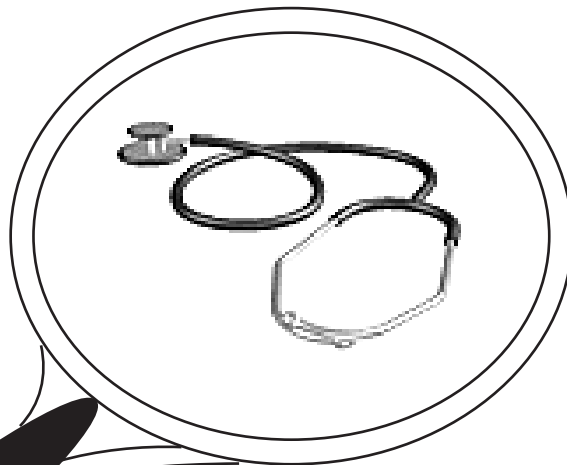


Cluster Busters:

A Game of Disease Mystery Solving

DRAFT



<http://swehsc.pharmacy.arizona.edu/coep/clusterbusters/>

Original Concept & Curriculum Content Developed By: Stefani D. Hines, M.A., M.S.
Website Development By: John King, M.S. and Joanna K Norman, Ph.D

The University of Arizona, Center for Toxicology,
Southwest Environmental Health Sciences Center,
Community Outreach & Education Program, 1999, 2000.

Re v.6/2000

TABLE OF CONTENTS

Section I: Introduction

Directions	1
Setting Description	2
Scenario Description	3
State Health Department Response Protocol for Reports on Disease Clusters	4

Section II: Teacher Guides

Teacher Reference Sheet	5 – 22
Cluster Significance Chart	23
Supplementary Activity: CEPUP “Chance & Probability	24-28

Section III: Student Materials

Interview	29-30
Student Note Sheet	31 – 41
Supplementary Activity: CEPUP “Chance & Probability	42
Response Protocol	43
Cluster Significance Chart	44

DIRECTIONS

1. To begin the game, read the “Setting Description” out loud to the class.
2. Read to the class the “Biography” description for the scenario named after a specific person (in this case Dr. Wanda B. Better).
3. Review the “State Health Department Protocol for Reports on Disease Clusters” with the students.
4. Refer to the “Cluster Details: Teacher Reference Sheet” for the directions related to a specific scenario.
5. Hand out the “Student Note Sheet”

Alternative Approach: You can have the class break-up into groups and run several different scenarios at once. At the end of the game, each group can share their results and you can discuss some of the similarities and differences between the scenarios. If you play multiple scenarios at once you may want someone in each group to read the “Background” description.

SETTING DESCRIPTION

You all are epidemiologists, or scientists that study the incidence, distribution, and control of disease in a population, and you work for the state department of health. You just received a phone call from (Dr. Wanda B. Better) who was reporting a potential disease cluster. A disease cluster is when a specific type of illness occurs in an identified population within a given period of time and the number of people with the illness exceeds the normal disease rate.

Your supervisor has assigned all of you to this case to work as a team. You need to determine if this cluster is real or if it falls within the same incidence rate as the rest of the nation's population (i.e. there is no difference between the people you study and everyone else). At the end of the game you decide if this disease cluster report merits further investigation.

SCENARIO: DR. WANDA B. BETTER

Dr. Wanda B. Better works at the Children's Hospital in Somewhere Town, Ohio. She has worked there for 11 years and noticed that in the past 2 years it seemed there were considerably more cases of infants admitted to the hospital for pulmonary hemorrhaging (bleeding in the lungs) and hemosiderosis (an indicator of repeated hemorrhaging). In order to confirm her suspicions, Dr. Better searched the hospital data base to determine how many infant pulmonary hemorrhages were diagnosed at this hospital over the past 11 years. She found that 12 cases were admitted in the past 2 years alone, compared to only 3 cases in the 10 years previous. Because of the sudden increase in infant pulmonary hemorrhages, Dr. Better decided to report the cases the Ohio State Health Department. Your team will determine if the Center for Disease Control should be contacted and if a full investigation should be conducted.

State Health Department Response Protocol for Reports on Disease Clusters

Adapted from: Fiore, B.J., Hanrahan, L.P., and Anderson, H.A. (1990) State health department response to disease cluster reports: A protocol for investigation. American Journal of Epidemiology, 132, supplement (1), S14-S21.

Use this protocol as a guide for your investigation

Describe the Cluster

Determine geographic locations, type, number of cases, age, time period

Verify the Cases

Obtain medical records of disease cases, death certificates

Risk Assessment

Characterize disease group cases by age and time lived in defined region. Determine comparison population's disease rate (via state, county, zip code, or US census)

Statistically Analyze Disease Rates

Compare the disease rates of the "cluster" vs. the comparison population

Examine Potential Exposure

Investigate environmental or occupational exposures

Determine if the Disease is

Biologically Plausible

Does the type of disease match the exposure?

Determine Cluster Significance

Determine if this investigation merits further action
Refer to the "Cluster Significance Chart"

Report Results

Write a report on the investigation. This is kept on file and used for future reference, especially if a full investigation is done.

Scenario: Dr. Wanda B. Better

TEACHER REFERENCE SHEET

SCENARIO: DR. WANDA B. BETTER

Teacher Reference Sheet

Dr. Wanda B. Better

<http://swehsc.pharmacy.arizona.edu/coep/clusterbusters/>

I. DESCRIBE THE CLUSTER & VERIFY THE CASES

Conduct Phone Interviews:

1. Assign students to be an interviewer or interviewee (1 interviewee -- Dr. Better, 1 interviewer – Health Department Official).
2. Hand out a copy of the “Phone Interview Question/Answer Guide” to the interviewee and interviewer
3. Hand out interview sections of the “Student Note Sheets” to the rest of the students so they can take notes during the interview.

