Exploring Databases:

Conducting Authentic Research Using the Smoking Behavior Database

Maureen Munn, PhD, Megan T. Brown, PhD
Jessica Aronson, MEd, Mark Gallivan, MPH

University of Washington Department of
Genome Sciences

and

Katie Van Horne, Hiroki Oura,
Andrew Shouse, PhD and Philip Bell, PhD

University of Washington College of Education

Table of Contents

Appendices

      A. Student Sheets
      B. Research Project Pages
      C. Directions for Research Project Pages
      D. Research Questionnaire
      E. Manual for Database

Introduction

Exploring Databases immerses high school students in conducting authentic research on smoking behavior using a scientific database. Through their participation in the project, students learn about genetics and neuroscience, develop skills in formulating and testing hypotheses and applying scientific reasoning, and use information and communication technology. The project is conducted by the Department of Genome Sciences and the Institute for Science and Mathematics Education at the University of Washington and is funded by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) award, # DRL-0929181, which we gratefully acknowledge.

Exploring Databases builds on a previous science education project, which was funded through a Science Education on Drug Abuse Partnership Award from the National Institute on Drug Abuse. In the previous project, called *StarNet*, students designed and conducted aspects of an epidemiological study of smoking behavior, in conjunction with scientists and teachers (http://gsoutreach.gs.washington.edu/programs/starnet-investigating-smoking-behavior). The smoking behavior research study collected genetic and environmental data from nearly 300 adult smokers and nonsmokers, which were entered into the "smoking behavior database." Over several years, students contributed to this research project in two ways: 1) by writing questions for the questionnaire used to collect demographic and environmental data about subjects, and 2) by genotyping subject DNA at three genetic loci that may affect smoking. In *Exploring Databases*, students use the unmined database to answer their own research questions about factors that affect smoking behavior.

The primary goal of *Exploring Databases* is to help students develop an understanding of what it means to conduct scientific research by involving them in formulating research hypotheses and testing them using the data from the smoking behavior research study. The focus on smoking behavior engages students in considering an area of research that may affect them personally and challenges them to consider social and ethical issues related to human subjects research. This epidemiological study does not follow the experimental design traditionally taught in high school textbooks, and thus it helps to broaden student understanding of how research is conducted and

what kinds of research questions scientists ask. The utilization of extensive data in a database mirrors research done by many scientists and allows students to engage in deeper scientific thinking than is usually captured in laboratory experiences.

In parallel to the way epidemiologists conduct research using a database from an existing research investigation, students will carry out their research in two parts. In Part 1, called Hypothesis Testing, students conduct queries on a limited number of items in the database to test a hypothesis they have previously developed based on their own observations and other research studies. In Part 2, Hypothesis Generation, students are able to "mine the database" to discover patterns within the data and develop new hypotheses that they would want to test in a new research study. The reasons for this approach and the distinctions between these two types of research are discussed in detail throughout the unit.

The *Exploring Databases* curriculum consists of seven lessons, as well as Student Sheets, Research Project Pages, Directions for Research Project Pages, and Final Project Ideas, as shown in Table 1. Together, these materials engage students in doing scientific research by formulating, testing, and generating hypotheses about how genetic and environmental factors influence smoking behavior. In Lesson 1, students learn that they will be conducting a research project as part of a research team, and each lesson helps them to develop the conceptual knowledge and skills to complete their research. Throughout the unit, students are challenged to examine evidence, carry out statistical calculations, evaluate whether their results support or refute their overarching hypothesis, and consider alternate explanations for their results. They discuss their ideas as part of their research team and as a class, at times engaging in the process of scientific argumentation as they discuss and critique each other's research.

All materials related to this curriculum, including the written materials, videos used during the curriculum, videos of lessons being taught in a classroom, and videos of workshop presenters are available on the Exploring Databases Professional Development Website at https://gsoutreach.gs.washington.edu/exda-pd/.

Table 1. *Exploring Databases* **curriculum**

Lesson 1. Why and how do people do science? (40-50 minutes)	Through several interviews with scientists, students learn about the many approaches scientists take to conduct their research.
Lesson 2. Why do some people become smokers and others do not? (100-120 minutes)	Students explore the wide range of smoking behavior and discover factors that influence people's smoking through profiles of actual smokers and nonsmokers.
Lesson 3. How do genes influence smoking behavior? (70-90 minutes)	Students learn how nicotine interacts with the body and discuss what genes might influence variation in smoking behavior
Lesson 4. How can we study genetic and environmental influences on smoking behavior? (70-80 minutes)	Students learn the characteristics of one commonly used epidemiological study design, the case control study; and they learn the details of the smoking behavior case control study.
Lesson 5. Analysis of data in case control studies: the odds ratio (100 minutes)	Students learn how to calculate the odds ratio and determine its statistical significance using the 95% confidence interval. They learn the difference between association and causality and how to apply the criteria for causality in their research.
Lesson 6. Database research: What can we learn from the smoking behavior data? (120-150 minutes)	Students develop and test specific hypotheses using the smoking behavior database. They analyze results and make conclusions about their overarching hypotheses. Students learn the difference between hypothesis testing and hypothesis generation.
Lesson 7. Hypothesis Generation: Exploring the database to propose new studies (100 minutes)	Students analyze many questions and exposures to generate new hypotheses. Students analyze all query results to propose future studies.
Final Project (100-200 minutes)	This section includes suggestions for the final project format, a template for the final project, and samples of a grading rubric and self and peer evaluation tool
Appendices **A. Student Sheets** **B. Research Project Pages** **C. Directions, Research Project Pages** **D. Research Questionnaire** **F. Database Manual**	The Appendices include various resources to guide students through their research and final projects, as well as a manual for accessing the database.

Connections to Six Strands of Science, Washington's Essential Academic Learning Requirements for Science and the National Framework for Science Education

All too frequently, science classes focus on the learning of facts, with little time devoted to the practices of science, including the importance of evidence and the tenuous nature of scientific knowledge. *Exploring Databases* integrates six strands of science learning defined by scientists and science educators convened by the National Research Council of the U. S. National Academies (Bell et al, 2009, Fenichel and Schweingruber, 2010). Although the six strands were originally developed for informal science learning, they are equally applicable to formal science classes. The six strands are:

Strand 1 – Sparking Interest and Excitement

Strand 2 – Understanding Scientific Content and Knowledge

Strand 3 – Engaging in Scientific Reasoning

Strand 4 – Reflecting on Science

Strand 5 – Using the Tools and Language of Science

Strand 6 – Identifying with the Scientific Enterprise

Table 2 describes how each strand is addressed in the *Exploring Databases* curriculum.

The *Exploring Databases* curriculum helps teachers to address many science standards for Washington State, as well as the newly released National Framework for Science Education, as shown in Tables 3 and 4.

Table 2. How does the Exploring Databases curriculum integrate the 6 strands of learning science?

	Strand 1: Sparking Interest and Excitement	Strand 2: Understanding Scientific Content and Knowledge	Strand 3: Engaging in Scientific Reasoning	Strand 4: Reflecting on Science	Strand 5: Using the Tools and Language of Science	Strand 6: Identifying with the Scientific Enterprise
Lesson 1: Why and how people do science?	Students learn that they will be conducting authentic research using a database			Students learn about the research of several scientists, the different approaches these scientists take to answer their research questions, and the social/collaborative nature of science and the workplace	Students learn that scientists use a wide variety of tools, language, and processes to conduct and describe their research	Students identify and reflect on how science impacts their lives.
Outcomes/Evidence	*Students report a positive attitude toward science and are engaged.*			*Students report a sophisticated notion of scientific research that extends beyond experiments.*	*Students are able to articulate the ways scientists in the videos go about conducting research.*	*Students report on how science is applicable in their lives.* *Students report that they can learn science and are able to use science in their lives.*
Lesson 2: Why do some people become smokers and others do not?	Students are exposed to the wide variation in smoking behavior through profiles of actual smokers	Students learn about the stages of smoking and the extent that genetic and environmental factors contribute to each	Through comparison of smokers, students make preliminary observations, identify patterns, and develop ideas about factors that influence smoking			As students select research topics that interest them, they begin to identify with the science

Outcomes/Evidence	*Students report a positive attitude toward smoker profiles and are engaged in collecting evidence from them.* *Students extend participation by bringing in their own knowledge and experience.*	*Students are able to identify whether a factor is genetic, environmental, or both.* *Students are able to express how these factors may influence smoking behavior.* *Students recognize that there are different stages of smoking*	*Students are able to document sets of both genetic and environmental factors that affect smokers' behavior from the profiles.*		*Students systemically identify patterns in the smoker profiles and are able to use the language of genetic and environmental factors to write and discuss those patterns.*	*Students show deep engagement with their chosen topic and articulate why they are studying that topic.*
Lesson 3: How do genes influence smoking behavior?		Students learn about neurons, neurotransmission, and the reward pathway; how nicotine interacts with the body; and candidate genes	Students consider how a change at the DNA level might affect a complex behavior, and they discuss why certain genes might be good candidates		Students learn terminology related to neuroscience and candidate genes	
Outcomes/Evidence		*Students are able to describe neurotransmission and how nicotine affects the reward pathway.*	*Students are able to apply their knowledge of the reward pathway to predict how genetic variation might affect smoking behavior.*		*Students clearly articulate how genetic variation in certain genes affect the proteins expressed by those genes, and how this might influence variation in smoking*	

				behavior	
Lesson 4: How can we study genetic and environmental influences on smoking behavior?		Students learn about one study design used in epidemiology called the case control study	Students identify the characteristics of the smoking behavior case control study. Students discuss the implications of not matching cases and controls and how to apply the criteria for causality	Students reflect on scientific practice as they compare and analyze different experimental designs	Students learn vocabulary related to case control studies; they learn the design of a case control study	
Outcomes/Evidence		*Students are able to define and use the basic aspects of a case control study including: case, control, sample, exposure, and outcome.*	*Students are able to explain why the study was set up as it was. Students are able to recognize the implications of working with real data and real life constraints.* *Students are able to describe why a case control study design is used in epidemiological research.*	*Students are able to compare and contrast case control studies with experimental studies.* *Students believe case control studies count as scientific research.*	*In their final product, students use appropriate terminology to explain the nature of the data they analyzed.*	
Lesson 5: Analysis of data in case control studies: the odds ratio		Students learn to calculate odds ratios, 95% confidence intervals, distinguish	Students learn to analyze data in an epidemiological study.		Students learn two statistical calculations used in case control	

		between association and causality, and identify sources of error in case control studies.			studies, odds ratio and 95% confidence interval, and how to interpret them.	
Outcomes/Evidence		*Students are able to use the odds ratio and 95% CI to make a statement about association.* *Students discuss whether an identified association meets the criteria for causality.*	*Students are able to articulate why data must be analyzed using an odds ratio calculation.*		*Students are able to identify and use the tools used to analyze epidemiological data (database).*	
Lesson 6: What can we learn from the smoking behavior data?	Students have the opportunity to conduct their own research using an authentic scientific database.	Students apply what they have learned about odds ratios and confidence intervals to test their research hypotheses.	Students evaluate their results and develop arguments based on them.	Students reflect on how different kinds of evidence contributed to their claims about smoking behavior.	Students use the statistical tools of case control studies and the database to test their own hypotheses about smoking behavior; they defend their arguments in front of the class.	By conducting real research using a database, students will deepen their understanding of science and their ability to conduct authentic research.
Outcomes/Evidence	*Students report a positive attitude toward database research and are engaged in data analysis and interpretation.*	*Students are able to use what they learned in the previous lesson to describe how a genetic or environmental factor affects smoking*	*Students develop arguments to explain how genetic and environmental factors affect smoking behavior that include*	*Students demonstrate a sophisticated understanding of the process they followed to conduct their research.* *Students believe*	*Students collaborate between and across groups to complete their research.* *Students are able*	*Students express interest in STEM careers, additional research opportunities.* *Students express ways in which*

