VOLUME 5, ISSUE 4

FOCUS on Field Epidemiology

Cluster Investigations of Non-Infectious Health Events

CONTRIBUTORS

Author:

Alyssa J. Mansfield, MPH

based on materials by

Jeanette K. Stehr-Green, MD

Paul A. Stehr-Green, DrPH, MPH

Reviewers:

FOCUS Workgroup*

Production Editors:

Tara P. Rybka, MPH

Lorraine Alexander, DrPH

Rachel A. Wilfert, MD, MPH

Editor in chief:

Pia D.M. MacDonald, PhD, MPH

*** All members of the FOCUS Workgroup are named on the last page of this issue.**

It's Friday afternoon, and you get a call from a concerned citizen. (Don't these calls ALWAYS come in on a Friday afternoon?)

She and her community action group are sure the effluent from a nearby toxic waste incinerator is causing cancers in the neighborhood—and they want to know what you're going to do about it.

What do you do, if anything? How do you determine whether there is a problem? Where should you start? Who should be involved? And what do you tell the local TV reporter who's calling to get your agency's reaction to reports of the "toxic waste cancer generator"?

Welcome to the world of non-infectious disease "cluster busters," a challenging and evolving facet of epidemiology and the topic of this issue.

Investigation of clusters of non-infectious disease is a critical public health function. Cluster investigations can result in major epidemiologic breakthroughs as they link specific exposures to diseases.

For example, in the early 1960s, several European countries saw a marked increase in the number of infants born with a severe limb deformity called phocomelia. The deformities were linked to maternal use of thalidomide, a sleeping pill and treatment for morning sickness during pregnancy. (1) The drug was never marketed in the US, but the problem, in part, led Congress to enact legislation in 1962 that began the rigorous testing now required for approval of pharmaceutical products in the US.

Several other notable investigations (Table 1), have linked a non-infectious disease cluster with a specific exposure, but these are the exceptions.

Table 1. Clusters leading to the identification of new exposure-disease relationships

Population	Year	Exposure	Outcome
expectant mothers	1962	thalidomide	phocomelia
workers	1968-1973	vinyl chloride	angiosarcoma
chemical workers	1978	kepone (pesticide)	infertility
jewelry wearers	1984	radioactively contaminated gold rings	dermatitis and skin cancer
drug users	1983	N-methyl-4-phenyl-1, 2, 5, 6 tertahydropyridine	Parkinson-like disorder
health food consumers	1989	L-tryptophan (contaminated)	eosinophilia-myalgia syndrome
dieters	1997	fenfluramine-phentermine	valvular heart disease

UNC SCHOOL OF PUBLIC HEALTH

NORTH CAROLINA CENTER FOR PUBLIC HEALTH PREPAREDNESS

The North Carolina Center for Public Health Preparedness is funded by Grant/Cooperative Agreement Number U90/CCU424255 from the Centers for Disease Control and Prevention. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the views of the CDC.

Few investigations have added to our knowledge of disease causes or provided a credible explanation of why the cluster of diseases occurred.

Why is this? First, in many non-infectious disease cluster investigations, an apparent geographic or temporal excess in the number of cases cannot be confirmed. In part that is because many non-infectious diseases reported as **clusters** (e.g., cancers, birth defects) are common in the community, and supposed clusters reflect normal disease patterns. Also, confounding factors such as older age might account for a seemingly higher **occurrence rate**. Some diseases like brain cancer might look alike to the community but have different pathogenic processes (i.e., a primary brain tumor or brain metastases that have spread from cancer in another organ). Although such cases might be grouped together in time and space, they are not really linked.

Second, establishing a definitive cause-and-effect relationship is often impossible because of small case numbers, problems in isolating a single potential **exposure**, and difficulty in reconstructing exposure histories due to long gaps between exposure and symptom onset. (2)

Finally, legitimate questions raised by a cluster investigation often require a separate large-scale epidemiologic study, which may be difficult to carry out.

To increase the chances of successfully investigating a non-infectious disease cluster and to ensure that the cluster investigation proceeds smoothly from one action to the next, an agency needs a standardized step-wise process for receiving and evaluating cluster reports. This generally includes a centralized tracking system, data collection tools, clear lines of communication, specific actions to be taken at each step and criteria for proceeding with or concluding the investigation.

