



Published in final edited form as:

Prog Community Health Partnersh. 2009 ; 3(2): 179–190. doi:10.1353/cpr.0.0065.

Data Completeness and Quality in a Community-Based and Participatory Epidemiologic Study

Leah Schinasi, Rachel Avery Horton, and Steve Wing

Department of Epidemiology, School of Public Health, University of North Carolina

Abstract

Background—The principles of community-based participatory (CBPR) research challenge traditional scientific standards of objectivity and neutrality. Little work has been done to evaluate the quality of data obtained from CBPR studies.

Objectives—We examined factors associated with the completeness and quality of data that participants collected for the Community Health Effects of Industrial Hog Operations (CHEIHO) study, a community-based, participatory, longitudinal, epidemiologic investigation.

Methods—Twice daily for 2 weeks, 101 eastern North Carolina residents collected data on odor from industrialized hog operations, physical health, and mood. Data collected at a single point in time constitute a record. For each record, participant responses were classified as error free or not and missing or not. We used mixed models to quantify associations between errors or missing values and time of day, odor rating, week-in-participation, and presence of a person to assist with data collection.

Results—Participants collected data out of order in 2% of 2,949 total records. On average, individual variables were incomplete in 2% of records. Errors and missing data were most common for lung function measurements. Missing data for lung function and blood pressure were less common after the first week of participation (odds ratio [OR], 0.41; 95% confidence interval [CI], 0.20–0.84). Saliva samples were more frequently missing when participants reported odor than when they did not (OR, 1.59; 95% CI, 0.97–2.59). For women, the odds that yes/no variables were missing in week 2 records were higher relative to week 1 (OR, 1.46; 95% CI, 1.01–2.12).

Conclusions—Community members collected relatively complete and consistent data. Better training in use of mechanical devices and more frequent input from researchers could help to improve data quality in CBPR studies.

Keywords

Epidemiology; quantitative methods; study design; environmental justice; validity

CBPR challenges traditional scientific principles and redefines the goals of investigation.¹ In CBPR, community members affected by an issue of concern actively participate in all phases of the research process.^{2–5} Principles of CBPR include valuing the resources that can be found within the community, complete collaboration between all partners throughout all phases of the research process, and addressing questions that are relevant to the local community.^{6–8}

Use of CBPR principles in epidemiology may conflict with the field's emphasis on neutrality, objectivity, and separation of research from political or social action.⁹ Therefore, it is important to examine the quality of data from epidemiologic studies that use CBPR.

Little work has been done to quantify and critically examine the quality of the data from community-based studies. We address this issue by assessing the completeness and quality of data that residents collected for a CBPR epidemiologic investigation. We also examine relationships between completeness and quality of data and several factors that, if modified, could help to improve data in future CBPR studies.

Materials and Methods

Study Design

CHEIHO was based in eastern North Carolina. The goals of CHEIHO were to evaluate community members' exposure to pollutants associated with swine confined animal feeding operations, their health-related outcomes, and relationships between exposures, health-related outcomes, and quality of life. Another study objective was to increase community members' understanding of research design.¹⁰ CHEIHO incorporated CBPR principles in several ways. The study addressed residents' concerns about the health effects of living near industrial hog operations. Community partners also helped to develop the study instruments and consent form, and hog operation neighbors collected data in their own homes.

Community organizers and researchers recruited participants in 16 neighborhoods. Any person was eligible to participate if they were 18 years of age or older, did not smoke, lived within 1.5 miles of at least one hog operation, had time to complete data collection activities, and had a freezer with sufficient room to store a box of tubes for saliva sample collection.

After completing an eligibility questionnaire, participants attended a 3-hour training session where they received a journal in which to record data, as well as the necessary equipment for taking blood pressures, collecting saliva samples, and measuring lung function. During the session, a staff person stood at the front of the room and, using laminated and enlarged journal pages that were propped on an easel, guided participants through the data collection activities. Participants completed sample journal pages and practiced using all data collection equipment. Two to four additional staff were present and assisted participants with difficulties that they encountered. Before or after the training, staff also administered a test to determine participants' odor sensitivity. Over time, CHEIHO staff adjusted training sessions in response to difficulties; however, there was never a formal evaluation of the sessions.

Participants chose morning and evening times, approximately 12 hours apart, to collect data. They collected data during these times for approximately 14 days. University researchers were available by telephone to answer questions and visited homes after approximately 1 week to offer feedback. At 10 sites, community-based representatives assisted participants. Representatives were volunteers from the community or informally identified community leaders who played important roles in recruiting neighbors to participate in CHEIHO. All representatives were particularly able study participants who received the same training as

other participants. Because representatives lived in the communities, they were able to assist participants in their homes, which was something that university-based staff could not do on a daily basis.

