Research Articles

Field Test of an Epidemiology Curriculum for Middle School Students

Mark A. Kaelin, Wendy W. Huebner, Mark J. Nicolich, and Maudellyn L. Kimbrough

ABSTRACT

The purpose of this study was to test the effectiveness of a middle school epidemiology curriculum called Detectives in the Classroom. The curriculum presents epidemiology as the science of public health, using health-related issues that capture the interest of young students and help prepare them to make evidence-based health-related decisions. The curriculum was field tested among seventh-grade urban students using a quasi-experimental design. Analysis of covariance of pre- and post-test scores examined five outcomes, including students' perceptions of their abilities in science as inquiry, scientific literacy, and knowledge about five enduring epidemiologic understandings; their self-reported interest in science, and assessment of students' epidemiological reasoning ability. The 378 experimental students, compared to 620 controls, had generally higher post-test improvements in epidemiology-related outcomes and smaller increases in the other measures. A dose-response was suggested by higher scores among students exposed to more than 10 lessons. Strengths of this evaluation include a large sample and availability of data to account for differences in demographic and school performance variables. Limitations of this evaluation include randomization by school as opposed to student, the relatively short-term and generally self-reported outcomes, and inconsistencies in proportion of the curriculum actually taught. The findings offer encouragement about the potential for Detectives in the Classroom to improve students' perceptions of their science abilities and scientific literacy, their interest in science and their abilities in basic epidemiologic reasoning. Further tests of this and other epidemiology curricula are needed to respond to the growing interest in teaching public health science to younger students. And while it is important to test near-term impacts, an additional challenge from a curriculum evaluation standpoint will be to follow students over several years to examine subsequent choices concerning selected courses, college majors, and career paths.

INTRODUCTION

From both an individual and societal perspective, it is important for students to develop scientific literacy. Recent data suggest that schools in the U.S. are not preparing a scientifically literate citizenry. These data are collected by the National Assessment of Educational Progress, under the auspices of the National Center for Education Statistics of the U.S. Department of Education.[1] Representative samples of children at ages 9, 13, and 17 have been tested at least eight times between 1970 and 1999 in areas of science, mathematics, and reading. The testing has tracked students' science achievement by five performance levels since 1977. Although there has been improvement, by 1999, only 47% of 17-year-olds were able to perform at the fourth level ("has some detailed scientific knowledge and can evaluate the appropriateness of scientific procedures"), and only 10% at the fifth level ("can infer relationships and draw conclusions using detailed scientific knowledge").[1]

Scientific literacy has important public health implications. For example, at an individual level, it is reasonable to assume that greater scientific literacy will empower students to make evidence-based decisions about personal health behaviors. This is critical because as students become increas-

Mark Kaelin, EdD, is associate professor, Department of Health and Nutrition Sciences, College of Education and Human Services, Montclair State University, Montclair, NJ 07043; E-mail: kaelinm@mai.montclair.edu. Wendy W. Huebner, MPH, PhD, is senior scientific associate–epidemiology, Epidemiology and Health Surveillance Section, ExxonMobil Biomedical Sciences, Inc., Annandale, NJ 08801-0971. Mark J. Nicolich, PhD, 24 Lakeview Road, Lambertville, New Jersey 08530. Maudellyn L. Kimbrough, is supervisor for Management Informations Services, Paterson Public Schools, Paterson, NJ 07505.

This project is supported by a Science Education Partnership Award, Grant Number R25RR018584, from the National Center for Research Resources, National Institutes of Health.

ingly independent, they will be faced with many more health-related decisions that can profoundly affect their lives. At the societal level, it is reasonable to assume that greater scientific literacy will empower students to be informed participants in the democratic decision-making process concerning public health actions and policy. Greater scientific literacy and knowledge of epidemiology may also enhance students' understanding of public health messages; this is important at the societal level, too, in a world of expanding possibilities for large-scale epidemics and other natural and man-made disasters. Furthermore, better understanding of the science of public health has potential to improve the caliber and numbers of young people entering the public health professions.

A large portion of health-related evidence can be understood through the science of epidemiology, defined as the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to the control of health problems.[2] Teaching epidemiology emphasizes the importance of the scientific methods and challenges students to develop problem-solving and critical thinking skills. It also aligns with the *Benchmarks for Scientific Literacy: Project 2061*[3] and the *National Science Education Standards*,[4] particularly regarding fundamental abilities in science as inquiry.

Historically, courses in epidemiology have been offered primarily to graduate students in health-related fields, although some have advocated teaching epidemiology to undergraduates as a general education course.[5,6] Moreover, recognition of the possibilities of teaching epidemiology to even younger students appears to be increasing. The U.S. Centers for Disease Control and Prevention (CDC) has taken the lead in advocating the integration of teaching epidemiology in grades K–12.[7] A CDC program called EXCITE (Excellence in Curriculum Integration through Teaching Epidemiology) (http://www.cdc.gov/excite) is based on case studies that incorporate key aspects of epidemiology such as quantitative methods and causal reasoning.[8] Additionally, The Robert Wood Johnson Foundation and the College Board have collaborated to develop *Young Epidemiology Scholars* (YES) (http://www.collegeboard.com/yes) competitions and epidemiology teaching units for high school teachers and their students.[9] A primary reason given for the support of the YES competitions is that "... the skills and orientation of epidemiology—critical thinking and a problem solving mentality—are relevant across all sciences."[10] This increased recognition of the value of teaching epidemiology to younger students is further evident by the Epidemiology Education Movement (http://www.Epiedmovement.org), which identifies the "Top 8 Reasons for Teaching/Learning Epidemiology," five of which are directly related to health education.[11]

The authors of this article have developed a curriculum to support the teaching of epidemiology to middle school students. It is designed for use by health, science, and mathematics teachers. *Detectives in the Classroom (Detectives)* is based on pedagogical principles suggested in *Understanding by Design*.[12] The authors explain how effective curricula can be built by identifying "enduring understandings" and "essential questions." Enduring understandings are the big ideas that reside at the heart of a discipline and have lasting value outside the classroom. The essential questions are questions that can be answered when the enduring understandings are achieved. It is the curriculum developer's responsibility to create lessons that develop students' abilities to answer the essential questions and, in doing so, to achieve the enduring understanding.