A well trained staff and adequate resources are needed to respond to cluster reports in a coordinated, consistent, and timely way. Investigators should be familiar with the epidemiology, natural history, and diagnosis of the disease; possess skills in environmental or biological sampling; have access to adequate laboratories; and have experience in recognizing patterns in space and time.

In addition, any perceived problem must be addressed responsibly and sympathetically, even if no underlying disease cluster actually exists. Communications officials should be consulted about what, when, how, and to whom information is provided. Community residents, health professionals, and government representatives should be involved in both the investigation and communication efforts, because they can make communications more credible, bring a greater understanding of the situation, and help the agency make better decisions.

To investigate...or not investigate?

Clusters are usually based on small case numbers, cannot easily be investigated, and tend to involve conditions in which little is known about the usual causes.

Investigating a link between exposure and disease is often impossible because of the amount of time since the exposure or because it is not clear what to investigate. Nonetheless, it is important to respond to threats perceived by the public. Approach these investigations with the following in mind:

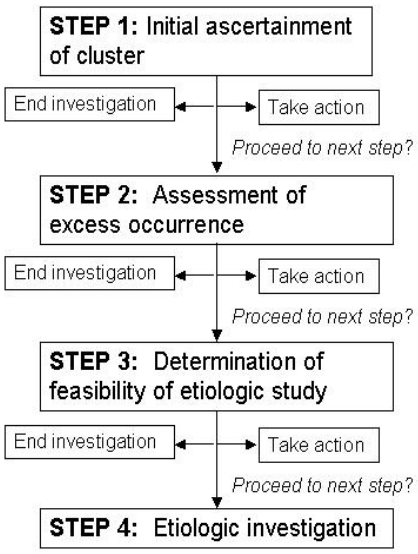
- Follow a step-wise process with decision points based on available information.
- Make your organization's policy of using a step-wise process widely known and understood among the medical community, the general public, the news media, and other key stakeholders *BEFORE* you get that next call reporting a cluster of disease X that "everyone knows" is related to exposure Y.
- Undertake any investigation with a deliberate and transparent approach.
- Take local concern about the cause of the disease cluster seriously, but stay within your pre-stated investigation process.
- Develop effective methods of communication, maintain your objectivity, and provide leadership on controversial and difficult issues.

Basic Steps in Investigating Non-infectious Disease Clusters

Each step in a cluster investigation requires investigators to collect and analyze data, then decide whether immediate action is needed and whether to proceed with the next step of the investigation (Figure 1). (3) Although these steps do not always need to be followed in order, they ensure consistent attention to the appropriate issues.

As you review the steps, keep in mind that while many cluster investigations require action only at the local level, others might require state or even national involvement.

Figure 1. Flowchart of cluster investigation



STEP 1: Initial ascertainment of cluster

To begin a cluster investigation, investigators should collect the following data:

- identifying information from the person reporting the cluster (name, address, phone number, organization);
- demographic information (person, place, time) for cluster cases (number of suspected cases, sex, race, age at diagnosis or death [if appropriate]); geographic area; and time period of concern;
- clinical information on cluster cases (diagnosis date, basis for diagnosis, relevant medical histories, status of cases [deceased, hospitalized, recovered], suspected exposures); and
- identifying information for cluster cases (name, location, phone number, contact person, attending physician's contact information).

Investigators should enter the information into a tracking system; computer programs such as EpiInfo, Microsoft Access, or Microsoft Excel can be used for this. They should also notify other health department staff (e.g., supervisor, subject matter experts, communications/ public relations staff) of the investigation as well as local health officers in the jurisdiction where the cluster occurred, and staff from agencies with jurisdiction or expertise in the area (e.g., departments of labor, environment, or ecology). Staff should then seek information on the causes of the disease under investigation and discuss how this information compares with the reported cluster.

After gathering this initial information, investigators must decide whether to continue the investigation. Criteria for continuing the investigation typically include:

- clinically similar health events without a plausible alternative etiology (e.g., a cluster of lung cancer cases with no history of smoking);
- an apparent **excess occurrence** of such health events;
- a plausible temporal association with the possible exposure(s);
- a disease present in a particular demographic group in which it is not normally found (e.g., Kaposi's sarcoma in young homosexual men); or
- one or more cases of a very rare disease.