Each journal entry included four pages of activities and questions (Appendix 1). To follow the protocol, participants first sat outside for 10 minutes and completed the first page of the journal, indicating the strength of any hog odor that they smelled during each of the previous 12 hours by choosing a number between 0 and 8, where 0 = *no odor* and 8 = *strong odor*. Participants also noted daily activities that they did not do or did differently or with difficulty because of odor.

Next, participants returned indoors and recorded the current time. They rated the strength of odor that they smelled during their 10 minutes outside using the same 9-point scale. They responded to 5 questions about stress and negative mood (see Step 5 in the Appendix), took 2 time-stamped blood pressure readings, and rated the extent to which they experienced 22 different symptoms during the previous 12 hours by circling a number between 0 and 8 (0 = *not at all* and 8 = *extreme*; see Step 7 in the Appendix).

After indicating any out-of-the-ordinary medications that they had taken in the past 12 hours, participants collected saliva samples, recorded the time at which they did so, and then measured lung function by blowing three times into an AirWatch Asthma Monitor. The monitor maintained the forced expiratory volume (FEV1) and peak expiratory flow readings internally and flagged them for errors if there was coughing, jetting, and/or too short of an effort during the trial. Participants wrote their third FEV1 and peak expiratory flow readings in their journals.

The Institutional Review Board of the University of North Carolina at Chapel Hill approved CHEIHO and all study participants provided informed consent. Assurance of participant confidentiality was important because, in a previous study, the hog industry attempted to identify participants.¹¹ We obtained a Certificate of Confidentiality from the U.S. Department of Health and Human Services.

Dependent Variables

We created several variables to indicate the completeness and consistency of data in each record. A sequence error occurred if participants provided health data before they went outside for 10 minutes and exposed themselves to the outdoor air (Table 1). We coded a record to have a missing saliva sample if no sample was submitted or if a sample weighed less than the minimum weight necessary to assay immunoglobulin A (0.25 g). We classified a record as having an AirWatch error if there was no usable reading because all trials were flagged for errors. We defined a diary response as missing if the participant did not answer a question or if the answer was illegible or illogical. We calculated the proportion of records with complete information for each individual question. We did not evaluate completeness of responses to the question about odor strength in the preceding 12 hours (Appendix, page 1) because the options were too complex to summarize.

To improve statistical power and aid interpretation of results, we created summary variables for analyses of completeness. Summary variables were defined based on response coding (yes/no, ordered response) or activity type (use of a machine). Summary variables were coded 1 if any of the individual variables that met its definition were missing from a record and 0 otherwise. Summary variables' names and definitions are given in Table 1.

Independent Variables

Data consistency and completeness might have improved in the second week of participation owing to increased experience or feedback from staff during midstudy visits. Therefore, we created a week-in-participation variable, scored 0 if the record was from the first week and 1 otherwise. Because strong odors might have motivated participants to collect higher quality data, we classified odor during the 10 minutes outdoors as either absent or present. To examine whether presence of a community-based representative improved data collection, we created a variable coded 0 if a representative was present and 1 otherwise. Finally, we considered time of day at which data were collected (evening, 2:00 pm to 1:59 am vs. morning, 2:00 am to 1:59 pm), which might have affected the extent to which participants experienced distractions or were able focus efforts on data collection.

Statistical Analysis

One participant who had trouble following the study protocol was excluded from all analyses.^{10,12} We calculated the total number of records that the remaining 101 participants produced and quantified complete participation as the percentage of participants that collected data for at least 14 days.

There were 2,949 records with complete information; approximately 28 records from 101 participants living in 16 neighborhoods. The records that each participant collected over time were not statistically independent. Standard logistic models treat observations as independent and yield incorrect standard errors if applied to non-independent data.¹³ Therefore, we used hierarchical logistic regression models to correctly model the variation in outcomes (1) within people over time, (2) between people within neighborhoods, and (3) between neighborhoods as a function of predictor variables.

We fit these models using the GLMM procedure in STATA.¹⁴ We included random intercepts in all models to account for different levels of errors and missing data between people and neighborhoods. We chose random over fixed slopes if the 1 degree of freedom likelihood ratio test comparing the fit of the two models was greater than 2.706. Random slope models accommodate differences in the associations between the predictor variables and outcomes, between people and neighborhoods. In random slope models, we assumed zero covariance between the random intercept and slope. Generally, we fit two-level models to accommodate variations (1) within people over time and (2) between people within neighborhoods. When a community-based representative was the independent variable, we added (3) a third level to accommodate between neighborhood variations because the community-based representative was measured at the neighborhood level.

We adjusted for a potential confounder if it produced a 10% or more change in the β coefficient. We explored potential effect measure modifiers (EMM) by adding the covariate

and an interaction term between the covariate and the main effect to the model. We report interaction terms with Wald test values greater than 1.282.