COMMENTS:

- You may want to have select students conduct the interview in front of the class or divide the students into groups of 3-4 and have them conduct the interviews within the group.
- The phone call is initiated by Dr. Better when she decides to report the possible disease cluster.
- Students can refer to the “Phone Interview Question/Answer Guide” or you can have them brainstorm to determine what information they need.
- Students need to take detailed notes of the phone interview. They can use their Cluster Buster “Student Note Sheet” to help them.

Note: This information is incomplete because it is *verbal* and based on recollection. It provides a beginning point from which to start the investigation. A more complete picture occurs when the students *Verify the Cases*. Verification is important because medical records are often more accurate than people’s memory.

4. Review with the students the major points of the Interview.

Interview Summary

Type of Disease: **Pulmonary Hemorrhage**

Total Number of Disease Cases Reported: **15 (12 in the past 2 yrs)**

Approximate Ages of People with Disease: **Infants**

Possible Cluster Locations: **Eastern Somewhere Town**

Other Information from the Interview: - **Use similar or same techniques to diagnose**

- **Has not been a significant increase in the population served**
- **Most of the cases were male**
- **Most of the cases lived with a family member who smoked**
- **5 cases re-occurred after the infant went back into the home**

5. Discussion questions

- a. Does a difference in 12 cases in 2 years compared to 3 cases in the previous 10 years sound suspicious?
Yes. There is probably a statistical significance between these numbers (we need to know what the disease incidence rate is to know for certain). But, because there is such a large difference in number of cases in a small time frame it calls for attention.
- b. What are some possible factors that might account for such an increase?
An environmental agent, a sudden increase in the population served, the ability to diagnose a disease.
- c. How does an increase in the number of people served potentially affect the apparent disease incidence?
You increase the likelihood of seeing a disease simply because you are seeing more patients.
- d. How do changes in the ability to diagnose a disease potentially affect the disease incidence rate?
If the ability to diagnose a disease becomes easier (and cheaper), then that technology becomes used more often and the diagnosis for a specific disease will increase. Conversely, if two very similar diseases were often "lumped" together in terms of diagnosis, the disease incidence rate for the "lumped" disease would go down.

Verify the Cases & Define the Cluster

- Hand out the Case Summary Table and have the students verify the cases by viewing the hospital records (on the website) and recording information into the Case Summary Table.

Case Summary Table

Case #	Date Years Ago	Address	Age	Sex	Etiology	Other Notes	Check Similar Cases
1	11	1700 City Blvd.	12 wks	F	Sudden Infant Death Syndrome (SIDS)	Died	
2	7	5900 Somewhere Ave.	4 wks.	M	Unknown, Possible environmental exposure		
3	3	Someplace Else (not in Somewhere Town)	3 yrs	F	Unknown, immune mediated?	Admitted 3 times previously for pulmonary hemorrhage	
4	2	4233 E. 33 rd Ave 202	8 wks	M	Unknown, Possible environmental exposure	Readmitted after 3 days, died	✓
5	2	5100 E. 25 th Apt. 2	4 wks	M	Unknown		✓
6	2	1500 Ohio Rd.	24 wks	M	Milk Allergy		
7	2	3016 Midwest Blvd A10	5 wks	M	Unknown, Possible environmental exposure		✓
8	1	4233 E. 33 rd #210	10 wks	M	Unknown	Readmitted after 2 weeks	✓
9	1	5404 E. 33 rd	12 wks	M	Unknown, smoking?	Previous condition febrile seizure, readmitted after 5 days	✓
10	1	5400 E. 33 rd	6 wks	M	Possible environmental exposure	Previous condition dehydration	✓
11	1	5100 E 25 th #52	6 wks	M	Unknown, Possible environmental exposure	Readmitted after 1 month	✓
12	1	5503 E. 25 th E3	8 wks	F	Unknown		✓

Scenario: Dr. Wanda B. Better

13	1	5000 Quick Ave.	6 yrs	F	Group A Streptococci	Fever, died	
14	1	3006 Midwest Blvd #42	9 wks	M	Unknown, possible exposure to toxin – cocaine?		✓
15	1	4800 Industrial Ave 2130	15 wks	F	Unknown	Previous conditions low birth weight, bacterial infection	✓

- Discuss with the students the similarities and differences between the cases. This discussion will help them determine which cases belong in the cluster investigation and which do not. The students should place a check mark next to the cases they identify as “similar” in the “Case Summary Table”

Similar cases help to narrow and define the cluster.

Based on age, the year the illness occurred, and the etiology (which is unknown/possible environmental exposure) cases that are similar include: 4, 5, 7, 8, 9, 10, 11, 12, 14, 15

Note: In the real cluster all of the infants were black, socio-economic are believed to be the “link”, not race susceptibility of Pulmonary Hemosiderosis, therefore race is not included in this scenario.

Map the Case Locations:

- Hand out the map and the “Map the Case Locations” page of the Student Note Sheet. Have the students map all of the addresses obtained from the hospital. This is an additional step to defining the cluster and the region for the investigation.
- Have the students highlight on the map the cases they identified above as being similar (based on age, year of illness, and the etiology).
- Ask the students how they might define the region for the investigation (also called a “cell”) based on what they see visually on the map. (In this case the cells are partitioned using zip codes. Which zip codes to use are easily defined here because of the clear visual “cluster” of cases).

Note: The hospital service boundary could potentially be used as the investigation boundary or “cell” for the case, however since it is a children’s hospital the service boundary covers all of Eastern Ohio – thus it is too large. Census tracts, water delivery boundaries, groundwater or contamination boundaries, or even specific buildings are also commonly used as investigation boundaries depending on the case.