	Students extend participation by bringing in their own knowledge and experience and seeking information from additional resources to support their research.	*behavior.*	*claims, evidence, and reasoning.* *Students are able to interpret their data analysis and identify possible false associations, shortcomings of the dataset.*	*database/secondary data analysis counts as scientific research.* *Students use multiple sources of evidence in their arguments.*	*to provide a description of how and why they conducted the research that led them to construct their arguments.*	*science is useful in their lives.* *In follow up interviews, students consider science conducted through computational analysis/computers useful in their lives or career.*
Lesson 7: Hypothesis Generation and Final Project	Students explore the database based on their interests and results of hypothesis testing.	Students use what they know about genetic and environmental factors to propose a future study to better understand how these factors might affect smoking behavior.			Students continue to use statistical tools of epidemiologists, and write up their proposal for future research using a modified NIH template.	
Outcomes/Evidence	*Students bring in their own knowledge to help frame their future research proposal/ hypothesis generation.*	*Students include information about genetic and environmental factors that affect smoking behavior in their future research proposal/hypothesis generation.*			*Student use the template and scientific writing to write and present their future research proposal.*	

Connections to Washington State Essential Academic Learning Requirements

The Exploring Databases curriculum has several connections to the Washington Essential Academic Learning Requirements. The curriculum connects most closely to EALR 1 (Systems), EALR 2 (Inquiry), EALR 3 (Application), and EALR 4 (Life Science: Structure and Function of Living Organizations and Biological Evolution). In the tables below, we have matched the relevant portions of the curriculum with the applicable science and math EALR's.

Table 3a. **EALR 1:** **Systems**
Big Idea: **Systems (SYS)**
Core Content: *Predictability and Feedback*

	Content Standards	**Performance Expectations**	**Connection to Exploring Databases**
	Students know that:	*Students are expected to:*	
9-12 SYSA	*Feedback* is a process in which the *output* of a *system* provides information used to regulate the operation of the *system*. Positive *feedback* increases the disturbance to a *system*. Negative *feedback* reduces the disturbance to a *system*.	• Give examples of a positive feedback system and explain its regulatory mechanism (e.g., global warming causes Earth's ice caps to melt, reflecting less energy to space, increasing temperatures). • Give examples of a negative feedback system and explain its regulatory mechanism (e.g., when a human body overheats, it produces sweat that cools the body by evaporation).	• Students learn about a positive feedback loop, as smoking behavior is reinforced by the reward pathway (Lesson 3).
9-12 SYSB	*Systems* thinking can be especially useful in *analyzing* complex situations. To be useful, a *system* needs to be specified as clearly as possible. **Content Standards**	• Determine if a systems approach will be helpful in answering a question or solving a problem. • Represent the system with a diagram specifying components, boundaries, flows, and feedbacks. • Describe relevant subsystems and the larger system that contains the system being analyzed. • Determine how the system functions with respect to other systems. **Performance Expectations**	• As students explore the complexity of factors related to smoking behavior, they use systems thinking to answer their overarching questions (Lesson 6).

			Connection to Exploring Databases
9-12 SYSC	In complex *systems*, entirely new and unpredictable *properties* may emerge. Consequently, modeling a complex *system* in sufficient detail to make *reliable predictions* may not be possible.	• Create a simplified model of a complex system. Trace the possible consequences of a change in one part of the system and explain how the simplified model may not be adequate to reliably predict consequences.	• Students recognize the limitations of their study (Lessons 5-7). • Students create a simplified model of a complex system as they make predictions based on their results (Lesson 6).

Table 3b. EALR 2:   **Inquiry**
                Big Idea: **Inquiry (INQ)**
                Core Content: *Conducting Analyses and Thinking Logically*

	Content Standards	Performance Expectations	Connection to Exploring Databases
	Students know that:	*Students are expected to:*	
9-12 INQA Question	Scientists *generate* and *evaluate* *questions* to *investigate* the *natural world*.	• *Generate* and *evaluate* a *question* that can be answered through a scientific *investigation*. Critique *questions generated* by others and *explain* whether or not the *questions* are scientific.	• Through videotaped interviews, students learn about the research of several scientists and the different approaches these scientists take to generate research questions (Lesson 1). • Students examine the questionnaire to develop a question related to smoking behavior (Lesson 2). • Students critique other students' results (Final project). • Students generate new hypotheses after they complete their investigations (Lesson 7).

	Content Standards	Performance Expectations	Connection to Exploring Databases
9-12 INQB Investigate	Scientific progress requires the use of various methods appropriate for answering different kinds of research *questions*, a thoughtful plan for gathering data needed to answer the *question*, and care in collecting, *analyzing*, and displaying the data.	• Plan and conduct a scientific *investigation*, choosing a method appropriate to the *question* being asked. • Collect, *analyze*, and display data using calculators, computers, or other technical devices when available.	• Students learn that scientists use a wide variety of tools, language, and processes to conduct and describe their research (Lesson 1). • Students reflect on scientific practice as they compare and analyze different experimental designs (Lesson 4). • Students identify the characteristics of the smoking behavior case control study and recognize why the study was designed as it was (Lesson 4). • Students use a computerized database to analyze and display data in an epidemiological study (Lessons 5-6).
9-12 INQC Explain	*Conclusions* must be logical, based on *evidence*, and consistent with prior *established* knowledge.	• Draw *conclusions* supported by *evidence* from the *investigation* and consistent with *established* scientific knowledge. • Analyze alternative explanations and decide which best fits the data and *evidence*.	• Students examine the evidence (odds ratios and confidence intervals) to draw conclusions about smoking behavior (Lesson 6). • Students develop alternative explanations for the evidence, whether or not the evidence supports their hypothesis (Final Project).
9-12 INQD Communicate Clearly	The methods and procedures that scientists use to obtain *evidence* must be clearly reported to enhance opportunities for further *investigation*.	• Write a detailed laboratory report that includes: the *question* that motivated the study, a justification for the kind of *investigation* chosen, *hypotheses* (if any), a description of what was done, a summary of data in tables and graphs, and a *conclusion*, based on the *evidence*, that responds to the *question*.	• Students prepare a final presentation containing background information, their overarching hypothesis, methods, results, and conclusion (Final Project).
9-12 INQE Model	The essence of scientific *investigation* involves the development of a *theory* or conceptual *model* that can *generate* testable predictions.	• Formulate one or more *hypotheses* based on a *model* or *theory* of a causal *relationship*. Demonstrate creativity and critical thinking to formulate and *evaluate* the *hypotheses*.	• Students develop and test a hypothesis (Lessons 2-6). • Students use the Criteria for Causality checklist to provide evidence of a causal relationship (Lesson 6). • After testing their hypothesis, students generate new hypotheses for future studies (Lesson 7).

	Content Standards	Performance Expectations	Connection to Exploring Databases
9-12 INQF Communicate	*Science* is a human endeavor that involves logical reasoning and creativity and entails the testing, revision, and occasional discarding of theories as new *evidence* comes to light.	• *Evaluate* an *investigation* to determine if it was a *valid* means of answering the *question*, and whether or not the results were *reliable*. *e • *Describe* the development of a scientific *theory* that illustrates logical reasoning, creativity, testing, revision, and replacement of prior *ideas* in light of new *evidence*.	• In their conclusions, students determine whether their study was a reliable means of answering their overarching question (Lesson 6). • Students develop new ideas in light of evidence they discover through data analysis (Lesson 6). • Students recommend new studies that might test their question more appropriately (Lesson 7).
9-12 INQG Intellectual Honesty	Public *communication* among scientists is an essential aspect of research. Scientists *evaluate* the *validity* of one another's *investigation*s, check the *reliability* of results, and *explain* inconsistencies in findings.	• Participate in a scientific discussion about one's own *investigation*s and those performed by others. • Respond to *questions* and criticisms, and if appropriate, revise explanations based on these discussions.	• Students learn about the social/collaborative nature of science and the workplace (Lesson 1). • Students defend their results as part of the final project (Final Project). • Students ask questions of other students to evaluate the validity of each other's investigations (Final Project).
9-12 INQH Intellectual Honesty	Scientists carefully *evaluate* sources of information for *reliability* before using that information. When referring to the *ideas* or findings of others, they cite their sources of information.	• Provide appropriate citations for all *ideas*, findings, and information used in any and all written reports. • *Explain* the consequences for failure to provide appropriate citations.	• Students provide citations for all sources (Final Project).

Table 3c.   EALR 3:   Application
Big Idea: Application (APP)
Core Content: *Science, Technology, and Society*

	Content Standards	Performance Expectations	Connection to Exploring Databases
	Students know that:	*Students are expected to:*	
9-12 APPA	*Science* affects society and *cultures* by influencing the way many people think about themselves, others, and the	• Describe ways that scientific ideas have influenced society or the development of differing cultures.	• Students identify and reflect on how science impacts their lives (Lesson 1). • Students research questions related to

	environment. Society also affects *science* by its prevailing views about what is important to study and by deciding what research will be funded.	• List questions that scientists investigate that are stimulated by the needs of society (e.g., medical research, global climate change).	public health (Lesson 2-7). • Students explore the implications of their research for the health of their community (Final Project).
9-12 APPB	The *technological design process* begins by defining a problem in terms of *criteria* and *constraints*, conducting research, and generating several different *solutions*.	• Work collaboratively with other students to *generate ideas* for solving a problem. Identify *criteria* and *constraints*, research the problem, and *generate* several possible *solutions*.	• Students work together to recognize the implications of working with real data and real life constraints. • Students understand the constraints of this particular study (i.e. matching, multiple comparison issues) (Lessons 3, 7). • Students apply the Criteria for Causality to their results (Lesson 6). • Students work together to discover several solutions to their results (Lesson 6).
9-12 APPC	Choosing the best *solution* involves comparing alternatives with respect to *criteria* and *constraints*, then building and testing a *model* or other representation of the final design.	• Choose the best *solution* for a problem, create a *model* or drawing of the final design, and devise a way to test it. Redesign the *solution*, if necessary, then present it to peers.*b	• Students analyze their results to choose a solution (Lesson 6). • Students redesign the study if necessary as they generate new hypotheses (Lesson 7).
9-12 APPD	The ability to solve problems is greatly enhanced by use of mathematics and information technologies.	• Use proportional reasoning, *functions*, graphing, and estimation to solve problems.*a*b*c • Use computers, probes, and software when available to collect, display, and analyze data.	• Students use a computerized database to analyze and display data in an epidemiological study (Lessons 5, 6). • Students learn to calculate odds ratios, 95% confidence intervals, distinguish between association and causality, and identify sources of error in case control studies (Lesson 4).
9-12 APPE	Perfect *solutions* do not exist. All technological *solutions* involve trade-offs in which decisions to include more of one quality means less of another. All *solutions* involve consequences, some intended, others not.	• *Analyze* a societal issue that may be addressed through *science* and/or technology. *Compare* alternative *solutions* by *considering trade-offs* and unintended consequences (e.g., removing dams to increase salmon spawning).	• Students consider the complications of applying their results to society (Final Project).