If none of these criteria are met, further evaluation might not be warranted. Whenever an investigation is halted, investigators should prepare a brief summary report to be shared with the person reporting the cluster and appropriate supervisors at the health department.

If the investigation is halted at this point, it is important to help the person reporting the cluster to understand the following:

- Cancer is a common illness (with a 1 in 3 lifetime probability). The risk increases with age, and thus cases among older persons are less likely to be true clusters.
- Variety in diagnoses (e.g., different types of cancers) argues against a common origin.
- Length of time in residence must be substantial to implicate an environmental carcinogen because of the long **latency period** required for most known carcinogens.

STEP 2: Assessment of excess occurrence

Estimating Excess Occurrence. Investigators must confirm whether the number of cluster cases is greater than expected in a given population during a certain time period. To determine whether there is an excess occurrence of a non-infectious disease, investigators must estimate an **occurrence rate** based on the number of people with the health event (numerator) and the total population at risk (denominator). The population at risk consists of all people in the geographic area where the supposed exposure occurred over a designated time period.

To establish the number of cases and population at risk, investigators must first select an appropriate geographic area and time period. The geographic area should be large enough to include all persons at risk for the health event, but not so large that it dilutes the occurrence rate by including those not at risk. The designated time period should be consistent with the time period during which the supposed exposure took place (if one has been noted) and allow for a long enough latency period. If the geographic area and time period are defined too narrowly or too broadly, the size of the problem might be over- or underestimated.

- For example, Figure 2 shows how changing the size of the geographic area affects the occurrence rate. The gray circles represent people with the health event and white circles represent those without the health event. If we focus narrowly on the neighborhood immediately surrounding the cases initially reported (left), the occurrence rate is 20% (7 cases among 35 people). But if we focus broadly on everyone in the town (right), the occurrence rate is 8% (8 cases among 100 people).

Investigators must also determine which cases from the reported cluster to include in a preliminary analysis. At this point, all reported cases are usually assumed to be real. However, cases might be excluded if they differ substantially from other cases. Examining the clinical, demographic, and epidemiologic characteristics of cases might help you decide which — if any — cases to exclude.

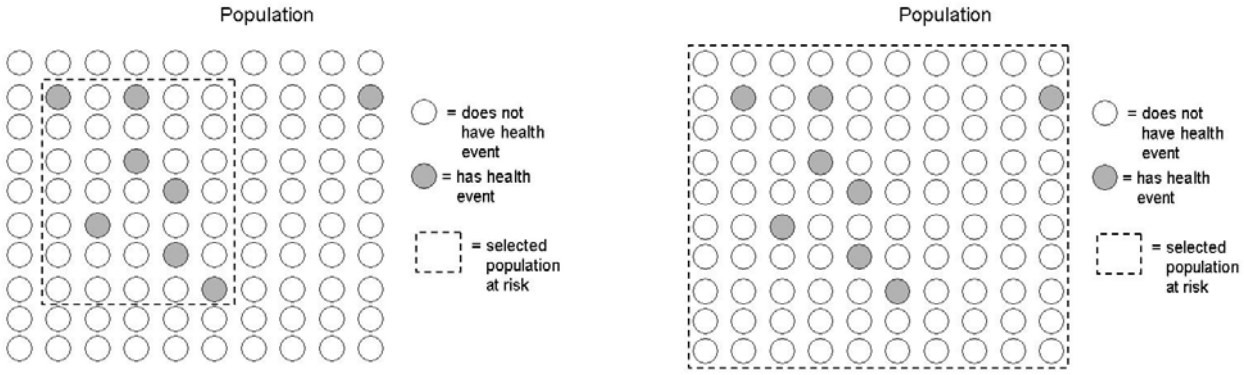
Investigators then find a reference population that is comparable to the population in which the cluster occurred. Typical reference populations include residents from a similar geographic area or data from the same area but in a different time period. Investigators estimate an expected occurrence rate for the reference population from existing surveillance data (cancer registries, birth defect registries) and data from other sources (hospital discharge databases, vital statistics records).

Finally, investigators compare the observed occurrence rate based on the cluster with the expected rate from the reference population to determine whether there is **excess occurrence** of the disease. (Appropriate ways to compare rates are discussed in *FOCUS Volume 3, Issue 6: Data Analysis: Simple Statistical Tests*.)

