR crop maps, while 100% ground truth, are surveys taken at one point in time in the annual cycle, and thus may miss crops with short growing periods. This may account for some of the error attributed to CPUR reporting in our study.
- Contrary to expected results, temporal variation in location of crops across a multiyear period was less for row crops (5-10%) than for more permanent crops (10-20%). Further research is needed to determine if this was a local anomaly in our study counties.
- The extent to which the temporal and spatial inaccuracy found in our study would effect location of pesticide use, and thus exposure classification in epidemiological investigations, needs further investigation.

Specific Aim 2. Develop and evaluate an exposure metric based on linkage of CPUR data with a spatial crop map database

2.1 Development of Metrics. We developed two exposure metrics for use in the epidemiological studies of agricultural chemicals. These metrics calculate potential residential exposure from application of agricultural chemicals in proximity to the target residence, and can be pesticide specific. The metrics we developed are (a) a subsection level metric using CDWR and CPUR, and (b) a subsection level calculation using a land cover database (LCD) such as NLCD or aggregated CDWR data and CPUR. A schematic diagram demonstrating the application of the metrics is presented in Figure 3.

The CDWR- and the NLCD-based exposure metrics we developed (Figure 3b and 3c) were based on a method reported by Ward et al (2000). That method used satellite imagery data and GIS technology to identify specific crops grown proximate to user-specified residences in Nebraska, and pesticide use (type, rate) on the crops was estimated from statewide survey data. We modified this method for use in California agriculture, where specific crop classification using satellite imagery is difficult to achieve, but pesticide use data is much more specific. Similar to the method we employed for accuracy assessment of CPUR, we used GIS technology to merge specific crop location data (CDWR) or broader crop classification categories (NLCD, Table 1) at the sub-Section level with specific pesticide use for the identified crops as reported in CPUR for the Section of residence. We calculated the metric within a 500 meter buffer around each residence using the following equation:

$$EM_k = \sum_{\substack{i=1 \\ j=1}}^{n} \left(\left(\frac{A_{ij}}{T_{ij}} \right) X_{ij} \right)$$

Where:
EM_k = exposure metric for a user-specified pesticide and residence, lbs
n = the number of sections within 500 meters of the residence.
k = pesticide type.
A_{ij} = the acreage of crop type i within section j, and within 500 meters of the residence.
T_{ij} = the total acreage of crop type i within section j.
X_{ij} = the total pounds of chemical k applied to crop type i within section j.

The CPUR-based metric we used in the comparative analysis (Figure 3c) is similar to the one used by Reynolds et al. (2002) and Gunier et al. (2001) in that it assumes uniform application of the target pesticide over the sections intersected by a buffer surrounding the residence. We used the following equation to quantify this exposure metric for the same 500 meter buffer used for the CDWR metric:

$$EM_k = \sum_{i=1}^{n} \left(\left(\frac{A_i}{T_i} \right) X_i \right)$$

Where:
n = the number of sections within 500 meters of the residence.
k = pesticide type.
A_i = the acreage of section i within 500 meters of the residence.
T_i = the total acreage of section i.
X_i = total pounds of chemical applied within section *i*.

2.2. Metric Evaluation.

2.2.1. CDWR-based metric. We applied the CDWR-based exposure metric (Figure 3b.) to a metric to one based solely on CPUR (Figure 3a.) We applied both metrics to 66 (11%) of a possible 582 residences in our six Central Valley county study area that were cases and/or controls in the CDHS database. The 66 residences were located in counties that had coincident CPUR and CDWR data in a year between 1988 and 1994, the birth years of the epidemiological study participants. These counties were San Joaquin (1988), Kings (1991), and Fresno (1994).

We calculated both exposure metrics for all residences for six (6) pesticides with high use in the study area: the herbicides Simazine and Trifluralin; the insecticides Dicofol and Propargite; and the fumigants Methyl Bromide and Metam Sodium (Gunier, 2001). We first compared the results of applying the metric for each pesticide across all sixty-six residences using a chi-squared test of a binary dataset (exposed / not exposed). In order to assess the effect of refining the location of pesticide use reported in CPUR using CDWR crop map, we converted the metrics to a common unit, pounds pesticide per square mile of buffer. We then used the Wilcoxon signed-rank test for significance (Weiss, 1997) to examine the differences of the predicted pounds of pesticide per square mile of buffer between the two metrics.

The results of the exposure classification analysis are presented in Table 3. We used the same definition for "exposure" used by Ward et al. (2000); the chemical was determined by the metric to have been applied within 500 meters of the participants' residences. When both metrics were in agreement, they classified more residences as exposed to herbicides (64% of all residences) than insecticides (17%) and fumigants (5%). The opposite trend resulted when the metrics agreed on classification as "not exposed" (17%, 61%, and 74%, respectively). Overall, four of the six pesticides, Simazine, Trifluralin, Dicofol, and Methyl Bromide, indicated similar categorical assignment of exposure for each metric., Cohen's kappa values for these pesticides ranged from 0.11 to 0.66. Although, overall agreement was relatively high for the herbicide Propargite (78.8%), it was found not to be statistically significant at the 95% confidence limit. There was very high agreement between the two metrics for the fumigant Metam Sodium (93.9%), but the data was so skewed toward no exposure (62 of the 66 residences) that the test for statistical association could not be applied.