9-12 APPF	It is important for all citizens to *apply science* and *technology* to critical issues that influence society.	• Critically *analyze* scientific information in current events to make personal choices or to understand public-policy decisions.	• Students are more prepared to make personal decisions after they learn about the environmental and genetic effects on smoking behavior.

Table 3d. EALR 4:   Life Science
Big Idea: Structures and Functions of Living Organisms (LS1)
Core Content: *Processes Within Cells*

	Content Standards	**Performance Expectations**	**Connection to Exploring Databases**
	Students know that:	*Students are expected to:*	
9-11 LS1C	Cells contain specialized parts for determining essential *functions* such as regulation of cellular activities, *energy* capture and release, formation of proteins, waste disposal, the *transfer* of information, and movement.	• Draw, label, and describe the functions of components of essential structures within cells (e.g., cellular membrane, nucleus, chromosome, chloroplast, mitochondrion, ribosome).	• Students learn about the structure and function of neurons (Lesson 3).
9-11 LS1E	The *genetic information* responsible for inherited *characteristics* is encoded in the DNA *molecules* in *chromosomes*. DNA is composed of four subunits (A,T,C,G). The sequence of subunits in a *gene* specifies the amino acids needed to make a protein. Proteins express inherited traits (e.g., eye color, hair texture) and carry out most cell *function*.	• Describe how DNA molecules are long chains linking four subunits (smaller molecules) whose sequence encodes genetic information. • Illustrate the process by which gene sequences are copied to produce proteins.	• Students apply their understanding of gene expression to explain how variation at the gene level can affect protein production and thus can contribute to variability in smoking behavior (Lesson 3).
9-11 LS1G	Cells use the DNA that forms their *genes* to encode *enzymes* and other proteins that allow a cell to grow and divide to produce more cells, and to respond to the *environment*.	• Explain that regulation of cell functions can occur by changing the activity of proteins within cells and/or by changing whether and how often particular genes are expressed.	• Students learn that nicotine stimulates the reward pathway, thus causing a change in the activity of proteins within cells as dopamine is released (Lesson 3)

Table 3e. EALR 4:   Life Science
**   Big Idea: Biological Evolution (LS3)**
**   Core Content:** *Mechanisms of Evolution*

	Content Standards	Performance Expectations	Connection to Exploring Databases
	Students know that:	*Students are expected to:*	
9-11 LS3B	Random changes in the *genetic* makeup of cells and *organisms* (*mutations*) can cause changes in their physical *characteristics* or behaviors. If the *genetic mutations* occur in eggs or sperm cells, the changes will be inherited by offspring. While many of these changes will be harmful, a small minority may allow the offspring to better survive and reproduce.	• Describe the molecular process by which organisms pass on physical and behavioral traits to offspring, as well as the environmental and genetic factors that cause minor differences (variations) in offspring or occasional "mistakes" in the copying of genetic material that can be inherited by future generations (mutations). • Explain how a genetic mutation may or may not allow a species to survive and reproduce in a given environment.	• Students consider how genetic and environmental factors affect a complex behavior (Lessons 2-7). • Students discuss why certain genes might be good candidates for causing variation in smoking behavior (Lesson 3).

Table 3f Mathematics EALR Connections

Math Standard	Connection to Exploring Databases
A1.8.A Analyze a problem situation and represent it mathematically.	• Students use the odds ratio and confidence interval calculations to analyze and represent data mathematically (Lessons 5-7).
A1.8.B Select and apply strategies to solve problems.	• Students apply mathematical concepts of odds ratios and confidence intervals to research their questions (Lessons 5-7).
A1.8.E A2.8.E Read and interpret diagrams, graphs, and text containing the symbols, language, and conventions of mathematics.	• Students interpret graphs in analyzing data from the smoking behavior database (Lessons 2-7).
A2.6.G Calculate and interpret margin of error and confidence intervals for population proportions.	• Students use the database to calculate and interpret confidence intervals (Lessons 5-7).

A1.6.B Make valid inferences and draw conclusions based on data.	• Students make inferences and draw conclusions based on their analysis of the smoking behavior data (Research Project Pages, Final Project).
A1.8.C Evaluate a solution for reasonableness, verify its accuracy, A2.8.C and interpret the solution in the context of the original problem.	• Students evaluate and interpret their results in the context of their original overarching question (Lessons 6-7, Research Project Pages, Final Project).
A1.8.F Summarize mathematical ideas with precision and efficiency A2.8F for a given audience and purpose.	• Students summarize their odds ratio and confidence interval results and present them to the class (Research Project Pages, Final Project).
A1.8.G Synthesize information to draw conclusions and evaluate the A2.8H arguments and conclusions of others.	• Students synthesize the information from their research study, and they evaluate their peers' conclusions (Research Project Pages, Final Project).

Framework for K-12 Science Education

In 2012 the National Academies Board on Science Education and the Committee on a Conceptual Framework for New K-12 Science Education Standards released *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (available at http://www.nap.edu/catalog.php?record_id=13165). The document is a first step in the creation of the Next Generation Science Standards. It is grounded in research on science and science learning, with the goal of identifying what all K-12 students should know about science concepts and practice. In the table below, we have matched the relevant portions of the curriculum with select practices, cross-cutting concepts, and core ideas presented in the framework.

Table 4

Framework for K-12 Science Education	Exploring Databases
1. Scientific Practice	
Asking questions	Students identify a research topic, an associated question and hypothesis, and choose four questions from the data to investigate (Lessons 2-7).
Planning and carrying out investigations	Students plan an investigation related to smoking behavior and carry it out through hypothesis testing (Lessons 2-7).
Analyzing and interpreting data	Students analyze and interpret data from the database during both hypothesis testing and hypothesis generation (Lessons 5-7).
Using mathematics and computational thinking	Students use statistical thinking to analyze data and to evaluate evidence gained from querying the database (Lessons 5-7).
Constructing explanations	Students construct explanations related to their question and hypothesis and apply the criteria for causality using data from their hypothesis testing (Lesson 6).
Engaging in argument from evidence	Students write, present, and discuss their arguments as they explain how their data support their claim (Lessons 6-7).
Obtaining, evaluating, and communicating information	Students communicate their arguments and culminating product to an audience (Lesson 7).
2. Crosscutting Concepts	
Cause and Effect: Mechanism and Explanation	Students apply the criteria for causality to determine the likelihood of cause and effect relationships found in their data analysis (Lesson 6).
System and System Models	As students explore the complexity of factors related to smoking behavior, they use systems thinking to answer their overarching questions (Lesson 6).
3. Disciplinary Core Idea	
LS1.D: Information Processing *How do organisms detect, process, and use information about the environment?*	Students understand how nicotine affects the reward pathway in the brain (Lesson 3).
LS3.B: Variation of Traits *Why do individuals of the same species vary in how they look, function, and behave?*	Students understand smoking behavior is influenced by both genetic and environmental factors (Lessons 1-7).

References

Bell, P, Lewenstein, B, Shouse, AW, Feder, MA, Editors, Committee on Learning Science in Informal Environments, National Research Council. 2009. *Learning Science in Informal Environments: People, Places, and Pursuits*. The National Academies Press

Committee on Conceptual Framework for the New K-12 Science Education Standards, National Research Council. 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. The National Academies Press

Fenichel, M and Schweingruber, HA. 2010. *Surrounded by Science: Learning Science in Informal Environments.* The National Academies Press

Office of Superintendent of Public Instruction, State of Washington (June, 2009). *K-12 Science Standards*, Retrieved July 18, 2011 from http://www.k12.wa.us/Science/Standards.aspx.

Office of Superintendent of Public Instruction, State of Washington (2008). *K-12 Mathematics Standards*, Retrieved July 18, 2011 from http://www.k12.wa.us/Mathematics/Standards.aspx

Basic flow of hypothesis testing and hypothesis generation

	Activity	Lesson
Hypothesis Testing*	**1. Develop your overarching hypothesis** ○ Utilize *smoking profiles* to think about how people become smokers ○ Draw on personal experience and observations ○ Use smoker profiles and personal experience to develop *a big idea* or *category* such as parents, friends, and education	Lesson 2-5
	2. Develop your specific hypotheses ○ Identify a few survey items related to the overarching hypothesis ○ Define the *exposure* and *non-exposure* using possible responses to each question ○ Define the study population	Lesson 5,6
	3. Determine whether there is a statistically significant association ○ Go to the database and set up the odds ratio calculation as defined in #2 ○ Estimate the *odds ratio* and *confidence interval* ○ Interpret the results and *determine* if there is a significant association or not	Lesson 6
	4. Judge whether association may be causal by applying criteria for causality ○ If there is a significant association, apply the criteria for causality ○ Think *critically* about results along with the criteria for causality and *judge* whether the exposure may increase or decrease the risk of smoking in a causal sense ○ Recognize that the purpose here is to make the inference, NOT to prove causation.	Lesson 6
	5. Develop your claims for each specific hypothesis and the overarching hypothesis (*Claim, evidence,* and *reasoning* are the basic components of a scientific argument) ○ If the claim indicates causation, present evidence and reason for how the evidence supports the hypothesis ○ If the claim indicates only association but not causation, give evidence for this claim (e.g. confounder) ○ If there is no association, provide explanations for why this is the case ○ Provide *alternative explanation*s and *rebuttals* (refutations to alternative explanations) that improve the quality of the claim	Lesson 6
Hypothesis Generation	**6. Develop new hypotheses and design a case control study** ○ Utilize the database to query *many* questions and combinations of exposed/not-exposed to obtain results to form *new hypotheses*. Look at the data to determine what should be exposed and not-exposed in order to optimize the odds ratio ○ Design a *new case control study* to test new hypotheses ○ Do not develop any scientific conclusion since you are performing many tests with the same data, making the results (odds ratio and 95% confidence intervals) less reliable	Lesson 7

* Note that the hypothesis testing in the curriculum is a set of two components: statistical tests (estimating odds ratio and confidence interval) for association and reasoning with the criteria for causality.

Lesson 1. Why and how do people do science?

Overview
Students learn that scientists study many different aspects of the natural world and there are many ways to go about conducting scientific research. They discuss why scientists study the questions they do, how they go about doing their research, and how they became scientists. They are also exposed to research approaches that are different from classical experimental research and that are frequently employed in studies involving human subjects. This lesson can be used as an introduction to the Exploring Databases unit or in the context of teaching about the nature of science.