Detectives consists of 34 lessons, in five modules, each focusing on one of five epidemiologic essential questions and its enduring understanding (Table 1). These have been described in detail in a previous *American Journal of Health Education* article entitled, "Epidemiology, Literacy, and Health Education."[13] In brief, the five essential questions and enduring understandings in Table 1 align, respectively, with the following epidemiologic content: 1) use of descriptive information to generate hypotheses; 2) use of analytic techniques and study designs to investigate associations; 3) evaluation of causality; 4) role of epidemiology in societal decisions about risk and prevention; and 5) assessment of prevention strategy effectiveness.[13] The lessons were shaped to help students "uncover" the epidemiologic principles and apply what they learn to health issues of interest in their personal and public lives. The modules may be taught in their entirety or in selected smaller sets of lessons that could be incorporated into classes of health, science, and/or mathematics.

Detectives was developed in partnership with a multidisciplinary team of teachers and health professionals and was pilot-tested in 2001 in two middle schools. During the 2002–2003 school year, the curriculum was field tested among seventh-grade science students in a New Jersey school district. This paper reports the results of this field test. Using a quasi-experimental design, we tested hypotheses that exposure to the curriculum would improve students' perceptions of their abilities in science as inquiry and scientific literacy, and their interest in science, knowledge about five enduring epidemiologic understandings, and epidemiological reasoning ability.

METHODS

Field Test Population

The New Jersey school district in which *Detectives* was field tested had more than 25,000 students, of whom approximately half were Hispanic, 40% African American, 6% Caucasian, 2.5% Asian/Asian American/Pacific Islander, and less that 1% Native American. Eighty percent of the district's students were eligible for free or reduced-cost lunch, a measure approximating socioeconomic status. The district is an "Abbott District," a school district administered by the state, as per the New Jersey Supreme Court decision in Abbott v. Burke, in order to provide urban education reform initiatives that ensure that public school children, including students with disabilities and students with limited English proficiency from the poorer

Table 1. *Detectives in the Classroom*: Five Essential Questions and Enduring Understandings

Essential Questions	Enduring Understandings
1. How is this disease distributed and what hypotheses might explain that distribution?	Health-related conditions and behaviors are not distributed uniformly in a population. Each has a unique descriptive epidemiology that can be discovered by identifying how it is distributed in a population, in terms of person, place, and time. Descriptive epidemiology provides clues for formulating hypotheses.
2. Is there an association between the hypothesized cause and the disease?	Causal hypotheses can be tested by observing exposures and diseases of people as they go about their daily lives. Information from these observational studies can be used to make and compare rates and identify associations.
3. Is this association causal?	Causation is only one explanation for finding an association between an exposure and a disease. Because observational studies are flawed, other explanations must also be considered.
4. What should be done when preventable causes of disease are found?	When a causal association has been identified, decisions about possible disease prevention strategies are based on more than the scientific evidence. Given competing values, social, economic, and political factors must also be considered.
5. Did the disease prevention strategy work?	The effectiveness of a prevention strategy can be evaluated by making and comparing rates of disease in populations of people who were and were not exposed to the strategy. Costs, trade-offs and alternative strategies must also be considered.

urban districts, receive educational entitlements guaranteed them by the Constitution.[14]

The potential target population for the field test consisted of approximately 2,200 students, in seventh-grade science classes held during the 2002–2003 academic year, in the district's 24 middle schools.

Field Testing Teacher Recruitment, Preparation and Implementation

The first phase of recruitment sought seventh-grade science teachers who would volunteer to attend five teacher preparation workshops, teach 30+ *Detectives* lessons with fidelity, and allow pre- and post-testing of their students. These teachers were randomly assigned to be field testing teachers (Experimental) or controls (Control Group A). Because the pool of "volunteer" teachers might be small, we also recruited a second control group (Control Group B) consisting of teachers who had not volunteered to teach 30+ *Detectives* lessons but did volunteer to allow pre- and post-testing of their students. This second control group increased the number of controls and allowed us to assess potential volunteer bias.

The teachers who were randomly assigned to the field testing group (Experimental) attended five teacher preparation workshops between October 2002 and April 2003. During each workshop, the project director taught the lessons for one module and the teachers discussed how they would teach the lessons and suggested revisions. The teachers returned to their schools and taught the lessons in that module before the next workshop. At the start of workshops two through five, teachers shared their experiences teaching the previous module's lessons. This training schedule was set up exclusively for purposes of the field test and is not intended to suggest a strategy for more widespread implementation.

Measures of Student Performance

During the fall of 2002, project staff administered a pre-test to students in the classes of the experimental and control groups. In the spring of 2003, project staff administered an identical post-test to students in these classes.

The test consisted of 62 items, assessing five types of information and is presented in its entirety in the Appendix. The basis for the 62 items was to address several endpoints that could reasonably be expected to improve with exposure to epidemiology teaching, with regard to perceived abilities in science generally, interest in science generally, and understanding of epidemiology specifically.

A. Students' perceptions of their abilities in each of seven fundamental abilities of science as inquiry, as derived from the National

Table 2. Reliability Testing Results Sample Size and Cronbach's Alpha for Test Subscores

	N	29 items	N	33 Items
Total	39	0.79	31	0.76
Class A	18	0.72	15	0.74
Class B	15	0.82	16	0.76

Science Education Standards Fundamental Abilities in *Science as Inquiry.*[4] A five-point Likert scale from "Definitely No" to "Definitely Yes" was used.

B. Students' perceptions of their abilities in each of six indicators of scientific literacy, derived from the *Benchmarks for Scientific Literacy: Project 2061.*[3] A five-point Likert scale from "Definitely No" to "Definitely Yes" was used.

C. Students' perceived understanding of *Detectives*' five enduring understandings, using a five-point Likert scale of "Definitely Do Not Understand" to "Definitely Understand."

D. Self-assessment of attitudes toward science using a six-point Likert scale from "Strongly Disagree" to "Strongly Agree" from *Attitudes toward Science Inventory.*[15]

E. Eleven questions to test students' epidemiological reasoning ability. These items, developed by the investigators, have common sense aspects to them in order to give students a reasonable chance of answering them correctly without specific training in epidemiology.

Reliability Testing of Instruments

Test-retest reliability of our instrument was determined in two groups of eighth-grade students. Students were tested twice, two weeks apart, as follows: 39 students were tested on 29 test items and 31 other students were tested on the remaining 33 items. (Testing was divided between two classes to reduce the amount of class time required.) A series of Cronbach's alpha coefficients were calculated on the summed item scores of each of the subtests.[16] The alpha was calculated for both classes combined and for each class separately. Only scores with no missing or unanswered items were considered. The range of Cronbach's alpha coefficients was 0.74 to 0.82 (Table 2). These results compare favorably to the cutoff of 0.7 indicating an acceptable reliability coefficient.[17]

Mean change scores for all individual test items were also tested. Among the set of 29 items and the set of 33 items, the mean change between the first and second test was statistically significant for only one item in each set. Furthermore, there was a good balance between positive and negative changes among the items. Mean differences between the first and second test were generally in the range of plus or minus 0.1 to 0.3 (among answers having a range of 1 to 5 or 1 to 6).