The results of the analysis to determine if significant differences occurred in the predicted pounds of pesticide per square mile of buffer by each metric are presented in Table 4. The Wilcoxon signed-rank test indicated significant differences in estimated lbs/mile^2 applied within the residential buffers of 500 meters between the CDWR metric and the CPUR metric for 3 of the pesticides analyzed. The medians and interquartile ranges in the metrics were: for propargite, CDWR = 0.04 (0.00-0.57) and CPUR = 89.41 (14.47-233.69); for simazine, CDWR = 0.00 (0.00-0.03) and CPUR = 8.77 (0.00-50.13; for methyl bromide, CDWR = 0.00 (0.00-0.00) and CPUR = 0.00 (0.00-342.27).

The medians and interquartile ranges in the metrics for those residences classified as "exposed" by both metrics are presented in Table 5. Sample size was reduced significantly using this constraint, with the exception of propargite and simazine (N = 50 and 34, respectively). The Wilcoxon signed-rank test indicated significant differences in estimated lbs/mile^2 applied within the residential buffers of 500 meters between the CDWR metric and the CPUR metric for these two pesticides at the 95% confidence level. The ranges in exposure predicted by the two metrics for these pesticides are presented in Figure 4.

2.2.2. NLCD-based metric. We applied the NLCD-based exposure metric (Figure 3c.) to a metric to one based solely on CPUR (Figure 3a.). We first compared the categorical assignments of exposure using the NLCD versus the PUR metric, and then the specific pesticide exposures assigned by each metric to the birth year only study population generated from the CDHS database (N = 582). The results of the comparative analysis of classification are presented in Table 6. The results of a correlation analysis between the exposure predicted by each of the two metrics is presented in Table 7. Using only the PUR data produced the most exposed subjects for each pesticide, on the order of 2-3.5 fold increase over the NLCD metric. The distributions of pesticide use density (LBS/MI^2) among the exposed subjects were generally the same order of magnitude for each pesticide and method. Treated as continuous variables, PUR exposure values had pretty consistent rank correlation with NLCD values (R=0.52-0.62), indicated the values calculated by the two different methods are moderately associated.

2.3. Feasibility of integrating the CPUR-CDWR metric with a more advanced computer models for pesticide transport in the environment . We applied a pesticide drift model developed by USEPA (cite) and modified by Miller et al. (1998) to a subset of 10 residences in the CDHS case-control study. We identified one weather station located in the city of Fresno with daily averages of wind direction for 1994 coinciding with CDWR land use data in that county. We randomly selected 10 residents for analysis, which were located within 10 miles of the weather station; assuming wind direction data collected at the weather station was representative of this small geographic area due to the relatively flat topography. Crop fields located within 1,300 meters of each resident were identified in the CDWR database. Pesticide applications within each section were identified and linked to each crop field using crop type and field size. The resulting database contained pesticide use information for each field identified to have an application, the information in the database included: the date of application, chemical applied, mass (lbs) of chemical applied, acres treated, and application type (ground, aerial, or orchard blast). The linkage of the CDWR dataset with the CPUR dataset resulted in the identification of two pesticides applied in proximity to the 10 residents, simazine and propargite.

Pesticide drift was modeled for both pesticides from each field identified to have an application. Pesticide drift was estimated using the amount of chemical applied to each field as inputs into the AgDrift model (CITE) assuming a wind speed of 10 mph. The AgDrift model is an empirical model, which predicts deposition amounts for off target pesticide drift. The results of the AgDrift model were distributed down wind of the field using the average wind direction for the day of application. This resulted in a GIS coverage identifying the areas potentially affected by off target drift and the mass of pesticide potentially deposited at the location. The total mass (lbs) of pesticide deposited within 500 meters of the residence resulting from all applications was calculated. We compared the results of the drift modeling with the CDWR and NLCD based metrics. The results of the drift modeling procedure and the comparison with CDWR and NLCD based metrics are presented in Table 8. The drift modeling procedure classified 3 residents as exposed to simazine with a mean exposure of 1,279.8 lbs. Propargite had similar results with 2 residents exposed and a mean exposure of 583.3 lbs. We compared these results with the exposure estimates for the metrics using CDWR and NLCD as land-use data. Estimates of exposure were substantially higher when potential pesticide drift was included in the estimates.

2.4. Transferability across the State of California. Only the CPUR and NLCD data have statewide coverage. Use of NLCD might be limited by the fact that the data is derived from limited satellite imagery data (1992-94 composite for California). CDWR data is in GIS format, but are presently only available for 38 counties in California for intermittent years between 1976 and 2000. Generally, an updated map for agricultural counties is published every 5-7 years.

2.5. Conclusions.

- We have demonstrated that a viable exposure metric can be developed by integrating spatial crop map data and the California Pesticide Use Reporting database.
- When compared to a metric based solely on PUR, we found similar categorical assignment of exposure for the CDWR-based metric, but significant variation in the magnitude of exposure for 3 of the pesticides analyzed: propargite, simazine, and methyl bromide. The variance in exposure captured by the CDWR-CPUR metric was much higher than for the CPUR metric alone.
- When compared to a metric based solely on PUR, we found the NLCD metric classified significantly less participants as "exposed", and that the degree of association between the magnitude of exposure predicted by the two variables was moderately associated.
- Use of the California Department of Water Resources (CDWR) and the NLCD crop map database can be used to improve the crop location attributes of the CPUR database, the effect of this improvement on estimating exposure to pesticide needs to be further analyzed.
- The pesticide exposure metrics developed by this study require field validation, but show promise in predicting potential pesticide exposure. If validated, they may reduce exposure misclassification for subjects with high or low exposure.