Learning objectives

Objective	Evidence
Students will understand the many ways of doing scientific research in addition to traditional experimental research designs.	Through their writing and discussion, students cite evidence from the videos that shows that science can include observational and experimental research.
Students will understand that scientists have many compelling reasons for doing their research.	Through their writing and discussion, students recognize reasons scientists do their research, such as: Interest in understanding how the natural world works Interest in learning to help people from diverse backgrounds

Lesson Preparation

Time
40 – 50 minutes

Materials
for each student
The invitation letter and Student Sheet 1.1
The Research Project Pages packet (to be completed throughout the entire unit)
for class, homework
Scientist Interview clips (video, audio, and/or transcript)

Getting Ready
Photocopy the Invitation Letter and Student Sheet 1.1, the Research Project Pages, and Scientist Interview transcripts if necessary. Make sure you can play the videos on your computer and projector. They can be accessed from https://gsoutreach.gs.washington.edu/database2/

1

Presenting the Lesson

A.   Discussing the Invitation Letter

1. If you do this lesson directly before the unit, provide students with the Invitation Letter at the beginning of this lesson or the previous day. Tell students they will continue to get more information about the research study described in the letter, and this is just the initial introduction. Make sure students save the letter with their individual notebooks in case they need to refer back to it later in the unit.
2. Ask students whether they have any questions about the smoking behavior study, and write them on the board or a chart paper. You may choose to answer some of the questions right away, or let students know that they will get answers later.

B. Viewing scientist interview clips

3. Tell students that they will be learning about different ways that scientists conduct research by viewing video-taped interviews.
4. Prompt students to share some of their ideas about science by asking the questions: (Keep track of student responses at the front of the room on chart paper or a document camera.)
    a. Can you describe what science is?
    b. What does a scientist do?
    c. Can you think of jobs and everyday activities that require you to use what you know from science?
    d. Do you know anyone who uses science in their job? Who is that person? What do they do?

Teacher Note: You may want to have the students quickly draw their understanding of science or a scientist, to use as a way to start the discussion.

---

5. Provide students with Student Sheet 1.1.
6. Review the questions on the student sheet and then show the clips from the scientist interviews. You can decide to show them all at once or discuss each video after it plays. As the class is watching the videos, the students should begin to take notes and complete Student Sheet 1.1. Replay videos as needed.

Ethical Consideration: The interview clips with LB include her speaking about her research on mice and doing injections into their brains. You may want to be prepared to discuss the standards and practices of research on animals.
The Northwest Association on Biomedical Research has a resource on animal research that you can consult if this comes up in your classroom: http://www.nwabr.org/curriculum/animals-research

---

7. Ask students to share their answers, and facilitate a discussion on how and why scientists do the work they do. You may go through each question on Student Sheet 1.1 or choose an overarching question like the ones below.

Discussion Question: Are there things the scientists talked about that seem different from how the students usually think about science?

Students should talk about scientists working on observational and experimental research projects. They may also discuss scientists' interests in understanding how the natural world works, helping all people from diverse backgrounds, and making new discoveries about the natural world.

Many scientists do work that is important to their community. Ask students if they can think of a research question/idea that would help their community or someone they care about.

Discussion Question: What issues in your community can be addressed through research? What research could you do to understand more about this issue?

Teacher Note: If you do this lesson in conjunction with the Exploring Databases unit, have students spend 5 minutes brainstorming about why people would study smoking behavior. How can it help society to have a better understanding of smoking behavior and addiction?

---

8. HOMEWORK: To prepare for lesson 2, ask students to complete RPP-1 in the Research Project Pages.

Acknowledgements

University of Washington
School of Medicine
Education Outreach
Department of Genome Sciences
University of Washington
Seattle, Washington

September, 2012

Dear Student,

We would like to invite you to be part of the scientist team for a research study of smoking behavior. Since 2007, the University of Washington (UW) and students around Washington have collaborated on this study. Your job will be to use the data from the Smoking Behavior Research Study to better understand the question, "**What genetic and environmental factors influence people to become smokers?**"

The study compares two groups of people called **cases** (smokers) and **controls** (people who tried smoking but didn't become smokers). Both groups of people came to the UW to fill out a research questionnaire and give a blood sample. The questionnaire helps us to learn about the subjects' lifestyles and environmental factors that might have influenced whether or not they became smokers. We purified DNA from the blood samples and analyzed three regions of the subjects' DNA using a process called genotyping.

In past school years, students contributed to two aspects of the study. During the design phase, high school students wrote questions for the research questionnaire. From 2007-2009, students from schools throughout Washington genotyped subject DNA in their classroom labs. All of the genotyping and questionnaire data were entered into a database.

Your role in this research will be to propose and test a hypothesis about how genes and the environment might influence smoking behavior. Then after exploring and analyzing all the data, you will generate hypotheses and ideas for future studies to share with your peers and the team at UW. In the future, you will also have the opportunity to submit your results to the University of Washington. We look forward to hearing about your findings, and we appreciate your help.

Sincerely,

The UW Smoking Behavior Research Team

Key to **Student Sheet 1.1**

Description of Each Clip

Theresa Madrid, Graduate Student, Public Health Genetics, University of WA
> *Theresa talks about the personal connections she has with her work and how she envisions her work will impact her community.*

Joel Grow, Graduate Student, Clinical Psychology, University of WA
> *Joel discusses how he made the transition from software engineering to clinical psychology through his teaching experiences, interests, and conversations with professors. He also talks about the interdisciplinary nature of psychology.*

Josh Akey, Associate Professor of Genome Sciences, University of WA
> *Josh describes his recent work on understanding the genetic differences between dog breeds. He discusses how databases are integral to his work and how they are used in genome sciences today.*

Lisa Beutler, Graduate Student, Genome Sciences, University of WA
> *Lisa describes how the science she undertakes is different than high school and undergraduate science. She also talks about how she studies addiction in mouse models.*

Tracie Delgado, Graduate Student, Microbiology, University of WA
> *Tracie talks about how collaboration and mentoring play a role in her lab and the challenges she faces in her work.*

Jeff Stanaway, Graduate Student, Epidemiology, University of WA
> *Jeff describes how he decided to pursue a career in epidemiology and how collaboration is essential to work in the field.*

To help answer each question, show the clips written under each question.

Possible responses students may give are shown in italics.

1. How did the scientists in the interviews become interested in science and research?
    (Theresa Madrid, clip 1; Joel Grow; Jeff Stanaway, clip 1)

    Their interest began in high school and continued into college and graduate school. Their research was based on the particular question that interests each of them and that was personal to them. Reading scientific books and outdoor activities also contributed to some of the scientists' initial interest.

2. What personal connections do these scientists have with their work?
    (Theresa Madrid, clip 1)
    Student answers will vary.

3. Using your experience and any information from the interviews, describe what you think happens in science. What do scientists do? How do they do experiments? Where do scientists do science?
(Josh Akey, clip 1; Lisa Beutler, clip 1; Tracie Delgado)

Science is the process of asking questions, testing and repeating experiments to answer the question you're curious about.

4. How are the ways that these scientists do work similar or different from how you do science in school or at home?
(Lisa Beutler, clip 2;Josh Akey, clip 2)

These scientists design their own experiments and carry them out according to their own schedule. They are usually new experiments that have never been carried out before. Lab experiments done at school are a lot more limited by time and creativity. It is a learning process that repeats other scientists' experiments.

5. What are some ways that you see science being applied in your everyday life?

Students should write a variety of ideas about how science is applied in everyday life. This question will help them explore the connections they are making to science.

6. How do scientists work together to solve problems and answer new questions?
(Tracie Delgado, clip 2; Jeff Stanaway, clip 2)

Scientists of different fields come together with their specialized skills and knowledge to contribute to the same problem or experiment. They find a common language to communicate with each other and make use of each other's skills to further the investigation or question.

7. What surprised you about these scientists? About the kinds of science they talk about?

Student answers will vary.

Lesson 2. Why do some people become smokers and others do not?

Overview
Students learn that both environmental and genetic factors can contribute to smoking behavior. Using profiles of actual smokers, students identify patterns in people's smoking behaviors. They begin to collect evidence from the profiles to inform their future database research.

This lesson introduces the driving question of the entire unit, "Why do some people become smokers and others do not?" Students learn that there are many factors that may contribute to smoking behavior and begin to identify ones that are interesting to them. Students form research groups of four students based on specific factors they would like to study further, and they work together in these groups for the rest of the unit. In this lesson, students are also introduced to the culminating research project.

Learning objectives

Objective	Evidence of Learning
Students will understand that environmental and genetics factors contribute to smoking behavior in humans.	Students recognize the types of smoking behaviors people describe and different factors that can be associated with smoking in the smoker profiles. Several factors that may be associated with smoking behavior include: having peers who smoked, feeling a buzz, being allowed to smoke at home, and having cravings.
Students develop preliminary hypotheses based on evidence from the smoker profiles.	Students write and discuss hypotheses about why people become regular smokers and justify them with supporting evidence from the smoker profiles.

Lesson Preparation

Time
100-120 minutes

Materials
for each student
Student Sheet 2.1, 2.2, 2.3, RPP-2-3, and Questionnaire
for each group
Copies of all 11 smoker profiles for the class so each student gets one profile and each group gets four different ones.

Getting Ready
1. Photocopy the smoker profiles and Student Sheets 2.1, 2.2, 2.3, RPP-2-3, and Questionnaire.
2. Make copies of Figures 2.1, 2.2, and 2.3 for use with document camera or overhead projector.
3. On Day 2, make sure that you can access and project the database on the classroom computer (see Manual).

Presenting the Lesson
Day 1
A. What factors contribute to smoking behavior?

1. Tell students that today they will be using profiles of real smokers to identify common factors that influence smoking behavior. Some of those factors will be environmental, such as a person having friends who smoke, and some will be physiological, such as a person feeling a buzz.

Teacher Notes

As the class discusses the smoker profiles, be sure that students understand the following concepts:

- Smoking behavior is influenced by both environmental and physiological factors.

- Physiological factors (traits) are determined wholly or in part by genes. Many of the physiological factors we discuss in this unit, such as feeling a buzz when smoking, are probably determined almost entirely by genes. In this unit we will equate physiological factors with genetic factors. Students may have difficulty making the connection that a physiological trait is caused by a gene. If so, review central dogma: DNA (a gene) codes for mRNA, which is translated into a protein, which results in a trait. How quickly students are able to equate physiological factors with genetic factors will depend on their familiarity with central dogma and how often they have applied it.

- Traits that are determined by multiple factors, including both genetic and environmental factors are called "multifactorial traits." You may want to encourage students to suggest other multifactorial traits. Some examples include behaviors like smoking behavior, intelligence, height, weight, and many common diseases, such as heart disease or type 2 diabetes.

- Environmental and genetic factors contribute to different stages of smoking to different extents, a feature your students should notice as they evaluate the smoker profiles. For example, initiation is believed to be more heavily influenced by environmental factors, while maintenance and quitting have a stronger genetic influence.

---

2. Have students get into groups of four, and give each group four different smoker profiles. Ask each student to read one profile (or all if there is time). Then ask students to respond to the questions on Student Sheet 2.1 as a group by pooling the information they have gleaned from each of the profiles they've read. When they finish the questions, have your students go back through the sheet and identify the factors that influence smoking behavior as either E (environmental) or G (genetic).

3. Provide each student with Student Sheet 2.2, or ask the students to draw Figure 2.1 in their notebooks. Project Figure 2.1, and fill out each section of the chart as a class. Ask each group to suggest one or two factors for each of the first three boxes, using their notes from Student Sheet 2.1. As teams present, also ask them to identify whether their statements might be genetic, environmental, or both. **Introduce terms such as addiction, drug abuse, nicotine, or withdrawal, and write the definition in the bottom box (definitions are provided in the glossary).** Encourage all students to take notes on Student Sheet 2.2 as the other groups present, as they will want to use this information later as part of their final project.