We made the following a priori decisions about potential covariates. In analyses in which week-in-participation was the independent variable, we evaluated the following as EMMs: community-based representative presence or absence (absent vs. present); gender (female vs. male); employment status (employed vs. unemployed); and age (>55 years of age vs. ≤55 years). For models with odor rating as the main effect, we considered time of day (evening vs. morning record) as a potential modifier and confounder and both odor sensitivity threshold (low vs. high) and gender as potential modifiers. For models with presence or absence of a community-based representative as the main effect, we considered age, gender, and employment status as EMMs. We examined gender and age as potential modifiers in analyses of time-of-day effects.

We report ORs and 95% CI because they are commonly used in epidemiology. We present Wald test statistics rather than *P* values because the latter are typically misinterpreted in nonrandomized studies.¹⁵ Similarly, the 95% CIs that we report should be interpreted as measures of precision, not as statistical tests of significance since they were estimated from data that derive from a nonrandomized study.

Results

Most participants collected data on 28 occasions over a 2-week period and produced 28 records. Eight community members produced fewer than 28 records but only two participated for fewer than 14 days. Ninety-eight percent of participants fully completed the study. In three neighborhoods, community members extended participation by 4 to 7 days; 15 participants produced more than 28 records.

Table 2 describes the 101 participants' demographic characteristics and the number of journal records that they produced. Eighty-five participants described themselves as African American and one as Latino. Ages ranged from 19 to 89 years (mean, 53). Two thirds were female and 58 were employed. There was a community-based representative at 10 study sites. Because community-based representatives were also study participants, their demographic composition was similar to that of the CHEIHO study population. In approximately half of records, participants smelled hog odor while they sat outside for 10 minutes. Participants produced roughly the same number of records during and after the first week of their participation and before as after noon.

We quantified the percentage of the 2,949 records from which each of the individual variables in the journals were missing. These percentages are not shown in the tables. Among the 52 individual variables in each record, the rating of nasal irritation was most complete; it was missing in 1% of records. Participants were supposed to record their third FEV1 reading in their journal. This variable was the least complete of the individual variables; it was missing in 20% of records. Table 2 shows percentages of records with AirWatch errors, sequence errors, as well as missing summary variables and saliva tube samples. On average, individual variables were incomplete in 2% of records. Only 2% of

records had a sequence error. This percentage is calculated out of 2,932 records because in the other 17 records, time information was missing, and we were unable to determine if the participant performed activities out of order. Of the summary variables, the ordered response group was most complete; one or more of these were missing in 12% of records. The least complete summary variable was the machine-use variable, with 26% missing. In 34% of records, the participant produced no error-free AirWatch trials. The denominator for this percentage was 2,918 because 31 records were excluded owing to failure of the monitor to record readings. Eight percent of the records were missing saliva samples.

Table 3 shows ORs and 95% CI as estimates of associations between predictor and outcome variables. A footnote indicates the main effect models that included a random slope. Including a random slope improved the fit of all bivariate and most multivariate models with week-in-participation as the predictor. Therefore, results from random slope models are reported for bivariate models in which week-in-participation was the main effect. However, when we added an interaction term between employment status and week in the model with sequence error as an outcome and between gender and week when ordered response was the outcome, the random slope component no longer improved the fit of the model and so we reported the simpler fixed slope model results. In analyses of predictor variables other than week-in-participation, bivariate and multivariate models that included a fixed slope fit the data as well as those that included a random slope. The fixed slope results are presented for these models. Only observations with complete data on odor were included in analyses of associations between odor and ordered response variable completeness ($n = 2,772$). None of the odor estimates are adjusted for time-of-day because this variable did not change the magnitude of the estimate by 10% or more.

The odds that machine-use variables were missing were lower after the first week of the study (OR, 0.41; 95% CI, 0.20–0.84). When odor was present, the odds that ordered response items were missing were higher (OR, 1.81; 95% CI, 1.22–2.68), but the odds that there were AirWatch errors were lower (OR, 0.76; 95% CI, 0.57–1.01). There were higher odds of missing saliva samples in records in which participants reported odor (OR, 1.59; 95% CI, 0.97–2.59). The odds that saliva samples were missing were higher in evening records (OR, 1.44; 95% CI, 0.98–2.12). Wald test statistics well below 2.0 indicate that, in most models, independent variables had little ability to predict errors or missing data.

Most EMMs contributed little to the fit of models. However, women (OR, 1.46; 95% CI, 1.01–2.12) and older participants (OR, 1.95; 95% CI, 1.29–2.94) had a higher odds of leaving yes/no variables missing in week 2 records. For these variables, participants indicated affirmative responses by writing long-hand, descriptive responses, or responded in the negative by marking the words “no” or “none.” For male participants, the odds that ordered response variables were missing were lower in week 2 compared with week 1 (OR, 0.58; 95% CI, 0.38–0.90). In communities with a community-based representative, we observed even higher odds that AirWatch readings were flagged for errors in week 2 (OR, 1.96; 95% CI, 1.51–2.52). For unemployed participants, there were fewer sequence errors in week 2 (OR, 0.32; 95% CI, 0.13–0.81). In morning records, the odds that machine variables were missing were higher when odor was reported (OR, 1.64; 95% CI, 0.98–2.75). Men had particularly high odds of leaving ordered response variables missing when there was odor

(OR, 5.27; 95% CI, 2.20–12.66). Finally, for employed participants, we observed lower odds of AirWatch errors if there was no community-based representative present than if there was a representative available (OR, 0.16; 95% CI, 0.02–1.15).

