Yale
Poorvu Center

Statistical Methods Section

Kelsie Cassell

Deziel (Environmental Health, 2015)

Statistical Methods

Statistical analysis

Using multivariable linear Tobit regression models for each pesticide, we determined the relationship between pesticide dust concentrations (dependent variables) and ever/never self-reported pest treatments in the past 12 months (independent variables). Tobit regression is an unbiased approach for analyzing measurement data when a substantial proportion of samples are below the limit of detection [19]. We modelled the natural log-transformed pesticide concentrations and estimated the relative change in pesticide concentrations by exponentiating the regression coefficients. We summed concentrations of pesticide isomers (cypermethrin I, II, III, IV, cyfluthrin I, II, III, IV, and cis- and trans-permethrin) be-cause they were highly correlated (rspearman > 0.9).

Define statistical method used, with outcome and predictor/exposure clearly labeled in a manner that would be reproducible.

Explain why this particular statistical method is chosen.

Clarify any data transformations.

Each insecticide-specific model included the following self-reported pest treatment variables: ants/cockroaches, carpenter ants/termites, fleas/ticks in home, fleas/ticks on pets, flying insects, lawn and garden insects, professional inside treatments, and professional outdoor insect treatments. These pest treatments were moderately correlated (Cramer's V ranged from 0.15 to 0.77; median = 0.42; Additional file 1: Table S1). We excluded flea/tick shampoo because it was highly correlated with other flea/tick treatment on pets (Cramer's V > 0.9). We excluded fogger/bomb and indoor plant treatments due to low prevalence (≤5%) and flea/tick collars because we did not measure relevant active ingredients. We combined bees/wasps/hornets and flies/mosquitoes into a "flying insects" category since we lacked data to discriminate active ingredients for these two groups [21]. The models for herbicides included weed treatments by a household member and by a professional as independent variables. A subset of participants who were not asked about professional treatments an early version of adjustment for a broad range of potential confounders of the relationship between self-reported use and pesticide concentrations in dust including the following demographic and household characteristics: child's age at diagnosis/reference,

List all covariables of the model.

Explain any exclusions from the model.

Explain any subsets of the model.

Explain any adjustments that were made, and define the variables in common terms.

child's sex, child's race/ethnicity, household income, sampling year, sampling season, when the sampled home was built, whether family members typically removed their shoes upon entering the home, mother's educational level, number of children residing in the home, residence type (single family home or other), whether a pet lived in the home in the first 2 years of the child's life (potentially increasing track-in of out-door pesticide applications [7]), dust sampling method (vacuum or HVS3), frequency of vacuuming, time be-tween diagnosis/reference date and dust collection date, and urbanicity of residential census block (based on population density) [22]. Because all 13 pesticides/synergists were used in both residential and agricultural products, we also considered the effect of nearby agricultural use. As described previously [23,24], agricultural use near the home was determined as the density (mass/unit area) of pesticides applied within a 1250-m buffer around the residence over the 12 months prior to dust collection based on the California Pesticide Use Reporting Database. We included all pest treatment, demographic, household, and agricultural density variables in our initial models. Retaining all pest treatment variables, we removed sequentially the demographic/household characteristics or agricultural density variable with the highest p-value until all remaining covariates had p-values <0.1.

Example of using common terminology.

How model fit was assessed and how best model was determined.

To assess whether the relationship between self-reported pest treatments and concentrations in dust differed by case–control status, we constructed our models separately for cases and controls as well as combined (including case–control status in combined models if $p < 0.1$). We also tested for interactions between pest treatments and case–control status.

Explain any effect modification or interactions that were assessed.

To assess whether the observed associations between treatment for a particular pest and pesticide concentrations in dust were consistent with the composition of commercial pesticide products used by the general pub-lic to treat that pest during the time frame of our study, we used information for the year 2000 from the NCI Pesticide Exposure Matrix (http://dceg.cancer.gov/tools/design/pesticide) [21]. This publicly available tool uses national data on product sales, active ingredient sales, and pounds of active ingredient from market planning reports to predict the probability that an active ingredient was used for each of 96 scenarios (12 pest types, whether the applicator was a household member ["consumer"]or professional, and 4 timeframes [1976, 1980, 1990, 2000])[21]. We categorized the probabilities as 0% (active ingredient not listed), 1-9%, 10-19%, and ≥20%.

If external data are linked, say where it comes from and how it is incorporated.