Specific Aim 3. Pilot Epidemiology Study

3.1 Feasibility – Mapping study population / statistical power. As part of the epidemiology study in the RO3 research, we demonstrated the ability to locate specific residences in a case-control study. We obtained registry data for 308 cases of cancer for children less than 5 years old that were diagnosed between 1988 and 1997, and 607 controls from the vital statistics registry for the six county study region. We were able to map the location of residence at time of birth for 852 subjects (93%) in a GIS. Controls were randomly selected from state birth certificates, and matched to cases based on sex and date of birth. Co-variables that we have collected for the study include neighborhood SES (block group median income), mother's age, and infant's birth weight and race. We have conducted a power calculation for a study of all cancers and for a study of leukemia (N=129). Assuming that all cases can be used in the study, power to detect an odds ratio of 2.0 for all cancer ranges from 0.52 (prevalence of exposure (PE) = 2%) to 1.00 (PE = 25%). Using the same assumption, power to detect an odds ratio of 2.0 for leukemia ranges from 0.32 to 0.91.

3.2 Feasibility – Application of the exposure metrics developed by this study in an epidemiological study.

3.2.1. CDWR-metric. We evaluated the feasibility of applying the exposure metrics developed by this project to this study population. A major impedance was identified in that, because the CDHS data was registry based, the residence mapped for the cases was the residence at time of diagnosis, while for the controls the residence mapped was the maternal residence at time of birth. Epidemiological studies of cancer require exposure estimates over the period of latency between exposure and clinical onset of the disease. In the case of childhood cancer, this latency period can be defined as date of conception until date of diagnosis. Therefore, there is a need to develop an exposure metric for intervening years between conception and diagnosis. This was not possible with the dataset that we had to work with.

In an attempt to adjust for this issue, we attempted to restrict our study, for demonstration purpose, to exposure during birth year only. This resulted in a subset of 582 (213 case, 422 controls) subjects for which we were able to match each subject's maternal address at birth, which was obtained from birth certificates for a study period between 1988 and 1994. We conducted a revised power calculation for a study of all cancers and for a study of leukemia (N=75 cases). Assuming that all cases can be used in the study, this procedure resulted in a power to detect an odds ratio of 2.0 for all cancer to ranges from 0.42 (prevalence of exposure (PE) = 2%) to 0.99 (PE = 25%). Using the same assumption, power to detect an odds ratio of 2.0 for leukemia ranges from 0.31 to 0.83. However, once we matched the residence at time and year of birth with years and counties for which available CDWR crop maps were available, the number of subjects was reduced to nine (9), which would not provide acceptable power for the epidemiological analysis, so this specific aim was not achievable using the CDWR-based metric.

3.2.2. NLCD metric. The NLCD data for California is based on satellite imagery, primarily dated 1992. Because the spatial accuracy for NLCD was low compared to that of CDWR (Specific Aim 1), and the temporal period of our epidemiological dataset so wide (1988 – 1997), we concluded that an epidemiological analysis using the NLCD would add little value to one conducted using the PUR alone without further validation that the temporal variation in broad crop classification categories was sufficiently low. One potential solution would be to use CDWR aggregated to the broad crop classification categories (Table 1) for locations where two "bookend" years of crop location data was available, and interpolated location of agricultural land use (row crops, orchards, vineyards, etc.) for the intervening years (Figure 5). Funding for development and validation of such an approach was beyond the scope of this study.

Under separate funding, a separate statewide epidemiological analysis of childhood leukemia and pesticide exposure based on a PUR only metric was conducted, and reported by Reynolds et. al. (2002). In that study, they reported elevated rates of childhood leukemia (RR = 1.48, 95% CL, 1.03 – 2.13) in census block groups composed of Sections with the highest use of propargite according to the CPUR database. The CPUR exposure metric was calculated based on the location of residence at time of diagnosis. Age of the population was < 15 years.

3.3. Conclusions

- There is a need to validate metrics used to express exposure to agricultural pesticides in a non-occupational epidemiological study. An important aspect of this validation should be to determine how well pesticide levels in current environmental or biological samples reflect historical agricultural pesticide use in proximity to the residence.
- Epidemiological studies of cancer require exposure estimates over the period of latency between exposure and clinical onset of the disease. In the case of childhood cancer, this latency period can be defined as date of conception until date of diagnosis. Environmental samples, even if sufficient as biomarkers of exposure, can only be collected after date of diagnosis. Therefore, there is a need to develop an exposure metric for intervening years between conception and diagnosis.
- If spatial crop map data are going to be combined with CPUR data in such an exposure metric, a method must be developed and validated for estimating location of crops for periods in the exposure period when crop maps are not available.
- There is a need to determine how potentially more spatially refined metrics than CPUR would effect risks such as those found by Bell et al. (2001) and Reynolds et al. (2002).

Cited Literature

Bell EM, Hertz-Picciotto I, Beaumont J. A case-control study of pesticides and fetal death due to congenital anomalies. Epidemiology (CDPR) California Department of Pesticide Regulation, Information Systems Branch, California Environmental Protection Agency. Pesticide Use Reporting: An Overview of California's Unique Full Reporting System, 1995.

California Department of Pesticide Regulation. Pesticide Use Report System Data, 1988-1994 (Data File). Sacramento, CA: California Department of Pesticide Regulation, 1996.

Fenske RA, Kissel JC, Chensheng L, kalman DA, Simcox NJ, Allen EH and Keifer MC. Biologically based pesticide dose estimates for children in and agricultural community. Environ Helath Perspect 108(6): 515-520; 2000.

Gunier RB, Harnly ME, Reynolds P, Hertz A, Von Behren J. Agricultural pesticide use in California: pesticide prioritization, use densities, and population distributions for a childhood cancer study. Env. Health Persp. 109 (10) 1071-1078. 2001.