Discussion Questions and Evidence of Student Understanding

- Is there a difference in the way that genetic and environmental factors seem to influence the stages of smoking shown in Figure 2.1 (i.e. trying, continuing, and quitting)?

Students should recognize that there are genetic and environmental factors present throughout the chart, but that trying smoking is more under the influence of environmental factors, while genetic factors play a strong role in later stages of smoking- continuing and quitting.

If students do not understand, use Figure 2.1 to show that even though environmental and genetic factors are present in each case, they identified mainly environmental factors for "Why people try smoking" and a larger proportion of genetic factors for "Why people continue to smoke after trying it."

- Do you think that some aspects of smoking behavior could be inherited?

Students should recognize that the physiological factors that affect smoking behavior are determined by genes (i.e. are genetic factors). From the smoker profiles, they will observe the wide variation among individuals in their physiological responses to smoking. For example, some, but not all people report getting a buzz when they smoke, some people can quit smoking the first time they try while others relapse, etc. This physiological variation is due to genetic variation. How people respond physiologically to smoking probably affects their smoking behavior— a person who has a strong craving between cigarettes is more likely to continue smoking than someone who doesn't crave cigarettes. And since those physiological/genetic factors are inherited, it's probable that there will be similarities in smoking behavior from one generation to the next.

- What are some aspects of smoking behavior that seem to vary among individuals?

A great example here is how easy or difficult it is for people to quit smoking. Students may discuss how long the person has smoked, number of cigarettes smoked, situations the person smokes in, and quitting strategy employed.

Glossary from the National Institute On Drug Abuse:
   http://teens.drugabuse.gov/blog/category/word-of-the-day

- **Addiction.** A chronic, relapsing disease characterized by compulsive drug-seeking and abuse and by long-lasting chemical changes in the brain.

- **Drug abuse.** The use of illegal drugs or the inappropriate use of legal drugs. The repeated use of drugs to produce pleasure, to alleviate stress, or to alter or avoid reality (or all three).

- **Drug.** A chemical compound or substance that can alter the structure and function of the body. Psychoactive drugs affect the function of the brain, and some of these may be illegal to use and possess.

- **Nicotine.** The addictive drug in tobacco. Nicotine activates a specific type of acetylcholine receptor.

- **Withdrawal.** Symptoms that occur after chronic use of a drug is reduced or stopped.

HOMEWORK: Provide each student with a copy of the questionnaire. Ask students to complete the questionnaire as if they were the person in the profile they read.

Day 2
B. What Factors Might Influence Someone Becoming a Regular Smoker After Trying It?
1. Present students with the diagram shown in Figure 2.2. Explain that this is one way to show the stages of smoking. Ask students, working in pairs, to show what each stage represents by placing the following 4 smoker profiles in one of the four stages on the diagram: Barbara (nonsmoker), Mark (former smoker), Drew (regular smoker), and Kayley (experimental smoker or regular smoker).

2. Ask students where they placed each person on the diagram and why. Point out that the data they will analyze targets the stage in which people are experimenting with smoking. Color the arrows that point from Experimental smokers to Nonsmokers and from Experimental smokers to Regular smokers. Explain that they will be comparing two groups of people, those who initiated and then became regular smokers (cases) and those who initiated but went back to being nonsmokers (controls).

3. Guide students in identifying related factors that might influence whether or not someone becomes a regular smoker after initiating, based on the smoker profiles and their own observations of smokers and nonsmokers. Record these as a list or in a simple concept map. These factors are the research topic areas that students may choose to investigate as part of this unit.

Teacher Notes
Some topic areas student may identify are:
- Family smoking
- Peer influences
- Public policy (restaurant smoking, age for buying cigarettes, etc.)
- Physiological/genetic effects
- Media influences

Topics covered well in the questionnaire
This list will help you to guide students toward research topics that can be investigated using the database.
- Features of their experimental smoking phase such as age of onset, duration, intensity of smoking, places where they usually smoked (questions 12-17)
- Physiological effects of smoking during experimental smoking (questions 18a-18i)
- Genetic factors (questions 103-105)
- Subject's attitudes toward smoking during experimental smoking stage (questions 19-21)
- Parent and family influences on smoking (questions 26-36)
- Peer influences on smoking (questions 37-40)
- Influences at school (questions 41-46)
- Influences at work (questions 47-50)
- Influences of religion and ethnicity (not a strong topic, questions 51-52)
- Influences of advertising and the media (questions 53-55)
- Influences of public policy (questions 56-61)
- Life during the teen years (lifestyle, diet, exercise, participation in organized activities, drinking) (questions 62-67)
- Emotional state during teen years (questions 68-70)
- Antismoking education during teen years and in school (questions 41-44, 71-72)
- Level of education and economic status (questions 80-83)

C. Introduction to the Questionnaire and Database

1. Ask students to get out their copy of the questionnaire.

2. Project the database in the front of the room, and select "Step 1.1. Subject-classification and matching questions"
    https://gsoutreach.gs.washington.edu/database2/

3. Select question 1 (How many cigarettes have you smoked in your entire life?). As you are displaying question 1, remind students that cases are smokers and controls are people who tried smoking but did not become smokers. Show the table and graph, and explain that the graph shows the number of cases and the number of controls that gave each response.

4. Now select "Step 1.3 Hypothesis Testing" in the left column. Near the top of the page, select "Show 100 entries" to show all the questions available for testing hypotheses. As a class, look through the questions to find connections to each topic area you just discussed. Let students know that they will explore four of these questions that relate to their research topic.

Teacher Notes
Explain to students that there should be at least four questions that can be used to investigate each research topic. If students can't find any questions that seem related to their research topic, they will need to select a different topic. Here are a few sample topics and a possible question for each topic.

Physiological/Genetic #18.1: *During your experimental smoking phase, did you experience any of the following regularly? - A pleasurable feeling*

Family #26: *While you were growing up, how many of your parents/guardians smoked at all?*

Education background #41: *When you were in school, were you taught about the dangers of smoking (for example, lung cancer, heart disease, fire hazards)?*

Media #54: *While you were growing up, how often did you see smoking advertisements on television, radio, billboards, posters, newspapers, and magazines?*

Peers #40: *During your experimental smoking phase, did your friends think that smoking was "cool"?*

D. Forming Research Teams and the Team Research Investigation

1. Remind students that we are trying to understand the driving question:
    "Why do some people become regular smokers and others do not?"

Teacher Note
Throughout the curriculum we will be coming back to this driving question. You may want to post it in your class or record it as a class.

2. Guide students in identifying the particular topic area that interests them, and group students in research teams of four students based on their shared research interests. Make sure that each research topic is well covered in the questionnaire—if not, it may be necessary to combine a few related topics. Throughout the unit there will be sharing among teams, but each team will stay focused on the topic they select. Make sure each group pursues a different topic so as a class you have a wide range of research ideas being investigated.

Teacher Note

The goal is for students to pursue their own interests within the limits of the UW Smoking Behavior Research Study. However, you know your students, and you may need to assist in the creation of groups that will function effectively together.

Note: Certain topic areas (for example, ethnic and religious background) are not well addressed in the database. As students choose research topics, encourage them to broaden their topic if you think there may not be sufficient information for an interesting investigation.

---

3. Explain to students that some of the constraints of using an existing database to test hypotheses is that you must first develop your overarching hypothesis BEFORE analyzing the data AND you can only ask a few questions. To explain the second point, show students Figure 2.3. Multiple Comparison Problem: Why we need to form hypotheses before looking at the data. The more tests you perform, the greater the likelihood that you will incur a false association, so we are limiting our study to only four questions.

4. After students are sitting with their groups, hand out the Research Project Pages. Read the first page of the document as a class, and give students opportunities to ask questions. Throughout the unit they will be collecting information and items such as graphs, research results, paragraphs they write, and quotes to use in their final project. The activities in this packet are intended to be completed with specific lessons, so students apply the concepts taught in that lesson to their own research question.

5. Ask each group to complete RPP-2-3, Lesson 2: Research Topic and Hypothesis. They should discuss their responses as a group, but each student should complete their own sheet.

    HOMEWORK: Hand out Student Sheet 2.3 and have students read about the stages of smoking. Ask students to highlight or record three points in the reading that might be important to their group's topic.

Figure 2.1. Observations about smoking behavior

WHY PEOPLE TRY SMOKING
WHY PEOPLE CONTINUE TO SMOKE AFTER TRYING IT
WHY PEOPLE HAVE DIFFICULTY QUITTING SMOKING
RELATED TERMS