Discussion

Community members collected fairly complete and high quality data and 98% completed the full 2 weeks of participation. We observed sequence errors in only 2% of records and, on average, individual variables were missing in 2% of records. It was necessary that participants understand study protocol to perform activities in their correct order. That we observed sequence errors in only 2% of records implies that participants understood the protocol. Not only does this finding have positive implications for the interpretation of results from CHEIHO and other similarly conducted studies, but it also suggests that one goal of CHEIHO, which was to increase community members' understanding of research design, was met.¹⁰

Our examination of incomplete responses offers information about the most challenging activities for study participants. Ordered response items were more complete than either yes/no or machine-use items. This implies that rating system use was less challenging than other activities. Machine-use responses were the least complete; FEV1 was the most frequently missing individual response and 34% of records did not have usable AirWatch readings owing to errors. These findings suggest that the most challenging activities involved machine use, particularly of the AirWatch monitor.

The high frequency of AirWatch errors reflects the fact that measurement of lung function is difficult, even in a clinical setting. Coughing, incorrect hand placement, and/or incorrect use of the mouthpiece produce errors that would have caused readings to be unusable. The AirWatch device has a faint display. Accessing FEV1 readings requires pressing a button for exactly 3 seconds; pressing for less time does not display the reading. Older participants and participants with less formal education, many of whom were less familiar with technical equipment, found the AirWatch difficult to use. Participants who had difficulty navigating the device would not have accessed their FEV1 readings and left that field blank in their journals. In most analyses, week-in-study, odor, presence of a community-based representative, and time of day showed little association with errors or missing data. The lower odds that machine-use variables were missing from records that participants produced in week 2 compared with week 1 might suggest that community members improved owing to increased experience or after receiving assistance from staff during midstudy visits. However, we observed higher odds that participants produced AirWatch errors in week 2; participants might have forgotten the correct AirWatch use technique after this length of time after the training. Feedback during midstudy check-ins would have done little to prevent AirWatch errors because the machine maintained readings internally and study staff and community-based representatives were not able to identify technique problems. It is unclear why, when there was a community representative present, we observed even higher odds of AirWatch errors in week 2. It is possible that representatives offered greater assistance in other parts of the journal that they were able to check and, as a result, participants focused less on AirWatch blowing technique.

Men left fewer ordered response variables missing in week 2 compared with week 1. We also observed higher odds that women and older participants left yes/no variables missing from week 2 records compared with week 1. These findings indicate that predictors of data quality and completeness are not uniform across different subgroups of participants. Also, the activity associated with completing the yes/no variables was to either write comments to indicate an affirmative response or to mark the word “no” or “none.” Women and older participants who left these responses missing more often in week 2 might have tired of writing long-hand responses to questions. Alternatively, to increase time efficiency, participants might have left the space blank when the question did not apply. A solution to this problem might be to provide a place for the participant to circle “yes” if they wish to respond in the affirmative but do not want to dedicate time to writing long-hand answers.

A common criticism of community based studies like CHEIHO is that participants have a vested interest in the topic and this might lead to bias.¹⁶ We examined associations between odor ratings and missing variables and sequence errors and hypothesized that odor might improve data collection efforts owing to participant concern about hog operations. There were no substantial associations between odor rating and sequence errors, incomplete yes/no variables, or missing machine-use variables in evening records. Ordered response items and, in morning records, machine-use items, were less complete when there was odor. Similarly, the odds that saliva samples were submitted were lower when participants smelled odor. Although there were fewer AirWatch errors when there was odor, these findings for the most part do not support the hypothesis that participants collected better data when they smelled odor.

Random intercepts in the mixed models reflect the different levels of errors and missing data between people and neighborhoods. In addition to these differences in levels, associations of some predictor variables with errors and missing data differed between people and neighborhoods, as indicated by the improved fit of models with random slopes. In models with sequence error and ordered response as the outcome, the addition of interaction terms between employment status and week of participation and between gender and week, respectively, eliminated the need for a random slope. This suggests that differences in employment status and gender explained much of the variability in the effect of week on the odds of producing sequence errors or leaving ordered response variables incomplete, respectively. That a random slope component did not improve the fit of other bivariate models indicates that the effect of the remaining predictor variables was relatively constant across people and communities.