Student Information

School district personnel linked computerized information on students' demographics and school performance to our test results. This information included gender, nationality, date of birth, most recent final grades in science, health, and mathematics, standardized test scores in science, math, and language, first spoken language, lunch code (free/reduced lunch), special education status, and number of unexcused absences in the previous school year.

Data Handling and Quality Assurance

Pre- and post-tests were administered via Scantron® bubble sheets, and individual paper and scanned output were checked before and after scanning for completeness and accuracy. Other checks assured completeness of the merge with school data. Data inconsistencies were resolved by an analyst/epidemiologist team.

Analysis Methods

The main analytic approach was an analysis of covariance (ANCOVA). This method of analysis provides a combination of regression and analysis of variance methods. It also uses the appropriate pre-test score as an independent variable so the dependent variable (outcome) is similar to, but more general than, a "change score." The same set of independent variables was used in the final models for all five of the outcomes analyzed. A variable was included in this set of independent variables if it was statistically significant in a preliminary analysis of at least one of the five outcomes (using an order independent analysis). This is a conservative approach. The independent variables used in all final analyses included study group, gender, nationality, first language, final grades in health, mathematics, and science, special education code, unexcused absences, and pre-test score. Variables not included in final models, because they were not statistically significant in preliminary analyses, included binary code for socioeconomic status derived from a school lunch code ("denied" and "not applicable" versus all others), age at pre-test, and standardized test scores in science, mathematics, and language. No further grouping of the variables was done for the analysis and students with any missing values for a specific analysis were excluded from that analysis. All variables, except pre- and post-test scores, were analyzed as distinct groups or classes.

The results were presented as least squares (LS) means, which are the statistically estimated means of the post-test scores that would be seen if all students were identical with respect to all of the independent variables except study group. That is, each student theoretically has the same pre-intervention score, the same bilingual code, the same health final grade, and so on, in order to account for the influence of these variables. The LS means are not the same as the observed means, but facilitate comparisons by creating a "level playing field." This approach therefore accounts for other student variability in addition to their differences in pre-test scores.

Groups Analyzed

We identified five groups of students to be considered in final models, including sub-divisions of the Experimental group to facilitate interpretation:

Experimental 1: Students of a single teacher who had participated in earlier *Detectives* pilot-tests and had more experience with *Detectives* than the other field testing teachers

Experimental 2: Students of other teachers who taught more than 10 *Detectives* lessons

Experimental 3: Students of other teachers who taught 10 or fewer *Detectives* lessons

As noted earlier, there were two control groups:

Control Group A: Students of teachers who volunteered to field test the curriculum and were randomly assigned to the control group

Control Group B: Students of teachers who did not volunteer to field test the curriculum but later agreed to allow their students to complete the pre- and post-tests

Human Subjects

The study protocol was approved by the Montclair State University Institutional Review Board. It received an expedited review because the evaluation protocol was within limits of normal educational experience and because of minimal risk to subjects. The requirement of informed consent was waived. The participating school district's Curriculum Committee approved implementation of the field test to "address the New Jersey Core Curriculum Content Standards," on February 13, 2002.

Each student's pre- and post-test results were linked by a unique Student Identification Number. Information on student characteristics was considered private and confidentiality was protected in all stages of data handling, analysis, and reporting. Access to the data was limited to those directly involved in the research project. After completion of the post-tests, the data file was sent to the school district, where demographic and school performance characteristics were added and personal identifiers (name and student identification number) were removed before sending the file back to the investigators. Thus, the final analysis database contained no information that could lead to the identity of individual students. All analyses and reporting were done in the aggregate, further protecting students' privacy.

RESULTS

Participation

Nine seventh-grade teachers from nine schools volunteered to participate in field testing. To ensure that a reasonably large number of students received the curriculum intervention, we randomly designated seven of the nine volunteer teachers as Experimental teachers. The other two teachers were assigned to Control Group A. One of the Experimental teachers left the school district during the study and the affected classes were excluded from the analysis.

Recruitment of additional "non-volunteer" controls resulted in eight Control Group B teachers and their seventh grade science students.

Overall, teachers/students in 16 of the 24 district schools were included in the evaluation. Among the 16 schools, each had participants of only one type (Experimental or Control). Some participating schools had seventh grade teachers/students who did not participate.

Among 2,192 students in the school district enrolled in the seventh grade during the academic year, 998 were in the final study group. The 1,194 non-participants included 727 students in classes of non-participating teachers, the 47 students in the two classes of the experimental teacher who left the school district, and 420 students of participating teachers who did not take both the pre- and post-test.

Figure 1 further illustrates the breakdown of the 998 final participants. For Experimental subjects: 88 were in classes of the more experienced teacher (Experimental 1) and received 16 lessons; 197 were in classes of the other teachers who taught 16–18 lessons (Experimental 2); and 93 had other teachers who taught 6–10 lessons (Experimental 3). Control A and B had 134 and 486 participants, respectively.

Participant Characteristics

Distributions of demographic and school-related characteristics among the five study groups are presented in Table 3. For completeness, all variables considered in the preliminary analyses are included in Table 3, with double asterisks (*) for those that were included in the final models. As the table shows, distributions in age and gender are similar between groups. Notable differences are seen for several other variables; the ANCOVA analyses took these differences into consideration (Table 3).

Findings

Table 4 and Figure 2 display the main analysis that generated least squares (LS) mean scores for the five mutually exclusive study groups, for each of the five outcomes under study. Among the 998 participating students, the number participating in each analysis varies because of missing data. Resulting sample count was particularly low (n=687) for analysis of *Attitudes toward Science Inventory* because there were many more questions to be potentially missed in this segment (33 of the total 62 items in the instrument). All statistical tests in Table 4 are at the $p<0.05$ level.

In Table 4, results are presented in the order hypothesized if the *Detectives* curriculum is associated with improved post-test scores. Thus, among the three experimental groups, students of the more experienced teacher would do best, followed by students who received more than 10 lessons, followed by students who received 10 or fewer lessons. Both control groups would not be expected to do as well as any of the experimental groups, and students in volunteer Control A might be expected to do better than those taught by teachers in the "non-volunteer" Control B (because of possible volunteer bias). The study results tend to support this hypothesis—all the analyses showed some differences among the study groups and if the study groups are ranked by LS mean scores, most are consistent with the hypothesized order.