Figure 2.2: Diagram of Smoking Behavior. Steps 1 though 4 show four stages of smoking that many smokers experience.
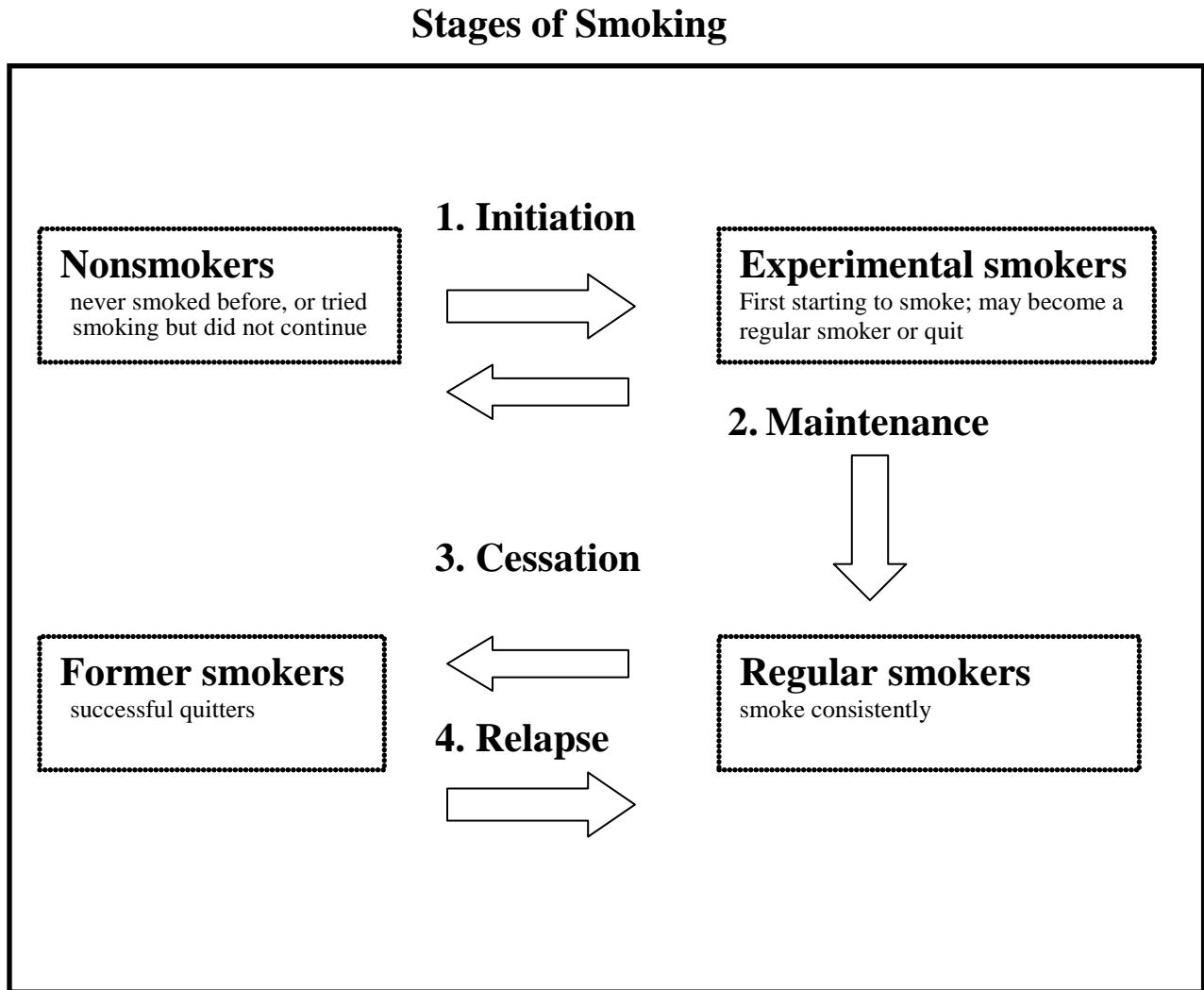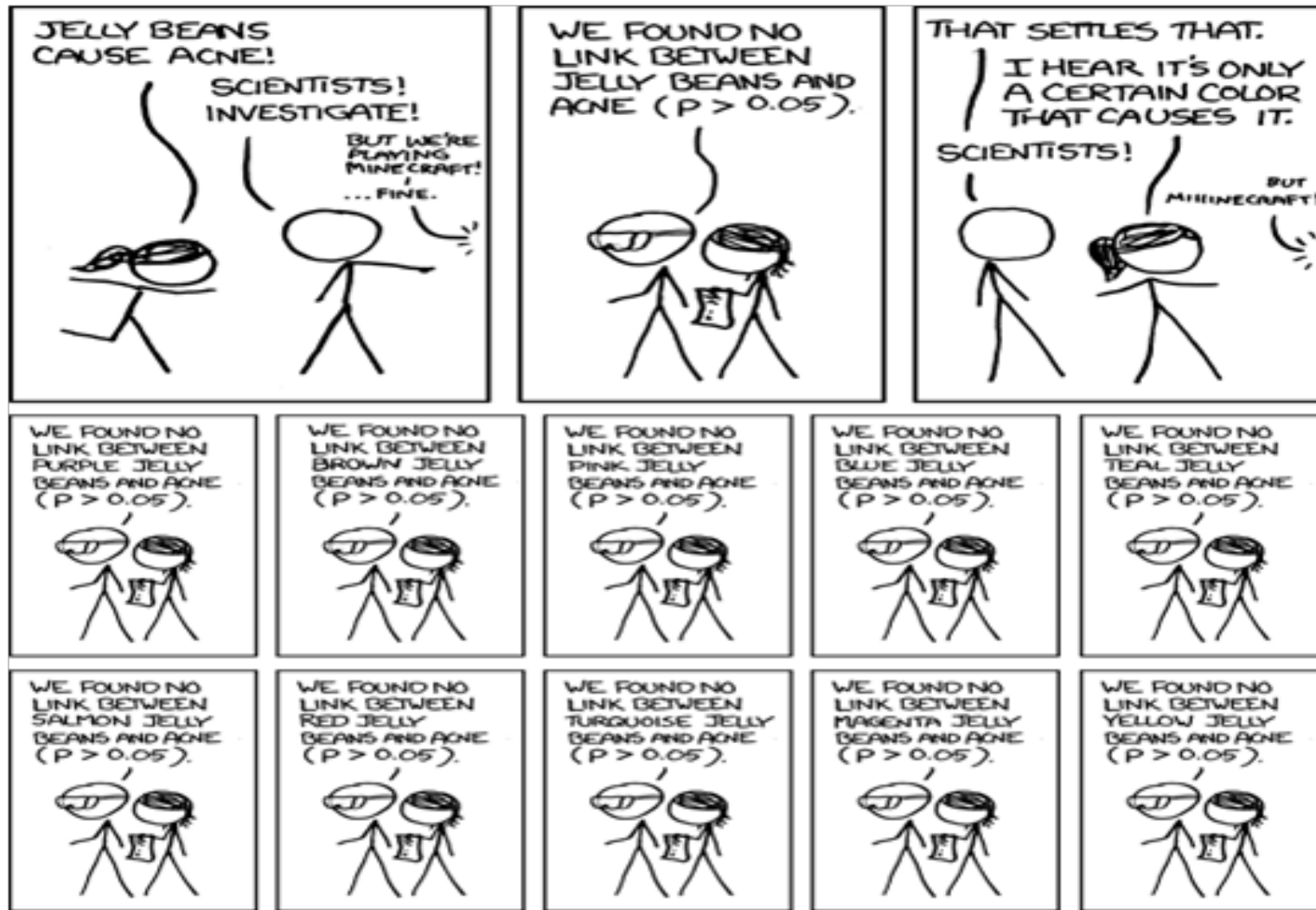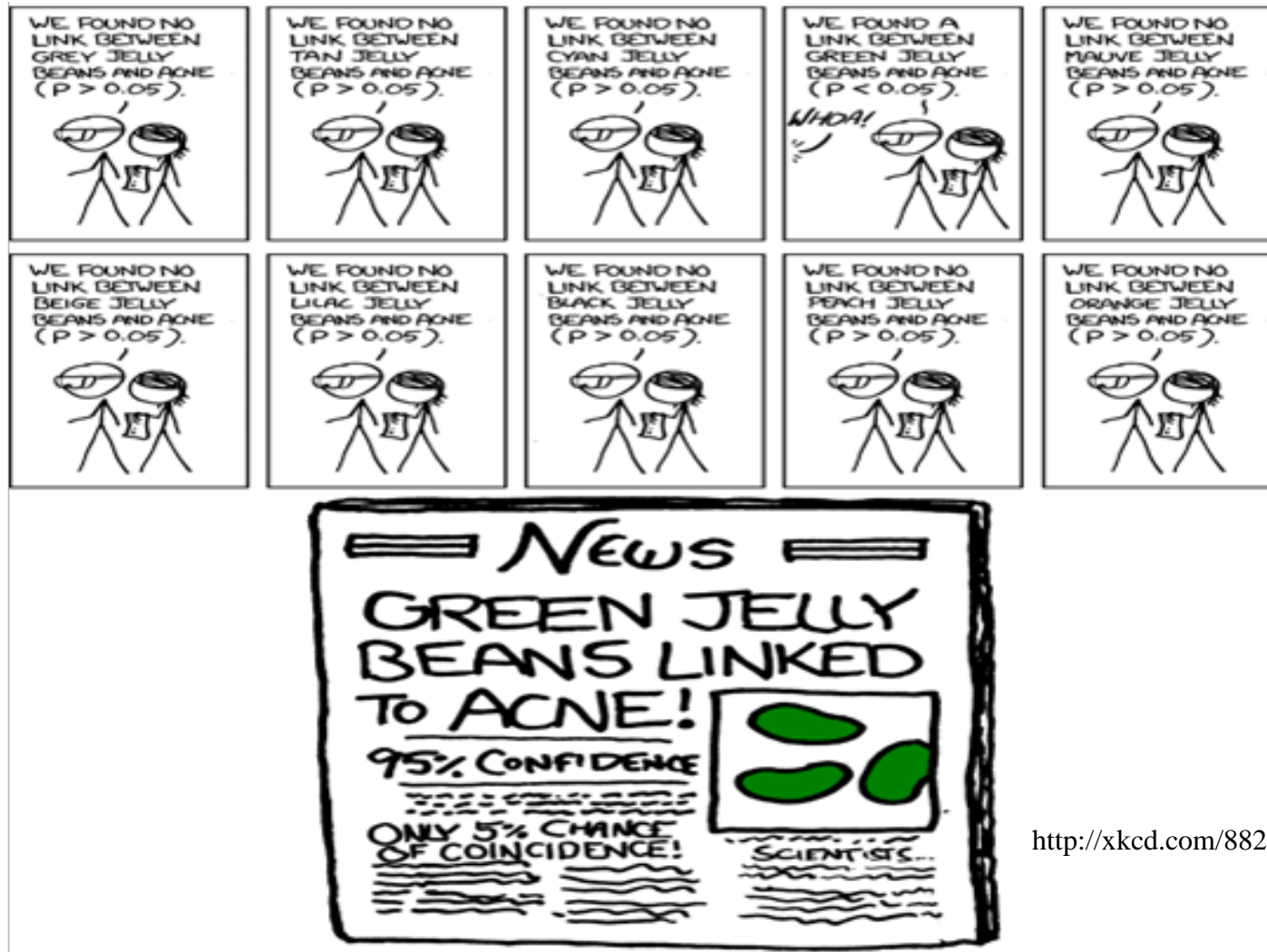
Stages of Smoking

Figure 2.3. Multiple Comparison Problem: Why we need to form hypotheses before looking at the data

P > 0.05 = 95% CI containing 1.00   P < 0.05 = 95% CI <u>not</u> containing 1.00

Figure 2.3, continued   **What makes the news?**

http://xkcd.com/882/

$P > 0.05$ = 95% CI containing 1.00 $P < 0.05$ = 95% CI <u>not</u> containing 1.00

Why People Smoke

Name: Barbara
Age: 55
Gender: female
Education level: College graduate
Job: Teacher

Barbara grew up in a culture where the dangers of smoking were off of the public radar. "It wasn't that big of a deal," she says. Despite this, Barbara had almost no exposure to cigarettes or smoking as a young girl. "I'm the third of four daughters, and neither my parents nor my older sisters smoked. I guess smoking just never seemed like something I wanted to do."

As a teenager, however, her attitude changed. "I remember spending summer afternoons with my best friend on the porch of our cottage at the lake. We used to sit there and smoke these long, thin cigarettes. It just seemed cool." Barbara says this small act of rebellion made her feel grown up and somehow more chic. "I actually hated the smoking, but the *idea* of smoking I liked." Barbara's smoking didn't last for long. "I smoked a few packs over the course of a summer or two, but it never really caught on for me. It didn't make me feel particularly good, and eventually I just thought it was stupid. So I stopped, cold-turkey. I guess I never really got hooked."

During college, plenty of Barbara's friends and acquaintances smoked occasionally, but she never felt much peer pressure to join in. "I never started again. I was pretty busy with work, but I was very happy, and the idea of smoking never appealed to me again." As far as she can remember, what pleasure she got from smoking cigarettes came purely from the image. "I don't remember a buzz at all, really."

Now in her fifties and the mother of four, Barbara hasn't touched a cigarette since her teenage summers, nor has she felt the craving to do so. "My husband is a doctor now, so I hear all sorts of horror stories about the dangers of using cigarettes. It makes me very glad I didn't get addicted."

Why People Smoke

Name: Mark
Age: 50
Gender: male
Education level: Ph.D.
Job: University professor

Mark spent his youth surrounded by people who smoked. He doesn't remember exactly why he started: "Smoking was just ubiquitous in my family and social circles growing up. It was before everyone knew it was stupid." When he was 17, Mark took up smoking, just as most of his friends and relatives had. "I really don't remember exactly why I began, other than that it was all around me. I guess I smoked about a half a pack a day, usually later in the day." Mark doesn't remember any particularly strong physical effects. "I would just get a physical craving that made me want to have a cigarette. I don't remember much of a buzz."