There were limitations in our examination of data completeness and quality. It was difficult to examine the effect of variables such as age, gender, or race because these demographic characteristics were not evenly distributed. There was reduced power in analyses that examined community-based representative presence as a predictor variable because this neighborhood-level variable did not vary over time, within people, or within neighborhood. Also, low proportions of records with missing variables or sequence errors resulted in low power to detect associations between errors and potential predictors. The last-stated limitation is only a weakness within the framework of this examination, because it derives

from community members having successfully produced complete and high-quality data. In the context of the CHEIHO study and CBPR, the low power represents a positive finding.

Comparison of the completeness of individual variables from the CHEIHO study, which ranged from 74% to 99% with that from other studies shows that participants collected high-quality data. For example, Warsi et al¹⁷ examined data in three clinical databases and found that completeness of individual items ranged from 11% to 100%. Lewis-Beck¹⁸ writes, “Problems of missing data are pervasive in ... social science research” and that a multivariate analysis of opinion survey data commonly reduces the original sample size by 50% owing to incomplete data.

The CHEIHO study integrated a number of the principles of CBPR.⁶ The basic study questions were based primarily on community concerns, rather than those of government agencies, industry, or academics. The study grew out of a partnership that emphasized environmental injustice, community self-determination, and social change.^{19,20} Also, active collaboration between community members and university researchers was inherent to the study process. Finally, the CHEIHO study considered health to include quality of life and not merely the absence of disease.

The conclusion that participants collected high-quality data is encouraging considering the potential gains that a CBPR approach offers to research, such as improved participant recruitment and retention^{3,4,8} and insight into the research topic and study materials.^{3,4,21,22} Other benefits of CBPR are improved capacity of community members to engage in research and participate in the political process²³ and to push for social change.¹⁶

Quantification of the fact that CHEIHO participants collected high-quality, consistent data is also important given the scientific culture within which CBPR must gain acceptance. Asking community members, who may have a personal and emotional interest in the research topic, to collect data on their own conflicts with scientific standards.¹ Incorporating such a participatory design feature opens the field to criticism of research findings.

A recent literature search for papers on CBPR methodology revealed that scant work has been done to examine the quality of data that participatory studies produce or to quantify the ability of community members to participate in research. Such examinations are important in addressing skepticism about the scientific validity of findings from community-based investigations and facilitate the exchange of information on ways to improve the rigor and quality of CBPR work. Our evaluation helps to fulfill this goal and could promote the adoption of CBPR principles into epidemiologic investigation.

Appendix

Day 1: _____ MORNING Date: I ___ / I ___ / 2004 No data collected

STEP 1. For the past 12 hours, rate the odor from bag operations in the box for --at home and outside, at home and inside, or not at home. If you did not smell an odor, enter a 0 in the box that corresponds to your location(s) at that time. If you were sleeping, enter a 7.

Use the scale to the right to rate the odor.

none	very faint...faint...moderate...strong...very strong
0	1 2 3 4 5 6 7 8

Beginning last night:

	7 pm	10 pm	1 pm	10 pm	11 pm	12 am	1 am	2 am	3 am	4 am	5 am	6 am	7 am	8 am	9 am
Home-outside															
Home-inside															
Not home															

Reminder: Did you enter a number between 0 and 8 in at least one of the three boxes below each hour of the day?

STEP 2. Since last night, please note:

- up to three things you wanted to do at your home but decided not to because of bag odor.

1. _____ 2. _____ 3. _____
- up to three things you did at your home that you did differently or with difficulty because of bag odor.

1. _____ 2. _____ 3. _____
- If bag odor did not alter your activities, circle: NONE

Day 1: _____ MORNING

STEP 3. Rate the odor from bag operations for the 10 minutes you spent outside.

Current Time	Circle number that best describes the strength of the odor when it was strongest	How would you describe the odor?	Was the odor continuous or did it come and go?
_____ AM	none very faint...faint...moderate...strong...very strong		<input type="checkbox"/> continuous <input type="checkbox"/> came and go?
	0 1 2 3 4 5 6 7 8		

STEP 4. Did you experience any irritation while outside? None (If yes, check all that apply) eyes? nose? throat? skin? cough?

STEP 5. How do you feel now? (Circle the number that best describes how you feel)

	not at all	a little	moderately	quite a bit	extremely				
Stressed or annoyed?	0	1	2	3	4	5	6	7	8
Nervous or anxious?	0	1	2	3	4	5	6	7	8
Clonny, blue or outgassed?	0	1	2	3	4	5	6	7	8
Angry, grouchy or bad-tempered?	0	1	2	3	4	5	6	7	8
Clothead or unable to concentrate?	0	1	2	3	4	5	6	7	8

STEP 6. Take your blood pressure twice, print the results, and tape the printout in this box.