For *Enduring Understandings* items, the scores for all experimental groups are statistically significantly different from one or both control groups and there is a logical increase in least squares mean scores with increasing exposure to the lessons. There is a 20% increase in mean scores between the

Table 3. Study Participant Characteristics

	TOTAL N=998		Experimental 1 N=88		Experimental 2 N=93		Experimental 3 N=197		Control A N=134		Control B N=486	
	Count	(%)*	Count	(%)*	Count	(%)*	Count	(%)*	Count	(%)*	Count	(%)*
Gender **												
Female	530	(53)	50	(57)	48	(52)	110	(56)	78	(58)	244	(50)
Male	468	(47)	38	(43)	45	(48)	87	(44)	56	(42)	242	(50)
Age (years)												
11	51	(5)	10	(11)	7	(8)	12	(6)	4	(3)	18	(4)
12	641	(64)	61	(69)	59	(63)	125	(63)	99	(74)	297	(61)
13	260	(26)	14	(16)	22	(24)	54	(27)	29	(22)	141	(29)
14	44	(4)	3	(3)	5	(5)	6	(3)	2	(1)	28	(6)
15	2	(<1)	0		0		0		0	-	2	(<1)
Standardized Language Test Score ‡												
Missing	263		20		21		37		54		131	
Partial	501	(68)	31	(46)	38	(53)	142	(89)	53	(66)	237	(67)
Proficient	229	(31)	36	(53)	34	(47)	18	(11)	26	(32)	115	(32)
Advanced	5	(<1)	1	(1)	0		0	1	(1)	3	(<1)	
Standardized Mathematics Test Score ‡												
Missing	262	21		21		36		54		130		
Partial	461	(63)	27	(40)	33	(46)	136	(84)	45	(56)	220	(62)
Proficient	235	(32)	30	(45)	36	(50)	25	(16)	33	(41)	111	(31)
Advanced	40	(5)	10	(15)	3	(4)	0		2	(2)	25	(7)
Standardized Science Test Score‡												
Missing	260		20		21		35		54		130	
Partial	280	(38)	15	(22)	23	(32)	94	(58)	25	(31)	123	(35)
Proficient	393	(53	43	(63)	39	(54)	64	(40)	47	(59)	200	(56)
Advanced	65	(9)	10	(15)	10	(15)	4	(2)	8	(10)	33	(9)
Health Final Grade **												
A	152	(15)	20	(23)	14	(14)	18	(9)	12	(9)	88	(18)
B	322	(32)	29	(33)	42	(45)	45	(23)	63	(47)	143	(29)
C	307	(31)	34	(39)	28	(30)	77	(39)	18	(13)	150	(31)
D	121	(12)	5	(6)	8	(9)	45	(23)	1	(<1)	62	(13)
F	42	(4)	0		1	(1)	12	(6)	0		29	(6)
Other	54	(5)	0		0		0		40	(30)	14	(3)
Science Final Grade **												
A	108	(11)	13	(15)	9	(10)	1	(<1)	27	(20)	58	(12)
B	234	(23)	27	(31)	27	(29)	36	(18)	40	(30)	104	(21)
C	321	(32)	26	(30)	30	(32)	86	(44)	47	(35)	132	(27)
D	243	(24)	19	(22)	16	(17)	57	(29)	13	(10)	138	(28)
F	80	(8)	3	(3)	11	(12)	17	(9)	6	(4)	43	(9)
Other	12	(1)	0		0		0		1	(<1)	11	(2)

lowest and highest group (Control B versus Experimental 1) and Experimental 1 has a considerably higher score than Experimental 2 and Experimental 3.

Knowledge of Epidemiology has statistically significant increases among Experimental 1 and Experimental 2 compared to all controls. These mean score increases are in the 15–20% range. As logic might dictate, the score for the group with 10 or fewer lessons (Experimental 3) falls between controls and the more exposed experimental groups.

For two indices, the *Attitudes toward Science Inventory* and *Science as Inquiry*, the

Table 3. Study Participant Characteristics (con't)

	TOTAL N=998		Experimental 1 N=88		Experimental 2 N=93		Experimental 3 N=197		Control A N=134		Control B N=486	
	Count	(%)*	Count	(%)*	Count	(%)*	Count	(%)*	Count	(%)*	Count	(%)*
Math Final Grade**												
A	107	(11)	7	(8)	11	(12)	9	(5)	15	(11)	65	(13)
B	200	(20)	18	(20)	22	(24)	23	(12)	34	(25)	103	(21)
C	298	(30)	19	(22)	31	(33)	56	(28)	40	(30)	152	(31)
D	237	(24)	16	(18)	20	(22)	71	(36)	33	(25)	97	(20)
F	136	(14)	28	(32)	9	(10)	28	(14)	11	(8)	60	(12)
Other	20	(2)	0		0		10	(5)	1	(<1)	9	(2)
First Language**												
All Other	56	(6)	27	(31)	0	4	(2)	6	(4)	19	(4)	
English	595	(60)	24	(27)	64	(69)	154	(78)	96	(72)	257	(53)
Spanish	347	(35)	37	(42)	29	(31)	39	(20)	32	(24)	210	(43)
Lunch Code												
Missing	74		0	13		33		2		26		
Denied	63	(7)	8	(9)	2	(2)	9	(5)	13	(10)	31	(7)
Direct Cert	183	(20)	10	(11)	16	(20)	62	(38)	18	(14)	77	(17)
Free	517	(56)	52	(59)	47	(59)	68	(41)	64	(48)	286	(62)
Not Applicable	44	(5)	6	(7)	0	5	(3)	13	(10)	20	(4)	
Reduced	111	(12)	12	(14)	14	(18)	19	(12)	24	(18)	42	(9)
Temp Free	6	(<1)	0		1	(1)	1	(1)	0		4	(<1)
Nationality **												
Asian/Pacific	24	(2)	2	(2)	1	(1)	2	(1)	5	(4)	14	(3)
African American	370	(37)	7	(8)	25	(27)	125	(63)	55	(41)	158	(32)
Caucasian	70	(7)	32	(36)	6	(6)	2	(1)	11	(8)	19	(4)
Hispanic	534	(54)	47	(53)	61	(66)	68	(35)	63	(47.0)	295	(61)
Special Ed Status **												
No	880	(88)	75	(85)	85	(91)	173	(88)	119	(89)	428	(88)
Yes	118	(12)	13	(15)	8	(9)	24	(12)	15	(11	58	(12)
Unexcused Absences (Days) *												
Missing	37		6		2		3		5		21	
0 – 5	350	(36)	51	(62)	34	(37)	58	(30)	56	(43)	151	(32)
6 – 10	275	(29)	19	(23)	29	(32)	54	(28)	31	(24)	142	(31)
11 – 20	267	(28)	9	(11)	19	(21)	70	(36)	36	(28)	133	(29)
21+	69	(7)	3	(4)	9	(10)	12	(6)	6	(5)	39	(8)