Johnson N, Kotz S. Discrete Distributions. Houghton Mifflin, now distributed by New York: John Wiley. 1969.

Keifer MC (editor). Human health effects of pesticides. Occupational Medicine:State of the Art Reviews Vol 12(2). Hanley and Belfus. Philadelphia. 1997.

Lillesand TM and RW Kiefer. Remote Sensing and Image Interpretation. New York: John Wiley and Sons, Inc. 1994.

Miller, RS; JR Nuckols; MH Ward; SK Maxwell; SJ Weigel. Estimating Individual Residences Exposure to Agricultural Chemicals. Proceedings Geographic Information Systems in Public Health 3rd National Conference, August 17-20, 1998, San Diego, CA.

Reynolds P, Von Behren J, Gunier RB, Goldberg DE, Hertz A, Harnly ME. Childhood cancer and agricultural pesticide use: an ecologic study in California. Env. Health Persp. 110 (3) 319-324. 2002.

Simcox NJ, Fenske RA, Wolz SA, Lee I, Kalman DA. Pesticides in household dust and soil: exposure pathways for children of agricultural families. Environ Health Perspect 103:1126-1134; 1995.

Teske, ME; SL Bird; DM Esterly; SL Ray; SG Perry. A Users's Guide for AgDRIFT™ 1.0: Tiered Approach for the Assessment of Spray Drift of Pesticides. Spray Drift Task Force. Stewart Agricultural Research Services, Inc. Macon Missouri.

Ward MH, Nuckols JR, Weigel SJ, Maxwell SK, Cantor KP, Miller RS. Estimating environmental exposure to agricultural pesticides using remote sensing and a geographic information system. Environ Health Perspect 108(1):5-12; 2000 .

Weiss, NA. Introductory Statistics. New York: Addison-Wesley Publishing Company, Inc. 1997.

Zahm SH, Ward MH, Blair A. Pesticides and Cancer. Occupational Medicine: State of the Art Reviews 12(2):269-289, 1997.

Publications and Related Materials

Published:

Nuckols JR. Use of a Geographic Information System (GIS) for Assessment of Exposure to Agricultural Chemicals in the State of California, USA. 13th Conference of the International Society for Environmental Epidemiology (ISEE). September 2-5, 2001, Garmisch-Partenkirchen, Germany. (abstract) Epidemiology. July, 2002.

Submitted:

Currently None.

In Preparation:

Nuckols JR, Miller RS, Gunier RB, Weigel SW, Reynolds P, and Ward M. Linkage of the California pesticide use reporting database with spatial land use data for exposure assessment and environmental epidemiology.

Proposals:

R01 CA92683-01. PI:John R. Nuckols. Geographic-based Exposure Assessment of Agricultural Chemicals and Childhood Cancer. Sponsor Name: HHS-NIH-Center for Scientific Review. Project Start: 9/27/2001. Project End: 8/31/2005. $1,031,673.

List of Tables

Table 1. Classification of CDWR, CPUR, and NLCD crop data into common land use categories.

General Category	Sub-Category
Crop Land	
	Agricultural Land
	Grain and Hay Crops
	Row and Field Crops
Natural Vegetation	
	Mixed Barren Land
	Pasture
	Native Vegetation
Riparian Areas	
	Riparian Vegetation
	Surface Water
Orchard and Vineyard	
Urban	

Table 2. Crops with the greatest number of pesticide applications: San Joaquin (1988) and Kings (1991) Counties.

Crop Type	*Number of Sections with Applications*	*Number of Applications*	*Percentage of Total Applications*	*Total Pounds of Pesticide Applied*	*Percentage of Total Pounds of Pesticide Applied*
Kings County, 1991					
Cotton	704	26,960	51.9	1,708,767.1	42.0
Peach[1]	78	5,257	10.1	274,745.4	6.8
Alfalfa	239	4,358	8.4	221,755.4	5.5
Nectarine[1]	60	3,162	6.1	136,620.3	3.4
Plum	70	2,359	4.5	78,924.0	1.9
Grape	60	1,784	3.4	182,519.4	4.5
Walnut	81	1,571	3.0	52,202.6	1.3
Safflower	96	1,026	2.0	67,595.7	1.7
Tomato	44	771	1.5	214,855.3	5.3
Corn	111	763	1.5	30,536.1	0.8
Apple	16	568	1.1	14,914.4	0.4
Almond	28	533	1.0	157,577.6	3.9
Total		48,011	92.4	3,141,013.3	77.2
Adjusted Total[1]		40,693	78.3	2,729,647.6	67.1
San Joaquin County, 1988					
Grape	239	3,043	15.1	782,290.7	13.3
Almond	200	2,760	13.7	445,124.9	7.6
Tomato	308	2,518	12.5	1,081,786.1	18.4
Alfalfa	318	2,006	9.9	115,898.2	2.0
Sugar Beets	262	1,256	6.2	2,510,762.6	42.7
Walnut	177	919	4.5	76,586.7	1.3
Corn	155	853	4.2	80,264.9	1.4
Cherry	99	750	3.7	67,655.3	1.2
Asparagus	138	608	3.0	83,004.7	1.4
Wheat	139	578	2.9	45,560.7	0.8
Total		15,291	75.7	5,288,934.8	90.0

[1] California Department of Water Resources database does not distinguish between peach and nectarine crops, making it impossible to calculate estimates of reporting error. Those crops are included here as reference information only. The adjusted total does not include peach and nectarine.