He has struggled several times with trying to quit. "My experiences varied. The first couple of weeks are always hard." Powerful cravings made quitting extremely difficult, but Mark has found nicotine gum to be very helpful in controlling cravings and allowing him to taper off nicotine gradually. "My problem is always relapse behavior. I'll quit for six months or so, and then for no particular reason, start up again. It's not rational at all. It's a yearning you can't control." It's particularly dangerous, Mark says, to be in a situation where others are smoking. "Someone will offer me a cigarette, and I'll think 'Hey, it's just one,' and then I'm right back where I started." Quitting permanently has not been easy. "The cravings always return, but usually less frequently with time. I haven't smoked for 3 months now, and I've promised myself that I won't start again. The hard part is, the cravings never go away."

Why People Smoke

Name: Kayley
Age: 16
Gender: female
Education level: In high school
Job: Student

Kayley, a 16 year old high school student, has her own reasons for smoking. "People say that kids smoke just to be cool or whatever, but that's bull. *I* smoke because I like it. It makes me feel good. After awhile you just get into that habit--I want something in my hand, I want the motion of smoking, of bringing my hand to my mouth. That's why I eat when I can't smoke." When discussing her friends' smoking behavior, Kayley returns to the idea of peer pressure. "Some of my friends don't smoke, and I'm not pressuring them to start or anything."

Kayley started smoking just last year. She says she was curious to see what it felt like. "I don't want to get addicted, but I figure if I don't do it that much, then I probably won't. Plus, if so many people do it despite the risks, then there's got to be something to it, no?" And so far, Kayley thinks there is. It feels good, she says. It relaxes her between classes or after her day, and it gives her "a little buzz." Kayley smokes 1–7 cigarettes a day, and sometimes none at all. "I probably average about two."

Kayley's father smoked as a younger man, and had some problems trying to quit, although she doesn't think he's had a cigarette in several years. Her mother didn't smoke at all. An only child, Kayley was the focus of her parents' "anti-smoking rant," and does believe that smoking is bad for her. "I'm not in denial; I don't think I'm invincible." Kayley notes, however, that there are lots of things adults do that are dangerous; they take a calculated risk. "People die from driving, too, but I'm not about to stop." Kayley says she doesn't intend to smoke all her life and thinks she'll taper off soon, maybe just smoking every so often. "I haven't really tried to quit yet, so I don't know what that would be like."

Why People Smoke

Name: Shirene
Age: 52
Gender: female
Education level: High school
Job: Prison guard

As a girl, Shirene was around smoking a lot, because both of her parents smoked. "It's funny—I can still remember going on car trips with my parents as a little girl. They both smoked in the car. I'd always get car sick, probably from the cigarettes. Almost strange to think that just a few years later I'd start smoking. On the other hand, maybe it makes sense." When she was 14, Shirene smoked her first cigarette. "It was cool, grown-up. Everyone I knew did it—all my friends, parents, relatives. Most of my relatives still smoke. The idea that smoking was "just something you did" definitely contributed to her smoking. But for Shirene, smoking also had a strong physical effect. The first cigarette she ever smoked, she says, "gave me a real buzz." And from there, the habit grew. "I started with just one cigarette a day, then it was three, then five, then half a pack, then a pack a day. Within a year or two I was totally hooked." In her young adulthood, short on cash, Shirene remembers hunting through old ashtrays, even the garbage, looking for butts that hadn't been completely smoked. "I wanted that buzz, to satisfy that craving, that much."

Shirene tried to quit smoking several times in the last four years, but it hasn't been easy. "I'll suddenly just *crave* a cigarette…. It's very powerful. It usually hits me when I'm doing something routine, like working at a computer. When I get a craving, it's very hard for me to focus on anything else." Shirene also thinks the pleasing menthol taste of her preferred brand may have made it even harder to quit.

A few years ago, Shirene had coronary artery bypass surgery. While in the hospital, she decided to quit cold-turkey. "I decided to quit once and for all. I couldn't smoke in the hospital, it hurt when I coughed, and when they tested my lung capacity, I couldn't perform at half the level they wanted me to. That scared me." Shirene hasn't smoked for 4 months now. Although she says that she still needs to stay away from other people who smoke, Shirene thinks it's going well. "Before my operation, I don't think I had it in my brain that I really wanted to quit. We'll see. I know some damage has been done, but it's better late than never. I still crave them, though."

Why People Smoke

Name: Donald
Age: 83
Gender: male
Education level: 10th grade
Job: Retired civil servant

As a boy in a small mining town, Donald started smoking at the age of twelve. He started mainly "because of peer pressure, to be a part of the crowd." Donald's mother was not a smoker, but his father smoked a pipe. "My parents weren't aware of my smoking, or at least didn't acknowledge it. I don't think they would've condoned it, either." Unaware of the health risks associated with smoking, he continued to smoke into his teens.

On his 17th birthday, Donald joined the army to fight in World War II and was sent to an army base in England. A government program allowed families to ship cigarettes inexpensively to their loved ones serving overseas. "They'd come a thousand at a time, sometimes once, sometimes twice a month," Donald recalled. Cigarettes served as a form of currency. They could be used to pay for services or goods, and even "led to black market-type stuff. I'd smoke cigarettes when I had them, and worry when I didn't. I can think of maybe one guy I knew at the time who didn't smoke." The soldiers must have had some idea that smoking was bad for their health, because the common saying on lighting up was "Put another nail in the coffin."

Returning from England at the end of the war, Donald started smoking less, partly because of the high "sin tax" on cigarettes. He quit for a while, but then started again when he was offered a cigarette at a dance. Donald smoked about a pack a day and would usually have his first cigarette around breakfast time. "It was just the habit that kept me going, although it was relaxing, too. I just got into the pattern."

At age 60, Dennis finally managed to kick the habit, quitting outright. He thinks that the personal decision to quit is the only way to stay off cigarettes, rather than doing it to please someone else. "I missed it at first, but I didn't have any of the physical effects like headaches."

Why People Smoke

Name: Drew
Age: 18
Gender: Male
Education level: In high school
Job: Student

Drew has smoked for two years, but his exposure to smoking began long before his first puff. "My mom bought me candy cigarettes when I was little, and I'd put one in my mouth and tell my mom to light me up." It was a natural thing for Drew to do, as both of his parents have smoked for as long as he can remember. His parents didn't even object much when Drew started smoking at 16, and they haven't been pushing him strongly to quit. "Other kids have it hard, man, trying to hide. Everyone in my family smokes and all my friends too. Starting was easy." Although legally unable to buy cigarettes, Drew found them easy to come by when he started. He could get them from older friends or from parents, and many stores would sell to him without checking his ID. With ready access to cigarettes, Drew's habit escalated quickly. He went from smoking a few cigarettes a day at first to a pack or more a day—his current level.

Drew says that because all of his friends smoke, he really hasn't had a powerful social incentive to quit. Nevertheless, he has tried to quit 3-4 times, but has failed to stay off cigarettes for more than a few weeks. "If I can't have a cigarette, I get headaches, and after a while longer I get hungry," he says. Bad moods and intense cravings also contribute to his difficulty quitting. "I'll get stressed or angry, and I just lose the nerve. After I have one, I'm back to where I started in no time."

Why People Smoke

Name: Sam
Age: 25
Gender: Male
Education level: College graduate
Job: 8th grade math teacher

Sam has smoked for about six years. "I had smoked a few times in high school, but I really started smoking in college. I was stressed all the time and I had to stay up late. My roommate smoked and told me that cigarettes help him focus and stay awake. So I started smoking with him to try it, and found out he was right." Within a few months, he was addicted. It was easy to stay that way, too. Sam was a fraternity member and the basements of the fraternity houses were always full of cigarette smoke. "No one there batted an eye when you lit up. I think it makes it way easier if smoking's acceptable in the places you hang out." When the cold winter weather arrived, Sam found his desire to have a cigarette so powerful that he would brave the zero degree weather for a smoke every few hours during the day, when the fraternity houses were closed. "I'd be standing there telling myself this is crazy--I want this so much I'll stand out here and freeze."

Sam hasn't tried to quit yet. "Look, I know smoking's bad. I haven't been doing it for that long, and I know it's having an effect. Fortunately I'm not a chain smoker or anything. I usually just smoke two or three cigarettes a day." Already Sam has a hard time doing anything physically strenuous. "But it's not something I'm ready to give up yet. I just can't."

Why People Smoke

Name: Tessie
Age: 16
Gender: Female
Education level: In high school
Job: Student

Tessie has smoked for about a year and a half, despite heavy pressure from her parents to quit. Tessie is adopted, and although neither of her parents smoke, her mother's father and brother both smoked and died from lung cancer. "My mom's really scared about smoking because of that," Tessie says. In spite of her parents' attitude, Tessie found smoking cool and sexy. Most of her friends smoked too. "I just kinda started because it was around me; it was the thing to do." Tessie gets her cigarettes from her sister and from convenience stores.

At first, Tessie wasn't too concerned about becoming addicted, but after a year and a half, it has become much more difficult for her to abstain from smoking. "I thought I could go three or four days without a cigarette, no problem, but I can't. In the morning, I *need* one." Tessie recently had the opportunity to see firsthand, in a pathology lab, the damage done to lung tissue and arteries by cigarette smoking. "It really horrified me, I started crying." Tessie has also begun to notice the physiological effects in herself. "I saw a girl who smoked wearing a letterman's jacket, and thought, 'You can do athletics?' After just a little bit of exercise, my lungs start to go out and I just can't breathe."

Tessie is currently trying hard to quit, but the withdrawal symptoms make it hard. After only six hours without a cigarette, Tessie feels intense pain in her temples, and she becomes irritable. "Sometimes I'll catch myself just staring at a cigarette, thinking how much I want it. I'm used to smoking almost a pack a day, so stopping would be a huge change for me." Tessie's attraction to cigarette smoke is so powerful that she finds herself drawn to smoking sections in restaurants, just to get a whiff of the smoke.

Why People Smoke

Name: Laura
Age: 52
Gender: Female
Education level: College graduate, some graduate work
Job: University administrator

Laura started smoking at age 17 for several reasons. "It was partly due to the influence of a friend who smoked, and partly because of the stress brought on by my mother's death." At this stressful and emotional time in Laura's life, she turned to cigarettes because they relaxed her and helped her deal with stress. "Smoking makes me feel calm and restored," she says. "My mother and grandmother smoked. I associate warm and comforting environments with coffee and cigarettes."

Now a regular smoker who usually starts in the morning and smokes a half a pack to a pack a day, Laura has been smoking for 35 years. Laura feels some of the physiological effects of smoking. "I can't run for very long," she says, though she hastens to add that she can still hike, and usually keeps up with non-smokers.

Even though she is aware of the dangers of smoking, Laura hasn't quit. "I've tried to quit many, many times. It's very hard. The longest I ever made it was six months. I drank huge quantities of tea and changed my diet to lighter fare. I found heavier foods made my nicotine cravings worse." Laura has relapsed every time she has tried to quit. The first relapse occurred after a stressful cross-country move. Now, she regularly tries to quit, but usually lasts only 2-6 weeks. "Whenever I start smoking again, it's not really the stress that makes me go back. It's mainly the withdrawal that is just so hard." Laura gets intense feelings of anxiety and discomfort during withdrawal, and finds herself very distracted. "The discomfort is hard to explain, but it's very powerful."