DO NOT PLACE TAPE OVER THE NUMBERS

Day 1: _____ MORNING

STEP 7. Since your last journal entry, have you had...? Circle the number that best describes how each problem was.

	not at all	a little	moderate	quite a bit	extreme		not at all	a little	moderate	quite a bit	extreme					
Cough	0	1	2	3	4	5	Frequency	0	1	2	3	4	5	6	7	8
Chest tightness	0	1	2	3	4	5	Shortness	0	1	2	3	4	5	6	7	8
Difficulty breathing	0	1	2	3	4	5	Fever	0	1	2	3	4	5	6	7	8
Wheezing or whistling	0	1	2	3	4	5	Pain	0	1	2	3	4	5	6	7	8
Runny nose	0	1	2	3	4	5	Nausea	0	1	2	3	4	5	6	7	8
Irritation or burning of the nose	0	1	2	3	4	5	Shortness of breath	0	1	2	3	4	5	6	7	8
Itchy or watery eyes	0	1	2	3	4	5	Cold or the flu	0	1	2	3	4	5	6	7	8
Sore throat	0	1	2	3	4	5	Stomach ache	0	1	2	3	4	5	6	7	8
Blushing or redness of the face	0	1	2	3	4	5	Aching or painful joints	0	1	2	3	4	5	6	7	8
Burning eyes	0	1	2	3	4	5	Headache	0	1	2	3	4	5	6	7	8
Itching nose	0	1	2	3	4	5	Lightheaded or dizzy	0	1	2	3	4	5	6	7	8

Day 1: _____ MORNING

STEP 8. In the past 12 hours, did you take any medication other than what you usually take? Yes No
If yes, what? _____

STEP 9. Collect 2 minute saliva sample. Record time you start collecting sample. _____ AM or circle: no sample
Record tube #: _____

STEP 10. Blow into Air Watch three times. Record the LAST reading below

Peak flow	LM
FEV-1	L

COMMENTS:

References

1. Parry O, Gnich W, Platt S. Principles in practice: Reflections on a 'postpositivist' approach to evaluation research. *Health Educ Res.* 2001; 16:215–26. [PubMed: 11345663]
2. Bradbury, H., Reason, P. Issues and choice points for improving the quality of action research. In: Minkler, M., Wallerstein, N., editors. *Community based participatory research for health.* San Francisco: Jossey-Bass; 2003. p. 201-22.
3. Leung MW, Yen IH, Minkler M. Community based participatory research: A promising approach for increasing epidemiology's relevance in the 21st century. *Int J Epidemiol.* 2004; 33:499–506. [PubMed: 15155709]
4. Minkler M. Community-based research partnerships: Challenges and opportunities. *J Urban Health.* 2005; 82(2 Suppl 2):ii3–12. [PubMed: 15888635]

5. Burhansstipanov L, Christopher S, Schumacher SA. Lessons learned from community-based participatory research in Indian country. *Cancer Control*. 2005; 12(Suppl 2):70–6. [PubMed: 16327753]
6. Israel BA, Schulz AJ, Parker EA, et al. Review of community-based research: Assessing partnership approaches to improve public health. *Annu Rev Public Health*. 1998; 19:173–202. [PubMed: 9611617]
7. Israel BA, Parker EA, Rowe Z, et al. Community-based participatory research: Lessons learned from the centers for children's environmental health and disease prevention research. *Environ Health Perspect*. 2005; 113:1463–71. [PubMed: 16203263]
8. O'Fallon LR, Deary A. Community-based participatory research as a tool to advance environmental health sciences. *Environ Health Perspect*. 2002; 110(Suppl 2):155–9. [PubMed: 11929724]
9. Calnan M. Commentary: The people know best. *Int J Epidemiol*. 2004; 33:506–7. [PubMed: 15155705]
10. Wing S, Horton RA, Muhammad N, et al. Integrating epidemiology, education, and organizing for environmental justice: Community health effects of industrial hog operations. *Am J Public Health*. 2008; 98:1390–7. [PubMed: 18556620]
11. Wing S. Social responsibility and research ethics in community-driven studies of industrialized hog production. *Environ Health Perspect*. 2002; 110:437–44. [PubMed: 12003746]
12. Wing S, Horton Avery R, Marshall SW, et al. Air pollution and odor in communities near industrial swine operations. *Environ Health Perspect*. 2008; 116:1362–8. [PubMed: 18941579]
13. Subramanian, SV., Jones, K., Duncan, C. Multilevel methods for public health research. In: Kawachi, I., Berkman, LF., editors. *Neighborhoods and health*. New York: Oxford University Press; 2003. p. 65-111.
14. Rabe-Hesketh S, Skrondal A, Pickles A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Statistics Journal*. 2002; 2:1–21.
15. Greenland S. Randomization, statistics, and causal inference. *Epidemiology*. 1990; 1:421–9. [PubMed: 2090279]
16. Nyden P. Academic incentives for faculty participation in community-based participatory research. *J Gen Intern Med*. 2003; 18:576–85. [PubMed: 12848841]
17. Warsi AA, White S, McCulloch P. Completeness of data entry in three cancer surgery databases. *Eur J Surg Oncol*. 2002; 28:850–6. [PubMed: 12477477]
18. Allison, PD., Lewis-Beck, MS., editors. *Missing data*. Thousand Oaks (CA): Sage; 2002. Introduction; p. v
19. Wing S, Cole D, Grant G. Environmental injustice in North Carolina's hog industry. *Environ Health Perspect*. 2000; 108:225–31.
20. Wing S, Grant G, Green M, et al. Community based collaboration for environmental justice: South-east Halifax environmental reawakening. *Environment and Urbanization*. 1996; 8:129–40.
21. Entwistle VA, Renfrew MJ, Yearley S, et al. Lay perspectives: Advantages for health research. *Br Med J*. 1998; 316:463–6. [PubMed: 9492683]
22. Lipscomb HJ, Argue R, McDonald MA, et al. Exploration of work and health disparities among black women employed in poultry processing in the rural south. *Environ Health Perspect*. 2005; 113:1833–40. [PubMed: 16330373]
23. Buchanan DR, Miller FG, Wallerstein N. Ethical issues in community-based participatory research: Balancing rigorous research with community participation in community intervention studies. *Prog Commun Health Partnersh*. 2007; 1:153–60.