Table 3. Results of Chi-square analysis to determine differences between CDWR and CPUR metrics in classifying (potential) exposure

Chemical	Type	Overall Agreement (%)	Chi –Square value	P-Value (95% Confidence Limit)
Propargite	Herbicide	78.8	0.80	0.371
Simazine	Herbicide	83.3	29.05	0.000
Trifluralin	Insecticide	75.8	8.85	0.003
Dicofol	Insecticide	80.3	19.55	0.000
Meth. Bromide	Fumigant	63.6	7.92	0.005
Metam Sodium[1]	Fumigant	93.9	-	-

[1] Small sample size did not allow chi-square analysis.

Table 4. Comparison of exposure metrics using CPUR with and without integration of CDWR crop maps.

Statistic	Herbicides				Insecticides				Fumigants			
	Trifluralin		Simazine		Propargite		Dicofol		Methyl Bromide		Metam Sodium	
	CDWR	PUR	CDWR	PUR	CDWR	PUR	CDWR	PUR	CDWR	PUR	CDWR	PUR
N	66	66	66	66	66	66	66	66	66	66	66	66
Mean *	21.67	16.94	0.97	49.87	29.77	168.72	48.52	18.58	8.57	617.19	0.00	26.40
StDEV	75.90	48.67	7.17	109.46	93.29	215.90	138.79	51.81	44.59	1513.06	0.00	152.65
Variance	5,760	2,368	51	11,982	8,704	46,614	19,262	2,684	1,988	2,289,341	0	23,302
Median	0.000	0.000	0.002	8.766	0.035	89.413	0.000	0.000	0.000	0.000	0.000	0.000
1st Quartile	0.000	0.000	0.000	0.000	0.004	14.469	0.000	0.000	0.000	0.000	0.000	0.000
3rd Quartile	0.001	0.386	0.034	50.129	0.573	233.693	1.000	7.872	0.000	342.270	0.000	0.000
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	365.94	270.10	58.23	585.93	567.10	888.84	671.96	323.95	314.40	8780.48	0.03	1053.55
Wilcoxon Signed Rank Test												
Z	0.4846		-5.602		-5.3917		-0.5784		-4.8281		-1.0231	
P-Value	0.628		0.000		0.000		0.563		0.000		0.306	

* All units = pounds of pesticide per square mile within a 500 meter residential buffer

Table 5. Comparison of exposure metrics for residences classified as exposed by both metrics.

Statistic	Trifluralin		Simazine *		Propargite *		Dicofol		Methyl Bromide	
	CDWR	PUR	CDWR	PUR	CDWR	PUR	CDWR	PUR	CDWR	PUR
N	9	9	34	34	50	50	14	14	6	6
Mean **	141.04	48.84	0.18	88.56	39.29	189.29	178.40	74.08	94.32	735.71
StDEV	162.56	27.06	0.68	141.44	105.66	199.96	239.82	88.08	127.05	1,429.65
Variance	26,424.82	732.32	0.46	20,004.89	11,164.10	39,984.62	57,512.24	7,757.39	16,141.63	2,043,891.83
Median	71.57	43.16	0.02	43.75	0.10	128.15	22.81	47.13	28.43	177.42
1st Quartile	0.03	35.28	0.00	10.17	0.02	33.71	0.01	15.89	8.77	14.35
3rd Quartile	343.97	62.74	0.10	75.23	16.57	284.96	435.72	89.75	215.80	1213.00
Minimum	0.00	10.88	0.00	0.07	0.00	0.15	0.00	1.86	0.04	0.63
Maximum	365.94	102.85	3.98	585.93	567.10	888.84	671.96	323.95	314.40	3633.78
DF	8		33		74		16		5	
P-Value	0.30		0.00		0.00		0.54		0.16	
Z	n/a		-4.99		-4.78		n/a		n/a	

* Indicates significant at both the .05 and .1 level of significance.

** All units = pounds of pesticide per square mile within a 500 meter residential buffer

Table 6. Comparison of an NLCD metric to one based only on CPUR for 582 residences at year of birth (estimated pesticide application within a 500 meter buffer around the residence)

Pesticide	Exposed Subjects (N, % of Total Subjects)		Median Exposure (LBS/MI^2)		90th percentile Exposure (LBS/MI^2)		Maximum Exposure (LBS/MI^2)	
	NLCD	CPUR	NLCD	CPUR	NLCD	CPUR	NLCD	CPUR
Dicofol	28 (5%)	72 (12%)	43	27	180	142	375	324
Propargite	85 (15%)	193 (33%)	37	38	258	286	786	888
Trifluralin	27 (5%)	94 (16%)	33	26	298	146	334	452
Simazine	60 (10%)	145 (25%)	25	11	77	82	451	586
Methyl Bromide	49 (8%)	119 (20%)	209	170	4373	4478	9352	8780
Metam Sodium	4 (1%)	13 (2%)	760	250	1614	1054	1614	2801

Table 7. Spearman rank correlation coefficients (R) between PUR and NLCD metrics for 582 residences at birth.

Pesticide	**Spearman Correlation Coefficient (R)**
Dicofol	0.59
Propargite	0.60
Trifluralin	0.52
Simazine	0.62
Methyl Bromide	0.62
Metam Sodium	0.55

Table 8. Results of pesticide drift analysis and comparison with CDWR and NLCD based metrics.