* All percentages are calculated based on total records present; so, for variables with missing records, denominators do not contain all subjects in subgroup
** Used in Model
‡ From the 4th grade Elementary school Proficiency Assessment; categorizations pre-determined by test score ranges

only significant difference among experimental students compared to controls is for the group with the more experienced teacher (Experimental 1). For the *Attitudes toward Science Inventory* only, experimental students who received 10 or fewer lessons had a LS mean score that was statistically significantly lower than those who had more than 10 lessons, with a 4% difference between the two.

Results for *Scientific Literacy* show a statistically significant difference between students with the more experienced teacher (Experimental 1) versus both control groups, and between Experimental 2 and 3

Table 4. Least Squares Means ‡

Outcome / Study Group +	Least Squares Means
Enduring Understandings 5 Items, n= 929	
Experimental 1	21.09 [1]
Experimental 2 (>10 lessons)	18.68 [2]
Experimental 3 (10 lessons	18.14 [2]
Control A	17.94
Control B	17.54
Knowledge of Epi 11 items, n=997	
Experimental 1	4.88 [1]
Experimental 2 (>10 lessons)	4.97 [1]
Experimental 3 (10 lessons	4.43
Control A	4.17
Control B	4.18
Attitudes toward Science Inventory 33 items, n=687	
Experimental 1	150.47 [1]
Experimental 2 (>10 lessons)	146.88
Experimental 3 (10 lessons	141.42 3
Control A	142.57
Control B	142.95
Science as Inquiry 7 items, n=962	
Experimental 1	28.09 [1]
Experimental 2 (>10 lessons)	27.18
Experimental 3 (10 lessons	27.13
Control A	26.53
Control B	26.68
Scientific Literacy 6 items, n=948	
Experimental 1	24.12 [1]
Experimental 2 (>10 lessons)	23.28 [2]
Experimental 3 (10 lessons	23.14 [2]
Control A	22.75
Control B	22.40

‡ Mean post-test scores, accounting for differences in pre-test scores and other variables

+ Experimental 1: More experienced teacher
Experimental 2: Taught more than 10 lessons
Experimental 3: Taught ten or fewer lessons
Control A: Volunteers who were selected to be controls
Control B: Non-volunteers who later agreed to participate as control classes

1 Indicates statistically different ($p<0.05$) from Control Groups A and B
2 Indicates statistically different ($p<0.05$) from Control Group B
3 Indicates statistically different ($p<0.05$) from Experimental Group 2

versus Control B only. Differences between mean scores among experimental versus control groups are in the 4–6% range.

Details in the model output show that overall, the two strongest influences on the post-test outcome were pre-test score and being in one of the experimental groups. For the most part, other model factors were not influential. However, occasionally other factors had an influence on the post-test outcome. These included the effects of health final grade and gender on the *Attitudes toward Science Inventory*, science final grade on perceived abilities in *Science as Inquiry*, science final grade and nationality on perceived *Scientific Literacy*, and math final grade on *Knowledge of Epidemiology*.

Field test teachers reported anecdotally that students in their classes had a favorable reaction to the *Detectives* lessons. They reported that students were actively engaged, particularly in lessons in which they were asked to work in groups to create materials or generate data to present to the class.

DISCUSSION

To evaluate whether a new curriculum effects change in students requires two elements: measurement of change over time and a determination of whether the change was due to the curriculum. We attempted to do this with a quasi-experimental design, pre- and post-tests, and a comparison of the scores of students who did and did not receive the curriculum. Our effort was enhanced by the willingness of the school district to provide student information regarding demographic and school performance characteristics. Accounting for these variables helped increase the validity of our results.

We hypothesized that exposure to a new middle school epidemiology curriculum would improve students' perceptions of their fundamental abilities in science as inquiry and scientific literacy, their interest in science, perceived knowledge about five enduring epidemiologic understandings, and their epidemiological reasoning ability. Analyses of these measures showed varying degrees of support for these hypotheses.

The greater percentage increases (up to 20%) in scores for the students in the experimental groups versus the control groups were seen in two areas: 1) perceived knowledge about five enduring epidemio-