Scenario: Dr. Wanda B. Better

11. Map the defined region: The students should highlight or outline the zip codes containing the majority of pulmonary hemorrhage cases.

Zip Codes for the Defined Region: 12303, 12304

Define the Time Period

12. Once the students have identified the potential cluster ask them what they would recommend as a time period to focus on for their investigation: **2 years**.

Two years is the time period for the investigation (as opposed to 1 because of the sudden increase in cases beginning 2 years ago and the patients' location within the "cluster"). However, the students' could argue to limit the investigation to one year since the majority of cases occurred within a one year period of time. The time period is based on the actual time period used for the "real" epidemiological investigation.

Note: Disease clusters are generally defined by time and location. Having a specific time frame helps to focus the investigation and helps determine any causative environmental agents. [CDC. Acute pulmonary hemorrhage/hemosiderosis among infants – Cleveland, January 1993 to November 1994. MMWR; 43:881-3.] And [CDC. Update: pulmonary hemorrhage/hemosiderosis among infants – Cleveland, Ohio, 1993-1996. MMWR; 46; 33-35.]

13. Have the students list the cases that will be included in the cluster investigation. Inform them that the rest of the investigation will only focus on the "cluster cases."

Cases: 4, 5, 7, 8, 9, 10, 11, 12, 14, 15

II. RISK ASSESSMENT

1. Hand out the Risk Assessment pages of the Student Note Sheets
2. Find the average age and age range (from the medical records).

Average Age: 8.3 weeks Age Range 4-15 weeks

Note: Age may play an important role in this particular type of hemorrhaging with this specific etiology.

3. Go to the website Housing & Urban Development office to obtain the length of time the patient lived at the location when the disease struck (include from time of conception).

Length of time: Since gestation

4. Define the populations you want to compare to the “cluster” population

- *You may want to ask the students why they need comparison populations, which populations they would select and why.*

Note: Comparison populations are needed in order to determine if a deviation from the “normal” disease rate exists. A number by itself does not tell you anything; but a number placed in the context of other numbers provides information.

5. Go to the website to obtain the disease incidence rate information (found in the Ohio State Health Department). Find the average disease incidence rate for Idiopathic Pulmonary Hemosiderosis per year for the defined region and the comparison population(s).

a. Disease Incidence Rate for the Nation

- **8.6 cases per 10^6 infants (<1 yr) per year**

b. Disease Incidence Rate for the Defined Region: (Zip codes 12303, 12304)

- # of births: **6600**
- Incidence of disease over 2 years: **10 (2 years ago=3 cases, 1 year ago=7 cases)**
- Average disease incidence rate per year: **10/6600 in 2 years \Rightarrow 5/3300 per year**

Note: Because of the very low incidence of idiopathic pulmonary hemosiderosis (IPH) we only have one comparison population (the nation). In cases such as this, we could also use other nation’s statistics. For example, the incidence of infant IPH in Sweden is 2.4 cases per 10^7 children per year. With other more prevalent diseases the city, county, or state can be used for comparison populations.

6. Next the students need to figure out how they can compare the various disease incidence rates. Ask the students if they can tell just by looking at the numbers if 5 out of 3300 infants is more, less, or the same as the national incidence rate of 8.6 per 10^6 infants.

- a. First, they must convert the numbers to the same ratio. You may want to use a denominator of 10^6 as opposed to dividing 5 by 3300, so they can easily compare the Somewhere Town incidence rates to the national incidence rates without dealing with decimals or very small numbers. You may want to encourage the students to figure out how to do this on their own.

Defined Region Cluster Incidence Rate: 5/3300

$$3300x = 10^6 \quad \Rightarrow \quad x = 303$$

303 is the number they need to multiply the numerator and the denominator by to convert their disease incidence rate to a population of 1,000,000 ($303 \times 5 = 1515$ & $303 \times 3300 \approx 1,000,000$)

Adjusted Cluster Disease Incidence Rate: 1,515/10⁶

Note: If you want to extend this lesson further, you can discuss rounding errors and significant figures (because $3300 \times 303 = 999,900$).

7. Ask the students if it looks like there is a difference between the disease incidence rates for the nation and the cluster. How many times greater is the cluster rate?

The cluster disease incidence rate is **176 times** greater than the national disease incidence rate. This is a highly significant difference, which indicates this is a disease cluster. However, in order to determine if this cluster merits a full investigation the students still need to examine potential exposures.

Note: A statistical analysis is typically performed to determine if the cluster disease rate is truly different from the national disease rate. In this case we are saving the statistics activities for some data analysis for the case-control study to help identify a potential environmental agent for this disease.

Summary

Average age: **8.3 weeks**

Age range: **4-15 weeks**

Length of time patients lived in same location: **Since conception**

Comparison populations for Idiopathic Pulmonary Hemosiderosis:

Nation: **8.6 cases per 10⁶ infants (<1 yr) per year**

Disease rate for the defined area: **5/3300**

Adjusted disease rate for the defined area: **1,515/10⁶**

III. EXAMINE POTENTIAL EXPOSURE & STATISTICAL ANALYSIS

1. Hand out the Student Note Sheets titled Examine Potential Exposure & Statistical Analysis.
2. Go to the website and click on the hospital to find “potential” and “known” causes for pulmonary hemorrhaging and hemosiderosis.

Summary:

“Known” Causes: Goodpasture’s Syndrome; Immune reaction to cow milk protein; Cardiac and vascular problems; Very low birth weight; Perinatal cocaine exposure; Smoking tobacco.