Metric	**Simazine**		**Propargite**	
	Mean Exposure (lbs/mi^2)	Exposed Subjects (N, %)	Mean Exposure (lbs/mi^2)	Exposed Subjects (N, %)
Drift	1,279.769	3 (30%)	583.322	2 (20%)
CDWR Based	0.003	3 (30%)	0.003	2 (20%)
NLCD Based	18.794	3 (30%)	18.220	3 (30%)

List of Figures

Figure 1.

Created By:

Ryan S. Miller
Environmental Health Advanced Systems Laboratory
Fort Collins, Colorado

/data13/calif/image-files/projects/cali.apr

Figure 2. Example of linkage capability between CPUR and crop maps using a GIS.

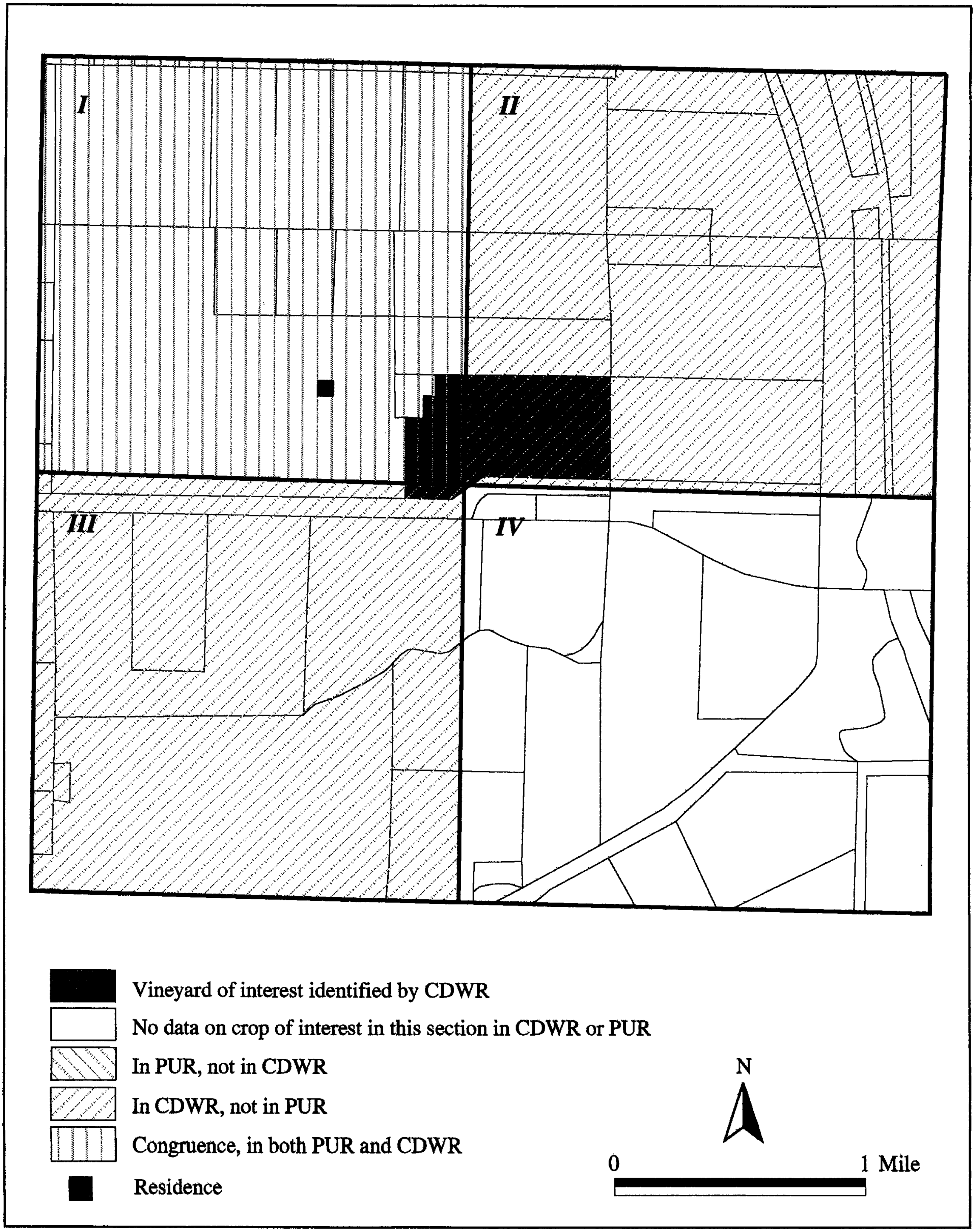


Figure 3. Metrics Evaluated in RO3 Project.