If there are 5 or more cases in the cluster and an appropriate denominator is available, you can compare the rates using Chi-square tests or Poisson regression. If the case numbers are too small, you might group cases across geographic areas (such as adjacent counties) or time periods (5 years versus 1 year). If grouping cases is not possible or does not make sense, or if denominator data are not available, you can use statistical tests developed to assess space, time, or space-time clustering. (See the Appendix to the CDC's "Guidelines for investigating clusters of health events.") (3)

Case Verification. If the occurrence of the disease is determined to be in excess, investigators should then verify that all cases have the disease in question using a clear, consistently applied case definition to be sure there is a biological basis for further investigation. Otherwise, the cluster might represent a mixture of health events with no common etiology.

Figure 2. Finding the occurrence rate in the population at risk



A case definition includes clinical criteria (signs, symptoms, confirmatory tests) and restrictions on time, place, and person. A *sensitive* case definition uses broad criteria and may include several different diseases or health events thought to be related by some pathogenic process. It typically captures more true cases but also includes "cases" who do not have the health event of interest (false positives). A *specific* case definition is based on narrower criteria, usually includes confirmatory testing and focuses on only one health event. Specific case definitions are likely to exclude true cases (false negatives) but are less likely to include false positives.

- For example, if a health department is evaluating a possible cluster of cancer cases apparently linked to long-term benzene exposure, they might use a sensitive case definition (individuals diagnosed with any form of blood cancer) or a more specific case definition (individuals diagnosed with leukemia).

Including multiple health events in a case definition increases the case numbers, but if the cases are not truly related (i.e., they are due to different etiologic agents), it becomes difficult to find an explanation for the cluster. Therefore, a case definition with multiple health events should be used only when there is reliable scientific information about the presumed causal exposure.

Investigators might also use more than one case definition.

- For example, in 1995, the New Jersey Department of Health and Senior Services investigated reports of childhood cancer in Dover Township, which includes the Toms River section of southeast New Jersey. Over several decades, chemical plants and other businesses released industrial pollutants into the Toms River. Contamination eventually reached the well supplying Dover's drinking water. (4)

  Investigators evaluated all childhood cancers combined and subgroups of selected childhood cancers. Cases were identified through the New Jersey State Cancer Registry. Expected numbers of cases were based on statewide average annual age- and sex-specific numbers for 1979 to 1995.
  The investigators compared observed and expected occurrence rates by calculating standardized incidence ratios (SIR) and 95% confidence intervals. (4)

$$SIR = \frac{\text{observed cases (or rate)}}{\text{expected cases (or rate)}}$$

where = 1 no excess occurrence
> 1 possible excess occurrence
< 1 observed is less than expected

  An excess occurrence of childhood cancer was found for all cancers combined and for acute lymphocytic leukemia in females in Dover Township; in Toms River, an excess occurrence was found for all cancers combined, for brain and central nervous system cancers, and for acute lymphocytic leukemia, particularly in female children under the age of 5 (Table 2). (4)

  In subsequent etiologic studies, no factor was identified as solely responsible for the general elevation in childhood cancer. However, findings supported prenatal exposure to township well water (1982-1996) and air emissions from a chemical company as risk factors for leukemia in female children. (4)

Table 2. Childhood cancer incidence in Toms River census tracts 1979-1995, children 0-4 years

Cancer type	No. observed	No. expected	SIR	95% CI
All cancers combined	24	14.4	1.7	1.07-2.49
Brain/central nervous system	4	0.6	7	1.87-17.8
Astrocytoma*	2	0.2	8.8	1.00-32.1
Acute lymphocytic leukemia**	4	0.4	9.4	2.52-24.0

*One specific type of cancer of the brain/nervous system
**Females only

Case-Finding. Once the case definition is clear and diagnoses have been verified, investigators must verify the cases. To determine if cases meet the case definition, investigators should:

- Contact the attending physician(s) to obtain permission to examine case-patients' medical records, and refer to any relevant health registries (e.g., surveillance reports, cancer registries, and/or medical charts).
- If possible, obtain copies of relevant pathology reports, laboratory or other diagnostic tests, and the medical examiner's reports.
- Obtain clinical/laboratory reevaluations (e.g., reexamination of biopsy specimens or retesting of specimens), if needed.

Case verification is often a multi-step process involving contact not only with the attending physician but also contact with the case-patients, their family members or friends.