“Potential” Environmental Causes: Pesticide exposure (very weak epidemiological evidence in children); Toxigenic molds/fungi such as *Stachybotrys spp.*, containing mycotoxins shown to cause bloody nasal discharge and cough in adults, causes hemorrhaging in very young animals especially males, no epidemiological links to children exposed. *Stachybotrys* is very uncommon it is found in <1% of the homes in the U.S.

Note: The cluster cases were diagnosed as “idiopathic,” which means the etiology (or the cause for the disease) is unknown. Therefore, the students will have to make some “guesses” and set up a case-control study to help them identify possible causes.

3. Go to the website and click on the Environmental Quality Office to find the town’s environmental assessment reports to begin assessment of environmental exposures.
4. Map potential problem locations

Summary:

Locations:

3310 S. Somewhere Ave. – Southeast 1.5 miles
6955 S. Somewhere Ave. – waste released downstream from town
33rd Ave. and Midwest Blvd. – soil contamination immediate area, vapors ¼ mile area,
2626 E. 25th Ave. – waste injected below potable water table

Types of contaminants:

Dichloromethane
copper, lead and zinc
pesticides
fertilizers

5. Go to the website to research the water quality supplied to Somewhere Town (Environmental Quality Office)

Summary:

- There are no identified chemical exposures for the Somewhere Town water supply

6. Discuss the results of the general environmental exposure investigation with the students.

Summary:

- Case #7 is near a site where there is a potential for exposure to air-borne pesticides. However, it is a very weak link due to the proximity to the site (it is just outside or on the edge of the exposure ring where concentration will be much lower).
- Since the general environmental exposure investigation did not reveal any strong toxicant exposures from the general environment, the next step is to focus the investigation into the patients' homes.

7. Discuss the next step with the students, which is to set up a case-control study looking at the patients' home environments compared with other infants. You may want to ask the following questions to stimulate thought and discussion.

Discussion Questions:

- a. Based on the information you have about known causes for pulmonary hemosiderosis, what do you think should be investigated to try and identify the potential cause(s) for this disease cluster?
 - *Cocaine exposure during gestation; immune responses to cows milk; pesticide exposure; mold exposure; cardiac/vascular history; exposure to tobacco smoke.*
- b. A case-control study means that you investigate two populations, the population with the disease (the "cases") and a comparative "control" population. Why do you need a control population?

- The case-control study is done to help identify potential environmental factors and/or toxicants that may cause or contribute to pulmonary hemorrhaging and hemosiderosis. As with any research or investigation, you want a "control" or comparison population to

help you identify a deviation from the norm. The control group needs to be as similar to the case group as possible. This eliminates variables that can complicate the study and make it difficult to identify any causative agent(s).

Case-Control Study Results

A. Discuss with the students the “matching” or “control” parameters for the study.

TABLE 1: Matching for Case and Control Groups

	Case Infants	Control Infants
Infant Ages	Matched w/in 2 weeks of birth	
Mean Maternal Age	21.2 years	24.3 years
Mean Maternal Education	11.4 years	11.5 years
Receives Medicaid Assistance	80%	83%
Average Age of Home	76 years	75 years

B. Ask the students what they would include in the interview.

Note: In the real investigation the study participants were asked over 200 questions. Table 2 includes only a few important parameters.

TABLE 2: Interview Survey Results for Exposure Assessment

Type of Exposure	CASE INFANTS N = 10		CONTROL INFANTS N = 20	
	Yes	No	Yes	No
Tobacco smoke	9	1	11	9
Cocaine during pregnancy	1	9	1	19
Breast feeding	0	10	7	13
Lived in same dwelling since conception to time of illness	10	0	18	2

TABLE 3: Comparison Blood & Urine Test Results

(significantly different at the P<0.05 level)

	Case Infants (Mean results)	Control Infants (Mean results)
Hematocrit	24.0%	32.6%
Hemoglobin	8.3 g/dL	11.2 g/dL
Red Blood Cell Count	3.5 million/mm ³	4.2 million/mm ³
Hemolysis	Yes	No

Milk-specific IgE	Negative	Higher levels
Glucose-6-phosphate dehydrogenase	Negative	Negative
Sickle Cell	Negative	One positive
Pesticides (organophosphate, carbamate, pyrethrins)	Below detection limits	Below detection limits
Volatile Organic Compounds (VOCs)	Below detection limits	Below detection limits

C. Discuss the environmental assessment results from the case and control infants' homes. Ask the students if they can see any trends in Table 4 (the numbers look higher in the case homes – this is a prelude to statistics).

TABLE 4: Environmental Assessment Results within the Homes

TEST/INSPECTION	CASE INFANTS N = 10	CONTROL INFANTS N = 20	
Water damage – visual inspection	10	5	
INDOOR AIR QUALITY (in patient's bedroom or room where majority of patient's time is spent)			
Pesticides (organophosphate, carbamate, pyrethrins)	trace amounts of chlorpyrifos, diazinin, & other carbamate or pyrethrin compounds found in 3 homes	trace amounts of chlorpyrifos, diazinin, & other carbamate or pyrethrin compounds found in 6 homes	
Volatile Organic Compounds (VOCs)	Below detection limits	Below detection limits	
Fungi – Air Samples <i>Stachybotrys atra</i> Units: Colony Forming Units (CFU)/m ³	Case 1: 33	C1: 0	C11: 30
	Case 2: 46	C2: 0	C12: 0
	Case 3: 43	C3: 0	C13: 0
	Case 4: 57	C4: 23	C14: 0
	Case 5: 36	C5: 0	C15: 0
	Case 6: 45	C6: 0	C16: 0
	Case 7: 49	C7: 0	C17: 0
	Case 8: 32	C8: 0	C18: 0
	Case 9: 0	C9: 0	C19: 0
	Case 10: 49	C10: 36	C20: 0
Fungi – Surface Samples <i>Stachybotrys atra</i> Units: Colony Forming Units (CFU)/m ³	Case 1: 1.6×10^7	C1: 0	C11: 3.2×10^4
	Case 2: 1.7×10^7	C2: 0	C12: 0
	Case 3: 2.1×10^7	C3: 0	C13: 0
	Case 4: 2.3×10^7	C4: 3.1×10^4	C14: 0
	Case 5: 1.7×10^7	C5: 0	C15: 0