Figure 1. Student Participation

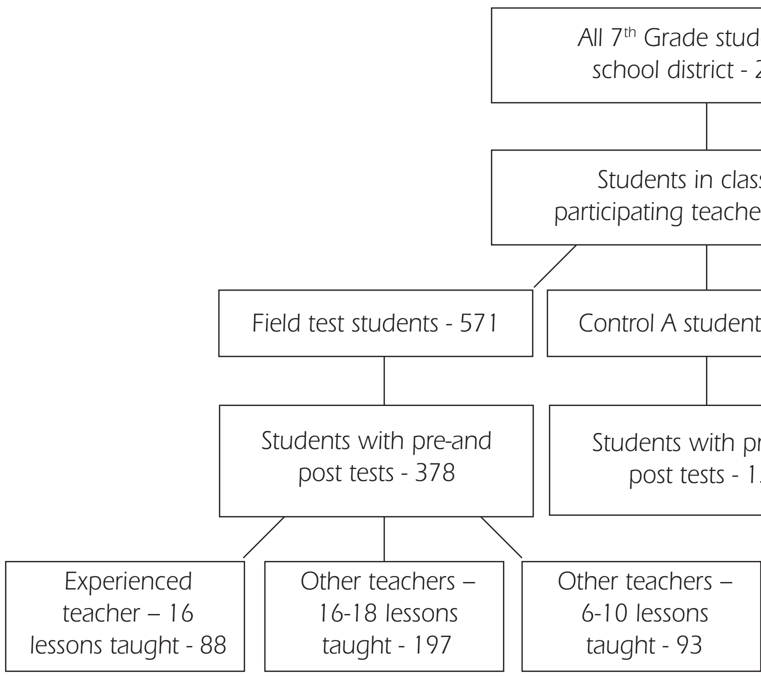
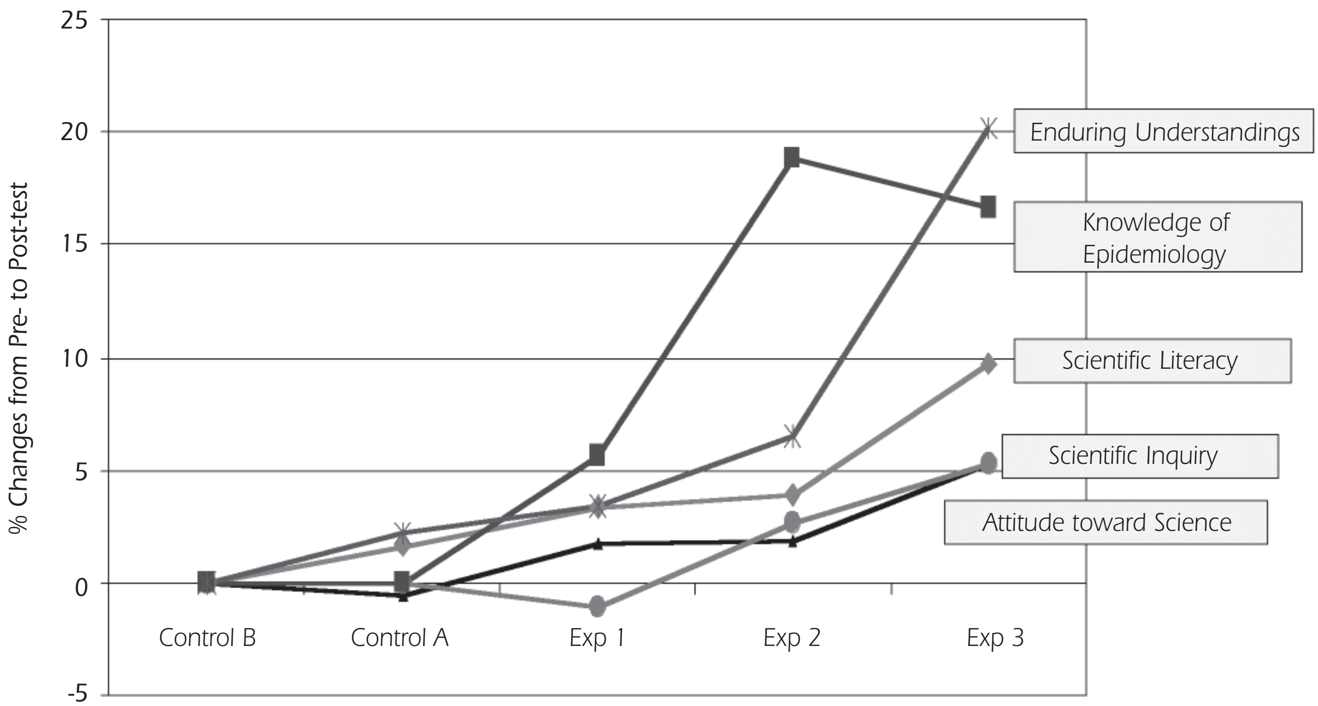


Figure 2. Percent Change at Post-test Compared to Control B Group



logic understandings; and 2) epidemiological reasoning ability. We note that these measures are more directly related to the curriculum itself. The other measures showed increases of 4–5% for experimental students. Overall, results suggest dose-response patterns between experimental versus control students, and among subgroups of experimental students with the more experienced teacher and/or exposure to 10 or more lessons. However, none of the differences are of great magnitude.

To help interpret these findings, we have considered several methodological and logistic factors including fidelity to teaching the curriculum, differential impact of the curriculum, selection issues, and measurement of outcomes. These are discussed below with consideration of their strengths and limitations:

Fidelity to Teaching the Curriculum: Of the 34 lessons, the number of lessons taught ranged from 6 to 18. Given our finding of higher scores among students exposed to more than 10 lessons, exposure to a greater proportion of the curriculum might arguably produce improvements of greater magnitude, but this needs to be tested. Evaluation of our curriculum's effectiveness in the classroom posed a number of practical and logistic challenges, not the least of which was the myriad of time pressures on middle school teachers. Although our findings are suggestive, we conclude that to more rigorously test the effectiveness of an epidemiology curriculum, more attention must be given to selecting a venue in which sufficient time can be devoted to teaching the entire curriculum.

Differential Impact of Curriculum: Measures of students' perceptions of their fundamental abilities in science as inquiry and scientific literacy and their self-reported interest in science are important first steps in the long-term development of science students and a scientifically literate public. We would expect that such outcomes would show less change compared to outcomes more closely related to our curriculum, because there are other curricular efforts aimed at these outcomes and because exposure to our curriculum was brief. The findings of this study are consistent with this expectation.

It is important to emphasize that measurement of student *perception* of abilities in science as inquiry and scientific literacy is not the same as measurement of *improvement* in actual abilities. Improved perception of abilities probably aligns with increasing confidence; other studies are needed to examine how this might be related to actual improvement.

We also note that our study is limited to short-range outcomes; the study of longer-term outcomes is of greater ultimate importance and would require additional time and more sophisticated tracking methods to follow students through high school and beyond to see, for example, what science courses and careers are chosen.

Selection Issues: We had one set of volunteer control teachers (Control A) who we can assume are reasonably comparable to volunteer experimental teachers. The second group of controls (Control B) is much less of a "volunteer" group, so there is potential for volunteer bias. However, results for Control Groups A and B are similar which tends to argue against this bias. In addition, with the small number of volunteer teachers, we were not able to achieve a true randomization process and completely control for self-selection. Furthermore, the randomization was, in effect, by school—no single school had both experimental and control participants. Thus, unmeasured differences between schools could have impacted our results in either direction. Another selection issue is the fact that a large proportion of students (420) only took one of the two tests. This extent of absence from class and/or school at time of testing showed no particular pattern, was not anticipated, and illustrates one of the logistical problems of a large field test in a school setting. It introduces another source of uncertainty regarding our results, because these 420 may have been different from the students who took both tests.

Measurement of Outcomes: Within the constraints of short-term evaluations of curricula, we have assessed various measures of content knowledge, problem-solving ability, confidence about science as inquiry, and science literacy. Instruments to measure such factors are not widely available, especially as related to epidemiology. We developed instruments that have adequate reliability as assessed by the Cronbach's alpha. The instruments have face validity in that the items directly relate to measures of attitude or content knowledge and problem-solving ability. But overall, more discipline-specific and validated instruments need to be developed if epidemiology curricula are to be more thoroughly tested.