Table 1
Names and Definitions of Dependent Variables and Journal Questions From Which These Are Derived

	Description	Individual Variables From Which the Variable Was Derived	No. of Individual Variables From Which Summary Variable Was Derived
Sequence Error	A sequence error occurred if a participant recorded an outcome variable before she or he sat outside for 10 minutes. If the time at which a participant indicated having sat outside was 5 or more minutes later than either the time at which she or he had taken their blood pressure, as indicated by time stamped blood pressure readings or the time that the participant recorded having collected a saliva sample, then we defined the record as having a sequence error.	Participant's recording of the time at which she or he returned from spending 10 minutes outdoors. If protocol was followed correctly, then this should have been the first data collection step because it ensured participants' exposure to the outdoor air. The times at which participants collected each blood pressure reading. These times appeared on print-outs from the blood pressure machine. The time at which the participant collected their saliva sample, which the participant recorded in their journal.	
Yes/No Variable	The yes/no variable was based on the completeness of variables for which the participant wrote long-hand, descriptive answers if their answer to a question was in the affirmative. Otherwise, the participant indicated the negative by either circling or checking the word no or none. A yes/no variable was defined as missing when the participant did not check the option "none" and did not write an affirmative statement.	Participant's indication of whether she or he changed their daily activities due to the hog odor over the past 12 hours. Participants either described the ways that they had changed their activities due to odor or circled the word "none" if they did not change their activities. Participant's indication of whether she/he experienced irritation of the eyes, nose, throat, or skin while she/he sat outside for 10 minutes. Participants either checked a box beside all applicable types of irritation that they experienced or checked a box beside "none" if they did not experience irritation. Participant's indication of whether she or he took any medications, other than those that they usually take, in the previous 12 hours. Participants either wrote in the medications that they took or circled the word "none" if they had not taken any medications.	3
Ordered Response Variable	The ordered response variable was based on the completeness of variables for which the participant circled a number on a scale of 0 to 8.	Participants rated the strength of hog odor that she or he smelled while sitting outside for 10 minutes by circling a number between 0 and 8, with 0 indicating no odor and 8 representing strong odor. Participants rated the extent to which she or he experienced 5 dimensions of mood by circling a number between 0 and 8. 0 corresponded to "not at all" and 8 corresponded to "extremely." The 5 mood variables are listed in step 5 of the Appendix. Participant's rated the extent to which she or he experienced 22 health symptoms in the previous 12 hours. 0 corresponded to "not at all" and 8 corresponded to "extreme." These 22 health symptoms are listed in step 7 of the Appendix.	28
Machine-Use Variable	The machine-use variable was based on the completeness of variables for which the participant used a machine. For the AirWatch monitor variables, the machine-use variable was not based on those numbers that the AirWatch maintained internally. Rather, the machine variable was based on the completeness of the AirWatch values that the participant recorded in his or her journal (Appendix 1, page 4).	FEV1 and peak flow readings that participants wrote in their journals. Two pulse, diastolic, systolic, time, and date readings that derived from 2 uses of the blood pressure monitor. These readings appeared on a print-out that the participant taped into his/her journal.	12

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2
Demographic Characteristics of CHEIHO Study Participants and the Proportion of Records Produced by Participants That Had Missing Variables, Sequence Errors, or No Usable Air-Watch Readings

