
Power Analysis in Association between Source-Specific Swine Markers and Acute Changes in Health Status Measures

This study is based on the Community Health Effects of Industrial Hog Operations (CHEIHO), a longitudinal health study of associations between daily airborne emissions and daily changes in respiratory symptoms, lung function etc, specifically, the impacts of exposure to hydrogen sulfide and PM from swine CAFOs on acute changes in odor perception and quality of life. It is of special interest then to identify source-specific PM exposure markers to swine CAFOs, and investigate their relations to public health. 12 communities near swine CAFOs in North Carolina participated in CHEIHO study. We need to find out how many sites were with detectable source-specific exposure markers of fecal pollution, decide the sampling plan based on the exposure prevalence, and then assess the exposure's impact. The main purpose here is to analyze the power of detecting source-specific markers' effects on the risk of physical symptoms (irritation and upper respiratory symptoms).

During CHEIHO, participant from each community recorded their physical symptoms twice-daily in personal diary for 14 days. The study periods vary among different communities. 12-hour PM filters were collected accordingly during the particular study period for that site. Those filters from CHEIHO having been analyzed can not be re-analyzed, and we are provided with 323 PM filters archived, together with 1497 diaries matched for analysis.

The first stage is an initial exposure assessment, screening PM filters pooled by week (14 twelve-hour filters per week) within each community to identify presence of source-specific fecal pollution markers. The following analysis will be adaptive to the results from initial screening. If no source-specific molecular markers are detected, we propose to perform exposure analysis including wider genera of fecal Methanogens on the remaining PM filters. If the source-specific swine markers are infrequently detected in the communities, we propose to conduct a cohort design using data only from those communities with source-specific exposure markers present. We expect being able to analyze up to 310 archived PM filter samples, and we can obtain the information about source-specific swine markers' presence/absence from PCR positive/negative analysis. Question raised here is

if there is an increased risk of symptoms(irritation, upper respiratory) during time periods when source-specific Methanogens are present compared to when absent.

If the source-specific molecular markers are detectable ubiquitously in the 12 communities, we propose to perform a nested case-control study. Again, we anticipate being able to analyze up to 310 archived PM filter samples, which means we can select 155 case time periods and 155 control time periods, each period of 12 hours, with and without physical symptoms. We would like to see if there is an increased odds of exposure to source-specific swine markers during case time periods compared to during control time periods. Questions of interest are how to conduct this nested case control study, given that we could not analyze as many filters as we want, and what would be the minimum detectable differences with at least 80% power and two-sided hypothesis tests with type I error rate 0.05. We wish to be able to detect 1.5-2 odds difference comparing period when source-specific swine markers are present with period when absent.

Instead of sampling cases and controls, we are actually sampling PM filters in our nested case control design. However, for one site, we can not always find a control time period in which no people reported symptoms or a case time period when all people reported experiencing symptoms. Therefore, we proposed to pick out case/control time periods based on the symptom frequencies for each site.

In this nested case control sampling setting, we perform power analysis to see if we can achieve desired power to detect an increase in odds of exposure, based on outcome of irritation symptoms. Some co-pollutants are of our concerns because they might interfere with presumed effects. We want to control for these pollutants, $PM_{2.5}$ and semi-volatile PM_{10} . We have the model:

$$\text{logit}(E(R_{ijk}|b_i, b_{ij})) = \beta_0 + \beta_1 \text{Marker}_{ik} + \beta_2 PM_{2.5ik} + \beta_3 SVPM_{10ik} + b_i + b_{ij},$$

with i indexing communities, j indexing individuals in a site and k indexing time periods.

To check the power, we set the model parameters as follows. Let $\beta_0 = -2.19$. Consider the community with no source-specific swine markers, $PM_{2.5}$ and

$SVPM_{10}$, and $b_i = 0, b_{ij} = 0$, then the probability of people experience irritation symptoms are 0.1. β_2, β_3 would be set to 0.049 and 0.043 respectively, according to relevant literature. $\sigma_1^2 = 0.25$, on the lower end of the range would allow most estimated exposure prevalence for the community in the range (0.04,0.23), while $\sigma_2^2 = 1.5$, would allow most estimated probability for people experiencing irritation symptoms in the range of (0.01,0.55) given $b_i = 0$.

This study indicates the possibility of the detection of meaningful difference in odds using the proposed sampling design for nested case control study, even when exposure prevalence is not high. In case we have probability of 0.3 to detect source-specific marker's existence, then we have at least 80% power to see 1.58 odds difference in the response. If no other co-pollutants exist, and random effects for site and individuals are zero, then people in that community have probability of 0.1 for experiencing irritation symptoms when no source-specific swine markers exist, while this probability goes up to 0.15 when markers do exist. If the probability of detection is 0.5, then we have power to find even smaller difference given the sample size we have. However, we might have some problems detecting the difference we expected to see when exposure is too rare, for example, around 0.05.

We assume source-specific exposure markers to be independent of other co-pollutants, e.g, $PM_{2.5}$ and semi-volatile PM_{10} . Since the filter analysis has not been done yet, it is hard to think about correlation with the copollutants we are controlling for. We are measuring the source-specific swine marker in the PM mass. It is possible that when a greater mass of PM is measured nearby CAFOs there could be a greater chance to detect the swine-specific marker in the PM mass. However, it can also be the opposite. Analysis was done only for binary exposure indicator, the presence/absence of source-specific swine markers. We can also perform quantitative PCT to estimate the airborne concentration of source-specific Methanogens. In this case, we are able to analyze the effect of source-specific markers on public health based on a continuous scale. More specifically, we can assess the effect associated with a one unit increase in concentration of swine specific markers.