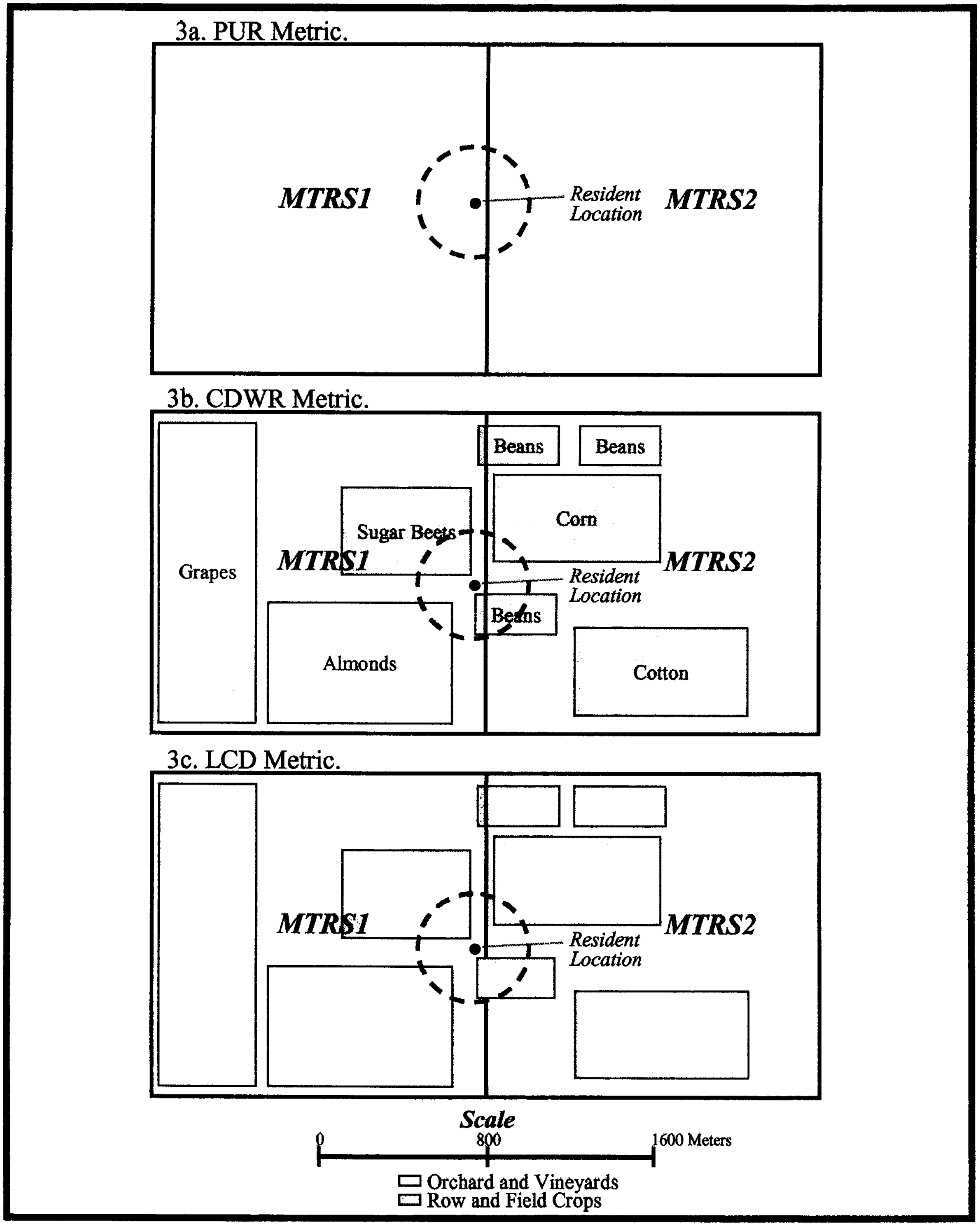


Figure 4. Range in exposure predicted by the two metrics for propargite and simazine.