	No. of Participants	No. of Journal Records	No. (%) of Records With Sequence Errors*	No. (%) of Records With No Usable Air-Watch Reading*	No. (%) of Records With Incomplete Yes/No Variable(s)*,†	No. (%) of Records With Incomplete Ordered Response Variable(s)*,‡	No. (%) of Records With Incomplete Machine-Use Variable(s)*,‡	No. (%) of Records With a Missing Saliva Sample*
Total	101	2,949	54 (2)	995 (34)	572 (19)	344 (12)	767 (26)	239 (8)
Gender								
Female	66	1,945	49 (3)	666 (35)	355 (18)	222 (11)	534 (27)	47 (5)
Male	35	1,004	5 (1)	329 (33)	217 (22)	122 (12)	233 (23)	192 (10)
Race								
Non-White	86	2,441	47 (2)	868 (36)	438 (18)	288 (12)	540 (22)	138 (6)
White	15	508	7 (1)	127 (25)	134 (26)	56 (11)	227 (45)	101 (20)
Age (yrs)								
55	55	1,522	21 (1)	461 (31)	290 (19)	167 (11)	183 (12)	118 (8)
>55	46	1,427	33 (2)	534 (38)	282 (20)	177 (12)	584 (41)	121 (8)
Employment Status								
Unemployed or Retired	40	1,248	24 (2)	411 (33)	185 (15)	152 (12)	402 (32)	128 (10)
Employed	58	1,617	26 (2)	560 (36)	377 (23)	185 (11)	329 (20)	111 (7)
Missing	3	84	4	24	10	7	36	0
Community Representative								
Present	67	1,931	35 (2)	697 (37)	368 (19)	258 (13)	448 (23)	63 (3)
Absent	34	1,018	19 (2)	298 (30)	204 (20)	86 (8)	319 (31)	176 (17)
Week-in-Participation								
1	101	1,374	28 (2)	697 (51)	252 (18)	165 (12)	380 (28)	95 (7)
2	101	1,575	26 (2)	298 (19)	320 (20)	179 (11)	387 (25)	144 (9)
Odor Rating								
0	88	1,419	27 (2)	479 (34)	240 (17)	58 (4)	322 (23)	89 (6)
1-8	96	1,353	24 (2)	447 (33)	275 (20)	129 (10)	372 (27)	129 (10)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	No. of Participants	No. of Journal Records	No. (%) of Records With Sequence Errors*	No. (%) of Records With No Usable AirWatch Reading*	No. (%) of Records With Incomplete Yes/No Variable(s) [†]	No. (%) of Records With Incomplete Ordered Response Variable(s) [†]	No. (%) of Records With Incomplete Machine-Use Variable(s) [†]	No. (%) of Records With a Missing Saliva Sample*
Missing	56	177	3	69	57	157	73	21
Time of Day								
Morning	101	1483	29 (2)	509 (35)	280 (19)	165 (11)	396 (27)	111 (7)
Evening	101	1466	25 (2)	486 (34)	292 (20)	179 (12)	371 (25)	128 (9)

* Total journal records in each category with complete outcome information serve as the denominator for the calculation of these percentages.

[†] This is a summary variable and is defined on the basis of the completeness of three or more individual variables

Table 3
ORs, 95% CI, and Wald Z Test Statistics for the Relationships Between the Log-Odds That a Journal Record Was Missing Variable(s) or Had an Error and Week-in-Participation, Odor Rating, Presence or Absence of a Community-Based Representative, and Time of Day at Which Data were Collected*

	Week 2 and Beyond vs. Week 1		Any Odor Rating vs. No Odor Rating		Representative Absent vs. Representative Present		Evening vs. Morning	
	OR (95% CI)	Wald Z Statistic	OR (95% CI)	Wald Z Statistic	OR (95% CI)	Wald Z Statistic	OR (95% CI)	Wald Z Statistic
Sequence error	0.40 (0.13–1.29) [‡]	-1.53	1.54 (0.74–3.19)	1.15	2.40(0.51–11.28)	0.93	1.17 (0.65–2.12)	0.53
Air Watch error	1.40 (0.92–2.14) [‡]	1.57	0.76 (0.57–1.01)	-1.90	0.39 (0.09–1.66)	-1.27	0.93 (0.75–1.14)	-0.70
Missing								
Yes/no variable	1.17 (0.86–1.59) [‡]	1.01	1.11 (0.83–1.48)	0.72	1.20 (0.62–2.31)	0.53	1.13 (0.91–1.41)	1.13
Ordered response variable	0.85 (0.62–1.16) [‡]	-1.03	1.81 (1.22–2.68)	2.96	0.58 (0.30–1.12)	-1.62	1.12 (0.87–1.44)	0.91
Machine-use variable	0.41 (0.20–0.84) [‡]	-2.45	1.18(0.79–1.78)	0.80	2.49 (0.29–21.66)	0.82	0.86 (0.64–1.15)	-1.01
Saliva sample	0.81 (0.38–1.72) [‡]	-0.56	1.59 (0.97–2.59)	1.85	5.21 (0.42–64.29)	1.29	1.44 (0.98–2.12)	1.85

* All estimates shown derive from bivariate models.

[‡] Estimate derives from a model that included a random slope component, in addition to a random intercept.