Broadly speaking, one hopes that curriculum assessment will answer the question, "Did the curriculum make a difference?" This, of course, is not possible to answer in a single study of a curriculum, or over a short follow-up period. On the plus side, our results are in a positive direction and some suggest a dose-response. However, improvements are mixed and of relatively small magnitude. Furthermore, an assessment of such short-term effects cannot reveal if any benefits "stick" or what other benefits might result. The crucial questions are about students' long-term interests and abilities as measured by increases in taking science classes and in science majors/careers and by some measure(s) of overall science-based gains.

CONCLUSIONS

No individual study can conclusively determine the effectiveness of a new curriculum. This field test of an epidemiology curriculum among middle school students offers encouragement of its potential to improve students' perceived abilities in science as inquiry and scientific literacy, self-reported science attitudes and their awareness and understanding of basic epidemiologic concepts and epidemiologic reasoning. Improvement in these areas has potential to help students make evidence-based personal health decisions and better understand public health messages and, ultimately, to contribute to a more enlightened society in determining public health actions and policies.

Considering the increasing interest in teaching public health science to younger students, it is important to continue testing this and other epidemiology curricula in real classroom situations. Field tests can assess receptivity of the material and near-term impacts; further tests would benefit from improved consistency in the amount of epidemiology material taught and identification and/or development of more specific evaluation instruments. An additional challenge from a curriculum evaluation standpoint is to follow students over several years to examine subsequent choices concerning selected courses, college majors, and career paths.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of many others to this study.

First, we would like to thank the students with whom we tested the curriculum and who willingly gave us feedback by the expressions on their faces, the enthusiasm with which they raised their hands, the accuracy and sophistication of their comments and questions, and the ways by which they corrected their misconceptions. Evaluating the curriculum would not have been possible without the dedication of the middle school teachers who pilot-tested (Tim Purnell and Lynn Tarant) and field-tested *Detectives in the Classroom* (Debra Falek, Paul Healy, Donna Kerwien, Rosa Kopic, Lucy Mancovich, Mary McClam, John Super, Lynn Tarant, and Veva Tronci). We also are grateful for the consulting and data work performed by Molly Weinburgh, Laura Dunn, and Gail Jorgensen, the reliability testing by Lori Levitt and her students, and for her assistance throughout the project by Anne DeSantis, Project Coordinator. Last but not least, we thank Ada Beth Cutler, Dean of the College of Education and Human Services, Montclair State University and former Director of the New Jersey Network for Educational Renewal, Barbara M'Gonigle, Director of the New Jersey Network for Educational Renewal, and Nestor Collazo and Vicki Madden—District Science Supervisors, for their enthusiasm and support of our plans to test the curriculum in the Paterson, New Jersey school district.

REFERENCES

1. Campbell JR, Hombo CM, Mazzeo J. *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics; 2000.

2. Gordis L. *Epidemiology.* 3rd ed. Philadelphia, Pa: WB Saunders Company; 2004.

3. American Association for the Advancement of Science. *Benchmarks for science literacy: Project 2061*. New York: Oxford University Press, Inc.; 1993.

4. National Research Council. *National Science Education Standards*. Washington, D.C.: National Academies Press; 1996:145–148.

5. Fraser D. Epidemiology as a liberal art. *N Engl J Med.* 1987;309–314.

6. Cowan MM. Epidemiology: A vehicle for thinking critically. *J Excel Coll Teach.* 1997;41–62.

7. CDC launches effort to teach epidemiology from kindergarten through high school. *Epidemiology Monitor.* 1997;1:4–5.

8. Excite: Excellence in Curriculum Integration through Teaching Epidemiology. Available at: http://www.cdc.gov/excite/. Accessed September 7, 2005.

9. Young Epidemiology Scholars. Epidemiology teaching units for the high school classroom. Available at: http://www.collegeboard.com/yes/ft/iu/home.html. Accessed September 7, 2005.

10. New national epidemiology prize competitions being created for high school students and teachers. *Epidemiology Monitor* (2002). Available at: http://www.epimonitor.net/EpidemiologyMonitor/Index.htm. Accessed November 21, 2006.

11. Promoters of Teaching Epidemiology to 6–12th Graders Discuss Implementation Challenges at APHA Session. *Epidemiology Monitor.* Available at: http://www.epimonitor.net/EpidemiologyMonitor/Index.htm. Accessed November 21, 2006.

12. Wiggins G, McTighe J. *Understanding by Design*. Alexandria, VA: Association for Supervision and Curriculum Development; 1998.

13. Kaelin MA, Huebner WW. Epidemiology, health literacy, and health education. *Am J Health Educ.* 2002; 33: 362–364.

14. New Jersey Department of Education. Improving Learning and Literacy in Abbott Classrooms. Available at: www.nj.gov/njded/abbotts. Accessed September 7, 2005.

15. Weinburgh MH, Steele D. The modified attitudes toward science inventory: Developing an instrument to be used with fifth grade urban students. *J of Women Minorities in Sci Engineering.* 2000; 6: 87–94.

16. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 16: 297–334.

17. Nunnally JC, Bernstein I. *Psychometric Theory*. New York: McGraw-Hill. 1978.

Pre-Test Instrument

A. Scientific Inquiry

Directions: Please read each statement carefully. Use the following scale, fill in the circle on the answer sheet that best describes the degree to which you think you can do that part of science.

DEFINITELY YES	PROBABLY YES	NOT SURE	PROBABLY NO	DEFINITELY NO
①	②	③	④	⑤

1. I can identify questions that can be answered by scientific study.

2. I can use appropriate tools and techniques to gather, analyze, and interpret data.

3. I can develop descriptions, explanations, and predictions based on scientific evidence.

4. I can think critically and logically about the relationship between evidence and explanations for that evidence.

5. I can recognize and analyze different explanations and predictions for the same observation.

6. I can talk about scientific procedures and explanations with others.

7. I can use mathematics to conduct and analyze the results of a scientific investigation.

B. Scientific Literacy

Directions: Please read each statement carefully. Use the following scale, fill in the circle on the answer sheet that best describes the degree to which you think you can communicate about that part of science.

DEFINITELY YES	PROBABLY YES	NOT SURE	PROBABLY NO	DEFINITELY NO
①	②	③	④	⑤

8. I can ask or find answers to questions based on my curiosity about everyday experiences.

9. I can describe, explain, and predict natural phenomena.

10. I can read and understand articles about science from newspapers and magazines and talk to my friends and family about what I think of the article's conclusions.

11. I can identify the scientific issues that are the basis for national and local decisions and explain my position based on what I know about science and technology.