If excess disease occurrence is still evident after verification of known cases, investigators might expand the assessment and undertake additional case-finding to better define cluster characteristics. Case-finding often requires a field investigation involving health care providers and the general community.

In an expanded assessment, investigators should:

- reconsider the initial case definition to determine if greater sensitivity or specificity is desired;
- reassess to determine the most appropriate geographic and time boundaries;
- ascertain all potential cases within the defined geographic and time boundaries;
- identify appropriate database sources for cases and the population at risk, their availability and quality;
- perform an in-depth review of the medical literature and consider whether the supposed association is epidemiologically and biologically plausible; and
- assess the likelihood that clustered health events are related statistically, temporally, and physiologically to the supposed exposure(s).

Ensuring complete case ascertainment may involve reviewing additional data sources or medical records and obtaining additional information from the community.

Although a community survey might find previously unknown cases, formal surveys are generally reserved for later stages. In earlier stages, investigators need to be cautious about providing public information to find cases, in order to avoid biasing people who eventually might be formally surveyed.

If investigators find excess occurrence of disease and compelling evidence for an association with the supposed exposure, they should consider the feasibility of an etiologic study. If excess occurrence is not confirmed, or is confirmed with no apparent plausible relationship to the supposed exposure, or if the evidence does not suggest an occurrence of potential biologic and public health importance, investigators should prepare a summary report and conclude the investigation. (5)

STEP 3: Determining the feasibility of an etiologic study

Before embarking on an etiologic study to examine the association between an observed cluster and a particular exposure, investigators must determine the epidemiologic and logistical feasibility of a study.

It is essential to first construct a testable hypothesis based on the investigation so far and knowledge of the relevant literature. The hypothesis should be clearly stated and should include the target population, health event(s), and exposure(s) of interest.

Factors to consider when determining feasibility include:

- The pros and cons of different study designs, including costs, and ways in which study results can be used.
- Potential challenges (e.g., sample size, use of already-identified case-patients, comparison group selection) and ways to address them.
- The potential for finding additional cases, expanding the case definition, and using a different or larger geographic area or time period.
- Ways in which additional data can be collected and the costs of that. Data collection will be driven by the hypothesized association between the health event and exposure, and might require clinical and laboratory tests, as well as data on other known or suspected causes (e.g., other medical conditions or environmental exposures) as well as the exposure and health event of interest.

Determining how to measure the putative exposure is critical to an etiologic study. Do clinical or environmental tests for the exposure exist, or can such tests be developed? How sensitive are they? Given the lapse of time since exposure, will the test be useful? Is reported exposure history a good predictor of true exposure?

While determining the feasibility of an etiologic study is relatively straightforward, determining whether the benefits of the study will justify the effort is not. Etiologic studies of disease clusters are not likely to be successful unless the disease is extremely rare or the frequency of the disease has suddenly increased. Furthermore, the etiologic agent must be measurable, and must leave a physiologic response in the bodies of those exposed. Additionally, an appropriate unexposed control group is needed, so levels of exposure must vary within the population to be able to carry out the study. (6)

Investigators must assess the epidemiologic and policy implications and likely community reactions to doing the study (or not doing the study). If an etiologic investigation is feasible and affordable, and the likely benefits justify the effort, get to it! However, if an etiologic investigation is logistically impossible, is too expensive, or is not likely to affect existing policies or programs despite results, you'll need to carefully explain why further investigation isn't feasible in a summary report.

STEP 4: Conducting an etiologic investigation

If you have successfully carried out all the steps described above, it's time to proceed with an etiologic study. For such a resource-intensive endeavor to be worthwhile, however, it should generate knowledge about the broader epidemiologic and public health issues raised, not merely explain a specific cluster.

Begin by writing a formal study protocol describing the data collection elements and methods to be used. Lay out the steps in data collection, processing, and quality assurance, and develop an appropriate plan of data analysis. At this point, study design decisions will be unique to the study, and further guidance is outside the scope of this *FOCUS* issue.

Conclusion

Most non-infectious disease clusters result from coincidence and chance, but they can still serve a useful purpose. They allow public health officials to interact with the community, be responsive to public needs, and learn about exposures. Occasionally, cluster investigations allow investigators to develop new hypotheses about previously unsuspected exposure-disease relationships. But they can also be an unproductive drain on scarce public health resources.