	Case 6: 2.0×10^7	C6: 0	C16: 0
	Case 7: 2.2×10^7	C7: 0	C17: 0
	Case 8: 1.8×10^7	C8: 3.0×10^4	C18: 2.5×10^4
	Case 9: 0	C9: 0	C19: 0
	Case 10: 2.5×10^7	C10: 2.4×10^4	C20: 0

Fungi – Air Samples <i>Cladosporium</i> Units: Colony Forming Units (CFU)/m ³	Case 1:	1339	C1:	0	C11:	88
	Case 2:	1400	C2:	0	C12:	0
	Case 3:	1555	C3:	103	C13:	0
	Case 4:	1222	C4:	66	C14:	0
	Case 5:	1756	C5:	0	C15:	93
	Case 6:	1544	C6:	0	C16:	0
	Case 7:	1222	C7:	0	C17:	0
	Case 8:	1652	C8:	0	C18:	0
	Case 9:	1199	C9:	195	C19:	0
	Case 10:	1433	C10:	0	C20:	0
Fungi – Surface Samples <i>Cladosporium</i> Units: Colony Forming Units (CFU)/m ³	Case 1:	1.0×10^6	C1:	1.5×10^6	C11:	1.0×10^6
	Case 2:	1.7×10^6	C2:	3.1×10^6	C12:	3.9×10^6
	Case 3:	2.1×10^6	C3:	1.1×10^6	C13:	1.0×10^7
	Case 4:	5.0×10^6	C4:	5.2×10^6	C14:	4.4×10^6
	Case 5:	1.7×10^6	C5:	3.6×10^6	C15:	2.7×10^6
	Case 6:	2.0×10^6	C6:	3.9×10^6	C16:	5.4×10^6
	Case 7:	2.8×10^6	C7:	4.1×10^6	C17:	2.0×10^6
	Case 8:	1.8×10^6	C8:	3.3×10^6	C18:	7.0×10^6
	Case 9:	5.0×10^5	C9:	3.5×10^6	C19:	3.8×10^6
	Case 10:	2.5×10^6	C10:	3.6×10^6	C20:	3.7×10^6

Data Analysis & Statistics Activities

1. Explain to the students that scientists cannot just look at numbers and say that one number looks *significantly* different from another number (the word *significant* is used to imply a meaningful difference).

For example, the chances of finding 10 CFU/m³ of a specific fungi may be the same as finding 20 CFU/m³ of that same fungi. Thus, 10 would NOT be significantly different than 20 and the difference between these two numbers is due to *random* events. In order to determine if 10 is significantly different from 20, scientists would need to take a large number of samples, and determine the *average* number of CFU/m³ found within most households. Lets say that after taking many air sample in many homes, the scientists found that the average CFU/m³ for a specific type of fungi was 15 ± 5 CFU/m³. This means that a normal range would be between 10 and 20 CFU/m³. Since 10 and 20 are within the range, the two numbers would NOT be significantly different from each other.

2. Also explain that scientists need to “reduce” the data into something that has meaning and is easy to work with. It is too hard to look at a bunch of different numbers and decide what they mean. Ask the students if they would rather compare 2 numbers or 200 numbers? It is easier to compare 2 numbers. So scientists “reduce” the number of numbers in order to make meaningful comparisons.

3. Have the students “reduce” some of the data “by hand” to help them understand the process. They will determine the *average*, the *deviation*, and the *standard deviation* for a small set of the air fungi samples. The students can fill in the following tables (blank tables with instructions are included on the “Student Note Sheet”) to guide them:

The *standard deviation* measures how closely data are clustered about the mean and is represented by the equation:

$$S = \sqrt{\sum(x_i - x)^2 / n - 1}$$

where x_i is an individual result, x is the mean, and n is sample number.

The table below is on the Student Note Sheet to guide them through the process step-by-step.

Summary

Years (=sample#)	Number of Cases	Deviation (cases – average)	Deviation Squared (deviation) ²
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
	Sum = <u>Average =</u>		Sum of deviations ² =

Standard Deviation [sum of deviations/(n-1)]:
(n = # of samples or years for the investigation)

You may want to ask the students what happens to the standard deviation when n gets bigger and then ask them if this is good or bad? Having a small standard deviation is good because you can better tell if your result is different from the norm. (See extension exercises below).

- Using EXCEL, find the *average* and *standard deviation* for the remaining fungi air samples in the Case and Control homes. Ask the students if they would want to “hand reduce” all of the data in Table 4. Explain that computers and calculators can do most of the work.

NOTES ON ENTERING FORMULAS INTO EXCEL:

- To obtain the average go to INSERT \Rightarrow FUNCTION \Rightarrow select AVERAGE \Rightarrow enter the first and last cell numbers you want to average separated by a colon. In the first “case” average box that would be (C4:C13). In the first “control” average box it would include the two columns separated by a comma (G4:G13, E4:E13).
- To obtain the standard deviation go to INSERT \Rightarrow FUNCTION \Rightarrow select STDEV \Rightarrow enter the first and last cell numbers you want to average separated by a colon. In the first “case” average box that would be (C4:C13). In the first “control” average box it would include the two columns separated by a comma (G4:G13, E4:E13).