12. I can evaluate the quality of scientific information based on the methods used to generate it.

13. I can make and evaluate arguments based on scientific evidence and come to a conclusion.

C. Enduring Understandings

Directions: Please read each statement carefully. Use the following scale, fill in the circle on the answer sheet that best describes how well you understand that aspect of science.

DEFINITELY UNDERSTAND	PROBABLY UNDERSTAND	NOT SURE	PROBABLY DO NOT UNDERSTAND	DEFINITELY DO NOT UNDERSTAND
①	②	③	④	⑤

14. Clues for formulating hypotheses can be found by describing the way a disease is distributed in a population of people, in terms of person, place and time.

15. Causal hypotheses can be tested by observing exposures and diseases of people as they go about their daily lives. Information from these observational studies can be used to make and compare rates and identify associations.

16. Causation is only one explanation for finding an association between an exposure and a disease. Because observational studies are flawed, other explanations must also be considered.

17. When a causal association has been identified, decisions about possible disease prevention strategies are based on more than the scientific evidence. Given competing values, social, economic, and political factors must also be considered.

18. The effectiveness of a strategy can be evaluated by making and comparing rates of disease in populations of people who were and were not exposed to the strategy. Costs, trade-offs and alternative strategies must also be considered.

D. Attitudes Toward Science Inventory

Directions: The following statements are about the study of science. Please read each statement carefully. Use the following scale to show how much you agree or disagree with each statement and fill in the circles on your answer sheet.

DEFINITELY YES	PROBABLY YES	NOT SURE	PROBABLY NO	DEFINITELY NO
①	②	③	④	⑤

19. Science is useful for the problems of everyday life.

20. Science is something that I enjoy very much.

21. I do not do very well in science.

22. Doing science labs or hands-on activities is fun.

23. I feel comfortable in a science class.

24. There is little need for science in most of today's jobs.

25. Science is easy for me.

26. When I hear the word "science," I have a feeling of dislike.

27. Most people should study some science.

28. I would like to spend less time in school studying science.

29. Sometimes I read ahead in our science book.

30. Science is helpful in understanding today's world.

31. I usually understand what we are talking about in science.

32. I do not like anything about science.

33. No matter how hard I try, I cannot understand science.

34. I feel tense when someone talks to me about science.

35. I often think, "I cannot do this," when a science assignment seems hard.

36. Science is of great importance to a country's development.

37. It is important to know science in order to get a good job.

38. It does not disturb me to do science assignments.

39. I would like a job that does not use any science.

40. I enjoy talking to other people about science.

41. I enjoy watching a science program on television.

42. I am good at working science labs and hands-on activities.

43. You can get along perfectly well in everyday life without science.

44. Working with science upsets me.

45. I remember most of the things I learn in science class.

46. It makes me nervous to even think about doing science.

47. Most of the ideas in science are not very useful.

48. It scares me to have to take a science class.

49. I have a good feeling towards science.

50. Science is one of my favorite subjects.

51. If I do not see how to do a science assignment right away, I never get it.

E. Knowledge of Epidemiology

Directions: Please listen to each question carefully and then fill in the circle on the answer sheet that is the best answer.

52. A study was conducted to examine the relationship between acne and eating chocolate. Two hundred teenagers with acne were in one group and 200 similar teenagers without acne were in the other group. Each subject was asked questions about his/her chocolate eating habits. In the analysis, percentages of teenagers who ate chocolate were calculated for the group with acne and the group without acne. Which of the following study questions is best answered by this design?

    1. Does drinking chocolate milk cause acne?
    2. Is eating chocolate more common among teenagers with acne than those who don't have acne?
    3. Is eating chocolate decreasing among teenagers?
    4. Is there more acne among teenagers than in the past?
    5. Is acne more common among teenagers who eat candy?

53. Which of the following describes the best plan to answer the study question: "Does playing violent video games cause violent behavior?"

    1. Study violent behavior in people who play violent video games.
    2. Study the use of violent video games in people who have violent behavior.
    3. Study changes in violent behavior from 1996–2002.
    4. Compare the amount of violent behavior in people who play violent video games to the amount of violent behavior in people who do not play violent video games.
    5. Study changes in the use of violent video games from 1996–2002.

54. Which of the following tools could you use to collect, analyze, and interpret data in a scientific investigation of an exposure / disease relationship?

    1. Questionnaires
    2. 2x2 tables
    3. Calculations of risks
    4. Calculations of relative risks
    5. All of the above

55. Let us say that a new disease has been identified that produces a swelling of the lips and tongue. The cause of the new disease is not known. At this point, all the reported cases have occurred among lifeguards at pools and beaches, referees at athletic facilities, and traffic police. Given the distribution of the new disease, what do you predict is causing it?

    1. Soap
    2. Toothpaste
    3. Whistles
    4. Chlorinated water
    5. Automobile exhaust

56. A study published in a scientific journal found that teenagers who had acne were three times more likely to be depressed than teenagers without acne. Which of the following may explain why this association was found?

    1. Acne causes depression.
    2. A medication used to treat acne might cause depression.
    3. There might not be an association between acne and depression; it might have been found due to chance.
    4. The researchers may have found their result because of bias in the way the study was done.
    5. All of the above

A study was conducted to test the hypothesis that carrying heavy backpacks causes back pain. 1,000 middle school students who did not have back pain were followed from January through June 2002. At the end of the study, among 100 students who carry heavy backpacks, 25 students developed back pain. Among the 900 students who did not carry heavy backpacks, 50 students developed back pain.

57. How many students carried heavy backpacks and developed back pain?

    1. 25
    2. 50
    3. 75
    4. 100
    5. 850

58. How many students carried heavy backpacks and did not develop back pain?

    1. 25
    2. 50
    3. 75
    4. 100
    5. 850

59. How many students did not carry heavy backpacks and did not develop back pain?

    1. 25
    2. 50
    3. 75
    4. 100
    5. 850

60. Do the data support the hypothesis that carrying a heavy backpacks causes back pain?

    1. Yes
    2. No
    3. Cannot answer based on data given

Town A has a population of 20,000 and Town B has a population of 40,000. In 2001, there were 240 cases of influenza in Town A and 360 cases in Town B.

61. What is the rate of influenza in Town A?

    1. 12 / 1000
    2. 24 / 1000
    3. 80 / 1000
    4. 120 / 1000
    5. 240 / 1000

62. Is the rate of influenza higher in Town A or Town B?

    1. Town A
    2. Town B
    3. The rates of influenza in Town A and Town B are the same.
    4. Cannot answer based on data given