The search for knowledge must be balanced with the need to fulfill ongoing public health responsibilities. In many cases, resources would be better invested in lab-based sciences and large epidemiologic studies in which small potential risks and complex biological phenomena can be investigated.

Glossary

Ascertainment: The process of identifying – from existing sources, using distinct case definitions – persons who have a particular health event or disease.

Cluster: An unusual aggregation, real or perceived, of reported health events grouped together in time and space. Even one case of a rare, serious disease may constitute a cluster. Clusters are occasionally detected through routine surveillance data (e.g., registries, death certificates) but are commonly first reported by health care providers, patients, or community members.

Etiologic study: An investigation that deals with the causes or origins of disease.

Excess occurrence: The occurrence of more cases of a disease than expected in a given area or among a specific group of people over a particular period of time. Also called an *epidemic*.

Exposures: Substances that a host ingests, inhales, or absorbs through the skin that might increase risk of disease. Can be man-made or natural, ongoing or past, and may include industrial chemicals, medications, pesticides, and radioactive materials. Usually the exposure has not previously been shown to be associated with the occurrence of the disease being reported. Concerns about an exposure are often the impetus for reporting a cluster.

Latency period: The time between exposure to a disease-causing agent and the onset of disease.

Non-infectious diseases: Health events without an obvious infectious etiology, including cancers, birth defects, chronic diseases (e.g., multiple sclerosis), and adverse events related to drugs or vaccines.

Occurrence rate: The number of cases of a health event in a defined population in a given period of time.

CONTACT US:

The North Carolina Center for Public Health Preparedness
The University of North Carolina at Chapel Hill
Campus Box 8165
Chapel Hill, NC 27599-8165

Phone: 919-843-5561
Fax: 919-843-5563
Email: nccphp@unc.edu

FOCUS Workgroup:

- Lorraine Alexander, DrPH
- Meredith Anderson, MPH
- Lauren N. Bradley, MHS
- Jennifer A. Horney, MA, MPH
- Pia D.M. MacDonald, PhD, MPH
- Gloria C. Mejia, DDS, MPH
- Amy Nelson, PhD, MPH
- Tara P. Rybka, MPH
- Rachel A. Wilfert, MD, MPH

REFERENCES:

1. Lenz W. Kindliche mißbildungen nach medikament-einnahme während der gravidat [Malformations in children after a drug taken during pregnancy]. *Dtsch Med Wochenschr.* 1961;86:2555–2556.
2. Cartwright RA. Cluster investigations: Are they worth it? *Med J Aust*. 1999;171(4):172. http://www.mja.com.au/public/issues/171_4_160899/cartwright/cartwright.html. Accessed August 13, 2008.
3. CDC. Guidelines for investigating clusters of health events. *MMWR Morb Mortal Wkly Rep*. 1990;39(RR-11):1-16. http://www.cdc.gov/mmwr/preview/mmwrhtml/00001797.htm. Accessed August 13, 2008.
4. New Jersey Department of Health and Senior Services and ATSDR. *Childhood Cancer Incidence Health Consultation: A Review and Analysis of Cancer Registry Data, 1979-1995, for Dover Township (Ocean County), New Jersey*. 1997. http://www.state.nj.us/health/eoh/hhazweb/cansumm.pdf. Accessed August 13, 2008.
5. Bender AP, Williams AN, Johnson RA, Jagger HG. Appropriate public health responses to clusters: The art of being responsibly responsive. *Am J Epidemiol.* 1990;132:S48-S52.
6. Rothman KJ. A sobering start for the cluster busters' conference. *Am J Epidemiol* 1990;132:S6-S13.

If you would like to receive electronic copies of FOCUS on Field Epidemiology, please fill out the form below:

- NAME: _____
- DEGREE (S): _____
- AFFILIATION: _____
- E-MAIL ADDRESS: _____
- May we e-mail any of your colleagues? If so, please include their e-mail addresses here:

  _____

  _____

  _____

Please fax to: (919) 919-843-5563

or mail to: North Carolina Center for Public Health Preparedness
The University of North Carolina at Chapel Hill
Campus Box 8165
Chapel Hill, NC 27599-8165

Or go online: http://nccphp.sph.unc.edu/focus/

UPCOMING TOPICS!

- **Basic Surveillance**
- **Advanced Surveillance**

We are on the web!
http://nccphp.sph.unc.edu/