Note: Excel writes exponents (1.3×10^6) as 1.3E+06.

Summary:

Stachybotrys atra air samples

	Case Data	Control Data
Average	39	4.5
Standard Deviation	± 15.8	± 11.1

Are the case and control data significantly different from one another? **Yes**

Stachybotrys atra surface samples

	Case Data	Control Data
Average	1.79×10^7	7.1×10^3
Standard Deviation	$\pm 6.94 \times 10^6$	$\pm 1.3 \times 10^4$

Are the case and control data significantly different from one another? **Yes**

Cladosporium spp. air samples

	Case Data	Control Data
Average	1432.2	27.3
Standard Deviation	± 192.4	± 53.6

Are the case and control data significantly different from one another? **Yes**

Cladosporium spp. surface samples

	Case Data	Control Data
Average	2.1×10^6	3.9×10^6
Standard Deviation	$\pm 1.2 \times 10^6$	$\pm 2.1 \times 10^6$

Are the case and control data significantly different from one another? **No**

Note: In real epidemiological studies the statistics are much more complex. You may consider incorporating additional lessons on statistics at this point.

Note: Here we did not address the *confidence interval* or *p*. The *confidence interval* represents how *confident* you are in your research results. A *p* of less than 0.05 means that you expect at least 95% of future results to fall within the range you determined in your study (e.g. 95% of the time the numbers will fall between 46.2 and 48.2). The *confidence* will generally increase as the number of samples increases. Scientists can plug numbers obtained from a preliminary study into a special equation to determine how many samples they will need to take to get within a certain confidence level.

EXTENSION

E1. Conduct CEPUP Risk Comparison Activity 1 “Chance & Probability” to illustrate the concepts of chance and probability (this activity also demonstrates why they want to take many samples)

E2. Once the students understand how to calculate the standard deviation, you can further explore why taking more samples is better. To illustrate this, have the students divide a deviation of 2 by 1, 3, 8, and 50. The answers are 2, 0.7, 0.25, and 0.04 respectively. Thus the variance decreases as the number of samples increase. Scientists want a small variance because it increases their certainty in the conclusions they draw from the data. For example, if the comparison population is 8 ± 2 (range of 6 to 10) and the result is 10.1, that is a very “borderline” result and the researcher cannot make any strong conclusions. However, if the comparison population is 8 ± 0.7 , the researcher can more confidently conclude that 10.1 is significantly different than 8. This is because the smaller the variance, or the larger number of samples, the more representative the sample is of the whole population.

Note: You can also have the students explore how the variance changes with the range and clustering of numbers. The variance will decrease as the range, or number of outliers, decreases. For example, the variance will be smaller if the range is 43-47 as opposed to 40-55. Generally, the variance will also decrease if the numbers are clustered. For example, let's say we have two sets of numbers with the same range of 40-55, but the distribution of numbers is:

Set 1: 40, 41, 42, 42, 42, 42, 55

Set 2: 40, 43, 47, 47, 53, 54, 55

The average & variance for Set 1 is: 43 ± 23

The average & variance for Set 2 is: 48 ± 29

VII. DETERMINE CLUSTER SIGNIFICANCE

1. Answer the following questions:
 - a. Does the cluster disease incidence rate significantly exceed the comparison population incidence rate? **Yes**
 - b. Was the disease population exposed to any known environmental or occupational contaminants? **Yes**
 - c. If an exposure did occur, is there a biological link between the disease and the contaminant? **Possibly**

Note:

2. Refer to the "Cluster Significance Chart" to determine if further action is needed.

Summary:

- High Disease Rate: Yes
- Documented Exposure: Yes
- Biologic Plausibility: Possibly
- Further Action: Yes

VIII. REPORT RESULTS

1. Students can write a report summarizing their findings and conclusions. Be sure to include the following information:
 - a. Total # of cluster cases: **10**
 - b. Time Period Investigated: **2 years**
 - c. Type of disease: Idiopathic **Pulmonary Hemosiderosis**
 - d. Age range: **4-15 weeks**
 - e. How Long People with Disease Have Lived in Area: **Since conception**
 - f. Comparison of disease rates of the cluster vs. comparison population: **The cluster rate is 176 times greater than the national disease rate.**
 - g. Potential exposures, Case-control study results & biological plausibility: **Pesticides are a potential exposure for several cases but it is unlikely that they are the cause of the hemorrhaging due to the low concentrations and the weak epidemiological links. However, pesticides may be a confounding factor along with cigarette smoke (9 of the 10 infants lived in homes with smokers). The most probable cause of the hemorrhaging is from a toxigenic mold species, in this case probably *Stachybotrys atra*. Even though at the time of the investigation there were no direct epidemiological links to infant pulmonary hemorrhaging, there is evidence that this mold species causes such health effects in mice and reduced effects in adult humans.**
 - h. Recommendations: **a) Recommend removing the infants from the homes with water damage and high fungi counts; b) Recommend further studies to investigate the toxicology of *Stachybotrys atra* and several other mold species that may be toxigenic.**

Note: The study that this scenario was based on was the first in a series of epidemiological studies to link *Stachybotrys atra* (and possibly some other toxigenic mold species) to infant pulmonary hemosiderosis.

Cluster Significance Chart

Use this chart to help you determine if further action is needed for your investigation. If you answer “yes” to more than one of the three criteria (high disease rate, documented exposure, or biologic plausibility), further action is probably needed. You may personally draw a different conclusion, which is fine, just be sure to justify your decision.