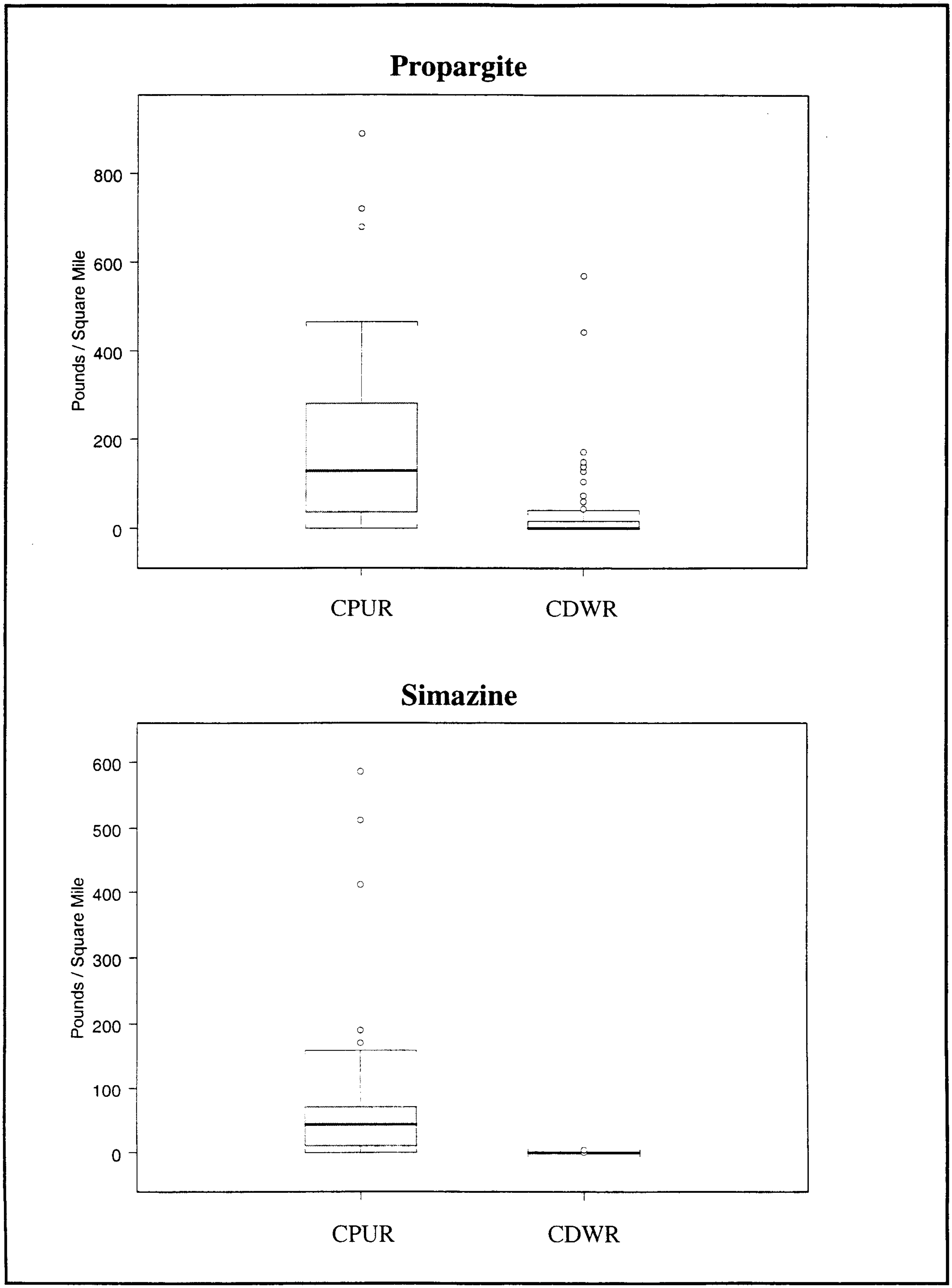
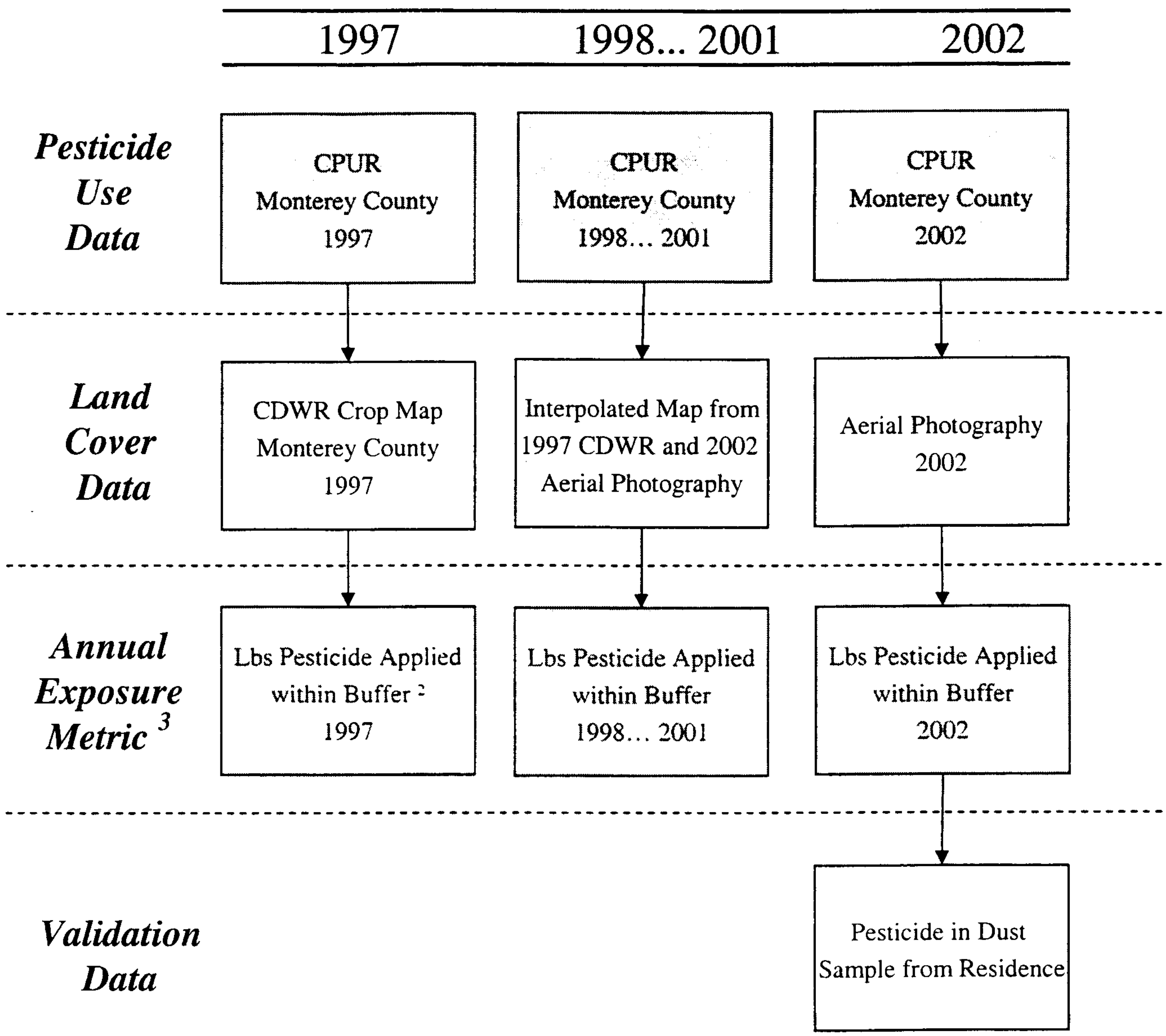


Figure 5. Exposure metric for person born in 1997 and diagnosed in 2002[1] in Monterey County.



[1] Resident residing at the same residence from date of birth until date of diagnosis.

[2] Buffer is defined as the source area for potential pesticide transport to the residence based on transport variables. Generally, primary pesticide drift distance ranges from 500 to 800 meters at wind speed of 10 mph.

[3] A cumulative exposure metric will be calculated as the $\sum EM$ for all years.