High Disease Rate +	Documented Exposure	+ Biologic Plausibility	= Further Action
Yes	Yes	Yes	Yes
Yes	No	Yes	Yes
No	Yes	Yes	Yes
Yes	Yes	No	Yes
No	No	No	No
Yes	No	No	No
No	Yes	No	No

Adapted from: Fiore, B.J., Hanrahan, L.P., and Anderson, H.A. (1990) State health department response to disease cluster reports: A protocol for investigation. American Journal of Epidemiology, 132, supplement (1), S14-S21.

Student Materials

Dr. Wanda B. Better

Phone Interview

Ohio State Health Department Official

Dr. Better places the phone call. The conversation begins with the health department official taking the call.

1. Good morning, Ohio State Health Department this is _____.
2. An increase of how much over what period of time?
3. Ok, that does sound suspicious. Have there been any recent medical advances that would allow for a better diagnosis of pulmonary hemorrhaging compared with previous years?
4. Has there been a sudden increase in the number of people in the region your hospital serves?
5. Dr. Better, have you noticed any similarities between the cases?
6. Really....I'm going to need access to the medical records to verify these cases and look at the addresses. Can you copy the medical records of all infant pulmonary hemorrhaging cases in your hospital over the past 11 years?
7. Did any of these cases result in death of the infant?
8. Were there any other recurrences after the infant was discharged?
9. OK. I'll arrange to come by the hospital to view the medical records. I'll also keep you informed about the status of the investigation. That's all the questions I have now. Thank you for help.

Phone Interview

Dr. Wanda B. Better

(The health department official begins the dialogue)

1. Good morning, my name is Wanda B. Better and I am a doctor over at the Somewhere Town Children's Hospital. I have noticed a sudden increase in infants being admitted to the emergency room for pulmonary hemorrhaging and thought I should report it.
2. I checked our hospital records and found that 12 cases were admitted in the past 2 years alone. Whereas, in the ten years prior we only have 3 cases recorded.
3. That's a great question and I wish that were the case. However, we basically still use the same techniques to diagnose the illness as we did 10 years ago.
4. Since we are a children's hospital we've served all of Eastern Ohio for at least the past 15 years. Though the area's grown, we still see about the same number of children per year.
5. Yes. Of the 12 cases in the past 2 years, nine were male, the majority of cases seemed to come from the East side of town, and most of them lived in a home where a family member smoked.
6. Sure. I'll have those ready within a couple of days.
7. Two of the 12 cases that occurred in the past 2 years. One of those was due to a reoccurrence after the infant was discharged and went back into the home.
8. Yes. A total of 5 including the infant who died.
9. Thank you.

CLUSTER BUSTERS : Student Note Sheet

CASE: Dr. Wanda B. Better

<http://swehsc.pharmacy.arizona.edu/coep/clusterbusters/>

I. DESCRIBE THE CLUSTER & VERIFY THE CASES:

Determine geographic locations, type, number of cases, ages, and time period.

Obtain medical records of disease cases and death certificates.

(You can obtain this information from the introduction, the interviews, your teacher, and the website)

INTERVIEW NOTES: Interview with Dr. Better

Type of Disease (be specific):

Total Number of Disease Cases Reported:

Approximate Ages of People with Disease:

Possible Cluster Locations:

Other Information from the Interview:

Verify the Cases & Define the Cluster:

Summarize the information you find in the hospital records (website). After you have completed the table, check mark the cases that have similarities (you may want to consider age, address, date, and etiology).

Case Summary Table

Case #	Date	Address	Age	Sex	Etiology	Other Notes	Check Similar Cases

Map the Case Locations: Map addresses from the case summary table.

Define Cluster Region – Based on what you see, how might you define the cluster region?

[Outline the defined region with a highlighter or marker]

Cluster Details:

Based on your map and Case Summary Table, list the case numbers to be included in the “cluster” investigation:

What is the defined time period for your investigation? (This is based on the cases)



II. RISK ASSESSMENT

Characterize the disease group cases by age and time lived in the defined region. Determine the disease rate for comparison populations (e.g. state, county, zip code, and/or US)

(This information you obtain from the website)
(You may need to calculate some of the averages yourself)

Average Ages (for defined time period):

Age Ranges (for defined time period):

Range of Time Patients Lived at the Location When the Illness Occurred (include gestation time, i.e. 9 months + age):

Identify Some Comparison Populations You Can Use. Briefly Explain Why You Selected Those Populations:

Compare the disease rates of the “cluster” versus the comparison population to determine if you have a real “cluster.”

(This information you obtain from the website)

1. Find the disease incidence rates for the nation and the defined cluster region.
 - a. Disease Incidence Rate for the Nation:

III. EXAMINE POTENTIAL EXPOSURE & STATISTICAL ANALYSIS

1. List “potential” and “known” causes for pulmonary hemorrhaging and hemosiderosis.

2. Define “idiopathic”

3. Go to the website and click on the Environmental Quality Office to find the town’s environmental assessment reports to begin assessment of environmental exposures.
 - a. List locations of potential concern and the contaminant(s)

 - b. Map potential problem locations.

4. Describe the water quality supplied to Somewhere Town.

5. Were any of your defined cluster cases exposed or potentially exposed to any general environmental contaminants (i.e. generated outside of the home)?

Data Analysis & Statistics Activities

1. Calculate the *Average* and the *Standard Deviation* for the Fungi Air Sample for *Stachybotrys atra*. Use the table below to help guide you.

Year (=sample#)	Number of Cases	Deviation (cases – average)	Deviation Squared (deviation) ²
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
	Sum = Average =		Sum of deviations ² =

Standard Deviation [sum of deviations/(n-1)]:
(n = the number of samples or years)

Are the case and control data significantly different from one another?
(You determine this by adding & subtracting the standard deviation from your sample averages and see if the ranges of the two samples overlap. If they do not overlap then they are different from one another).

2. Using EXCEL, find the *average* and the *standard deviation* for the rest of the fungi data and summarize your results below.

Stachybotrys atra surface samples

	<u>CASE DATA</u>	Control Data
Average		
Standard Deviation	±	±

Are the case and control data significantly different from one another?

Cladosporium spp. air samples

	<u>CASE DATA</u>	Control Data
Average		
Standard Deviation	±	±

Are the case and control data significantly different from one another?

Cladosporium spp. surface samples

	<u>CASE DATA</u>	Control Data
Average		
Standard Deviation	±	±

Are the case and control data significantly different from one another?

IV. DETERMINE CLUSTER SIGNIFICANCE

Determine if this investigation merits further action.

1. Answer the following questions:

- a. Does the cluster disease incidence rate significantly exceed the comparison population incidence rate?
- b. Was the disease population exposed to any known environmental or occupational contaminants?
- c. If an exposure did occur, is there a biological link between the disease and the contaminant?

2. Refer to your answers to the above questions and to the “Cluster Significance Chart” to determine if further action is recommended.

- a. High Disease Rate (Yes or No):
- b. Documented Exposure (Yes or No):
- c. Biologic Plausibility (Yes or No):
- d. Further Action Recommended (Yes or No):

V. REPORT RESULTS (Summary)

1. Write your conclusions about the presence of a disease cluster and your recommendations for further action. Justify your conclusions with the information and data you collected.

State Health Department Response Protocol for Reports on Disease Clusters

Adapted from: Fiore, B.J., Hanrahan, L.P., and Anderson, H.A. (1990) State health department response to disease cluster reports: A protocol for investigation. American Journal of Epidemiology, 132, supplement (1), S14-S21.

Use this protocol as a guide for your investigation

Describe the Cluster

Determine geographic locations, type, number of cases, age, time period

Verify the Cases

Obtain medical records of disease cases, death certificates

Risk Assessment

Characterize disease group cases by age and time lived in defined region. Determine comparison population's disease rate (via state, county, zip code, or US census)

Statistically Analyze Disease Rates

Compare the disease rates of the "cluster" vs. the comparison population

Examine Potential Exposure

Investigate environmental or occupational exposures

Determine if the Disease is

Biologically Plausible

Does the type of disease match the exposure?

Determine Cluster Significance

Determine if this investigation merits further action
Refer to the "Cluster Significance Chart"

Report Results

Write a report on the investigation. This is kept on file and used for future reference, especially if a full investigation is done.

Cluster Significance Chart

Use this chart to help you determine if further action is needed for your investigation. If you answer “yes” to more than one of the three criteria (high disease rate, documented exposure, or biologic plausibility), further action is probably needed. You may personally draw a different conclusion, which is fine, just be sure to justify your decision.

High Disease Rate +	Documented Exposure	+ Biologic Plausibility	= Further Action
Yes	Yes	Yes	Yes
Yes	No	Yes	Yes
No	Yes	Yes	Yes
Yes	Yes	No	Yes
No	No	No	No
Yes	No	No	No
No	Yes	No	No

Adapted from: Fiore, B.J., Hanrahan, L.P., and Anderson, H.A. (1990) State health department response to disease cluster reports: A protocol for investigation. American Journal of Epidemiology, 132, supplement (1), S14-S21.

Vocabulary

References: Dorland, W.A.N., The American Illustrated Medical Dictionary, 21st Edition, W.B. Saunders Company, 1949.

Webster's II New Riverside University Dictionary, The Riverside Publishing Company, 1988.

1. acute hemothysis – Severe bleeding for a short duration.
2. anemia – A condition in which the blood is deficient either in quantity or in quality.
3. bilateral pulmonary infiltrates – Foreign material deposited in both lungs.
4. bronchoalveolar lavage - Washing of the broncheal tube and lung air sac.
5. cyanotic – Blueness of the skin.
6. dyspnea – Difficult or labored breathing.
7. endotracheal intubation – Insertion of a tube into the trachea to open breathing passage.
8. erythrocytes – red blood cells
9. etiology – The study or theory of the causation of a disease
10. febrile seizure – seizure associated with a high fever
11. hemoglobinuria – The presence of hemoglobin in the urine.
12. hemosiderosis – A condition when hemosiderin (an iron containing pigment in blood cells) is deposited in the tissues due to the destruction of blood cells. It is generally in indication of chronic, rather than acute, hemorrhaging.
13. hypernatremic – Excessive amount of sodium in the blood.
14. idiopathic – of unknown causation.
15. jaundice – A syndrome characterized by deposition of bile pigment in the skin and mucous membranes resulting in a yellow appearance.
16. pallor – extreme or abnormal paleness.
17. pleural effusions – bloody fluid in one or both of the membranous sacs which line either side of the thoracic cavity and envelopes the adjacent lung.
18. pneumonia- an acute or chronic disease caused by viruses, bacteria, or physical and chemical agents and characterized by inflammation of the lungs.
19. pneumothorax – accumulation of air or gas in the pleural cavity, occurring as a result of injury or disease.
20. polycythemia – a condition characterized by an abnormally large number of red cells in the blood.
21. pulmonary hemorrhage – Bleeding in the lungs.
22. trachea – Tube extending from the larynx to the bronchi and carrying air to the lungs.
23. sickle cell anemia – A disease marked by anemia and by ulcers and characterized by the red blood cells of the patient a sickle-like or crescent shape. This disease is hereditary and appears to be confined to the black population.
24. sudden infant death syndrome – The unexpected death of an apparently healthy infant that usually occurs during the first four months of life while the infant is sleeping.