Though Laura has not managed to quit yet, she has had a lot of experience with different methods. She has tried nicotine substitutes such as the patch and gum, the anti-depressant Wellbutrin, which can reduce withdrawal symptoms, and hypnosis, all with little success. "For me, water, hot tea, and the desire to quit are the strongest non-smoking aids."

Why People Smoke

Name: Aaron
Age: 18
Gender: Male
Education level: In high school
Job: Student

Aaron started smoking three years ago because he thought it would help him with schoolwork. "I was doing really bad in math, and I figured it would help me relax. Also, a couple of my friends were doing it, and I kinda did it just to blend in." Aaron found the taste of cigarettes somewhat unpleasant, so he smokes menthol cigarettes, going through a pack every couple of days. Aaron is eighteen and can legally buy his own cigarettes now. But for the three years he smoked as a minor, it was surprisingly easy for him to get cigarettes.

Smoking is something Aaron does "under the radar" of his parents, although he thinks they are aware that he smokes. Neither of his parents smoke, but yet they haven't talked to him much about it either. During school, Aaron smokes between classes, during breaks, or at lunch. He says that his craving to smoke is so powerful that "there's no way I can quit." Cigarettes help Aaron focus and relax; without them he gets distracted, irritable, and fidgety. If he goes for more than a few hours without a cigarette, he gets an intense headache. "It doesn't go away, either, and all I can think about is having another cigarette." Aaron is trying to quit, "but these headaches are killing me."

Why People Smoke

Name: Frances
Age: 74
Gender: Female
Education level: Some college
Job: Retired

Frances has smoked for most of her adult life. "I started smoking so long ago, I can't quite remember exactly why I started. It was something that all my friends did, and my parents too occasionally, but I don't remember ever feeling pressured. It was almost as though it was something that was expected as you grew older." She remembers the early advertisements for cigarettes, and the appeal of the sexy women with long, thin cigarettes clasped in their fingers.

The physical effects of smoking on Frances have been profound. She now has severe emphysema and can't go anywhere without her oxygen tank. Tubes from the tank pass through her nostrils, carrying much needed oxygen to her lungs. She spends most of her time at home, or walking slowly through her garden. Despite all this, Frances still smokes. "It's one of the last pleasures left to me. I know I won't be around for long, and the damage is done, so I figure there's no sense in quitting." She has been smoking so long now, she says, that there isn't a strong feeling when she lights up. "It just feels... good. Calming, pleasurable. And it's unthinkable that I could stop, now."

In the early days of Frances' smoking, the harmful effects of smoking weren't known, but as more and more research revealed the dangers, she tried several times to quit. "Every time I'd come back to cigarettes. Sometimes I'd tell myself that I'd just have one, but a little slip would be all I'd need to head back right to where I'd been, and within a few days I'd be back to a pack a day," she says. Frances expects she will smoke until the day she dies.

References

Batra, Vikas, Ashwin A. Patkar, Wade H. Berrettini, Stephen P. Weinstein, and Frank T. Leone. "The Genetic Determinants of Smoking." *Chest* Volume 123 No 5 May 2003.

Bricker, Jonathan B., Brian G. Leroux, Arthur V. Peterson, Jr., Kathleen A. Kealey, Irwin G. Sarason, M. Robyn Andersen and Patrick M. Marek. "Nine-year prospective relationship between parental smoking cessation and children's daily smoking." *Addiction* Vol 98 Page 585, May 2003.

Li, Ming D. "The Genetics of Smoking Related Behavior: A Brief Review." *The American Journal of the Medical Sciences*. October 2003 Volume 326 Number 4.

Lloyd-Richardson, Elizabeth E., George Papandonatos, Alessandra Kazura, Cassandra Stanton, and Raymond Niaura. "Differentiating Stages of Smoking Intensity Among Adolescents: Stage-Specific Psychological and Social Influences." *Journal of Consulting and Clinical Psychology* 2002, Vol 70 No 4, pp 998-1009.

National Institute on Drug Abuse Research Report Series, "Tobacco Addiction" July 1998, revised July 2006. NIH Publication Number 06-4342. http://www.drugabuse.gov/PDF/RRTobacco.pdf

Rhee, Soo Hyun; John K. Hewitt; Susan E. Young; Robin P. Corley; Thomas J. Crowley and Michael C. Stallings. "Genetic and Environmental Influences on Substance Initiation, Use, and Problem Use in Adolescents." *Archives of General Psychiatry* Vol 60, Dec 2003.
Vink, Jacqueline M.; Gonneke Willemsen, and Dorret I. Boomsma. "Heritability of Smoking Initiation and Nicotine Dependence." *Behavior Genetics* Vol 35 No 4 July 2005.

Woodward, Kristin. "Quit before kids reach third grade." Fred Hutchinson Cancer Research Center Science Article. May 15 2003.

World Health Organization. World No-Tobacco Day, 31 May 1998. Geneva, Switzerland: World Health Organization, 1998.

Lesson 3. How do genes influence smoking behavior?

Overview
Students learn how nicotine affects the brain and consider how genes involved in the reward pathway might affect smoking behavior. The genes in the UW Smoking Behavior Research Study are introduced, and students discuss why they are included in the study.

Learning objectives

Objective	Evidence of Learning
Students will understand that certain genes might contribute to variability in smoking behavior and how this process might occur.	Students are able to draw and explain how a genetic change might change neurons in the brain and therefore affect a person's physiology.
Students will understand that there are genetic and environmental components inherent in smoking behavior.	In research teams, students explain possible genetic components that may affect their research topics.

Lesson Preparation
Time
70 – 90 minutes

Prerequisite concepts
This curriculum assumes that students understand the following concepts at an introductory high school level. You may need to review these concepts with your students prior to starting this activity.
1. Mendelian genetics
2. Gene expression: DNA codes for RNA, which codes for protein
3. What a gene is
4. What a protein is
5. How changes in a gene can result in a change in a protein, which can affect its function
6. Genetic variation (diversity in the inherited traits within a species)

Materials
for each student
Student Sheets 3.1, 3.2, 3.3, and optional 3.4

Copies of PowerPoint slides (optional)

for class
Neuron model
Copies of Diagram 3.1 and 3.2 for overhead projector or document camera

Getting Ready

1. Photocopy Student Sheets 3.1, 3.2, 3.3, and 3.4 (optional).
2. Download Lesson 3 PowerPoint from website.
3. To prepare a packet of PowerPoint notes pages for each student, follow the instructions below to print a master document you can photocopy:

    a. With the PowerPoint presentation open on your computer, use the mouse to select "File / Print."
    b. Select "Handouts" in the "Print what" pull-down menu in the Print dialog box.
    c. You can choose how many slides to print per page to the right of the Print what menu. Printing three slides on a page leaves room for students to take notes next to each slide.

4. Make sure you can access the database on your class computer and project it for the class to view: https://gsoutreach.gs.edu/database2
5. Review the teacher background notes.
6. Assemble the neuron model (using teacher background notes).

Presenting the Lesson

A. Neurotransmission and the Effects of Nicotine on the Brain

A.1: Tracing Nicotine's Path through the Body

1. Ask students to work in pairs for this activity.
2. Give them Student Sheet 3.1, and ask them to draw where nicotine goes in the body after a person takes a puff on a cigarette.
3. As you discuss the pathway with the class, draw it on the diagram:
    - From the mouth, down the trachea to the lungs
    - From the lungs into the bloodstream to the heart
    - From the heart to all parts of the body, including the brain
    - From the blood to the liver, where it is broken down and excreted from the body through the kidneys
    - **Make sure that students recognize that nicotine goes to the brain.**

A.2: The Giant Neuron and Overview of Neurons and Neurotransmission

1. Give each student a copy of Student Sheet 3.2.
2. Explain to students that they need to understand some basics about neurons and neurotransmission.
3. Use Diagram 3.1 to discuss the parts of the neuron (cell body, dendrites, axon, axon terminals, vesicles of neurotransmitter, and receptors) and the process of neurotransmission. Be sure to discuss how information is transmitted during neurotransmission— via an electric impulse within a neuron and via neurotransmitters between neurons.
4. As a class, use the model neuron to show how neurotransmission works.
5. You can assign people to have different roles. You may need to demonstrate neurotransmission more than once so everyone gets a turn.

6. Use large post-its to label the different cell parts (dendrites, cell body, and axon) on the neuron model, and ask students to label the sketch of the model neuron on Student Sheet 3.2.

A.3 PowerPoint Presentation: How Nicotine Interacts with the Brain
1. Give students the optional PowerPoint handouts, if you plan to use them.
2. The PowerPoint presentation introduces students to the concept that different parts of the brain have different functions. There is even a part for reinforcing behaviors that are helpful for survival. This is called the **reward pathway**.

Teacher Note
Students can become overwhelmed in the details of the reward pathway and the names of enzymes. Focus on the main concept—that genetic variation in the human population results in slight differences in the ways our bodies interact with chemicals like nicotine and can lead to variation in smoking behavior.

---

3. After completing the PowerPoint presentation, as a class, use the model neuron again, this time to model how nicotine interacts with neurons in the reward pathway.
4. Ask students which ping pong balls represent the nicotine (*the ones tossed to the students holding the "dendrite" ropes*) and which are the dopamine (*the ones in the "axon terminal" bucket*).
5. As a class, discuss where dopamine receptors are found (*on the hands of the students catching the balls at the axon end of the model, which represent the dendrites of the next neuron*) and where dopa decarboxylase is found (*in the bucket at the end of the rope, which represents the axon terminals where dopamine is released*).

Teacher Note
You may ask students to discuss some of the strengths and weaknesses of the giant neuron as a model for nerves and neurotransmission.

B. Viewing genetic questions in the database

1. Access the database from your class computer and project it for the class to see
    https://gsoutreach.gs.washingthon.edu/database2/
2. Provide students with Student Sheets 3.3.
3. As a class, view Questions 103-105 in the Hypothesis Testing view of the database to see what responses are available for each question.
4. Complete the responses to the questions in Student Sheet 3.3 as a class.

Teacher Notes
The questions in Student Sheet 3.3 are intended to guide student understanding of how variation in the three gene regions in the study might affect smoking behavior. This concept can be difficult for students, so this activity is best done as a class.
The intention of Question 3 is to show that everyone has two copies of the *DRD2* gene, one on each chromosome 11.

Teacher Notes, continued
The purpose of Question 4 is to help students understand that different genotypes might affect the amount of DRD2 receptor made.
In Question 5, students are then asked to consider how making less DRD2 receptor might affect their smoking behavior.

C. How do scientists choose which genes to include in a study?
   1. Show Diagram 3.2 to the class. As a class, discuss how a scientist might select good candidate genes.

Discussion Questions:
How do you think scientists choose genes to study?
Students should discuss:
- They choose genes that make sense biologically. In order to do this, they need to understand something about the process they are studying.
- Scientists often look at other research studies that have been done by themselves or others. This often means reviewing the published literature to find out what has been learned on the same or a related condition.

Based on what we've learned about nicotine and the brain, how might DRD2 and DDC affect smoking behavior?
Students should discuss that DRD2 and DDC are necessary in the reward pathway, and variation in these genes may be associated with variation in the effect of nicotine on the brain.

Based on what you gathered from the smoker profiles, do you think smokers are likely to have different variations of these genes than non smokers?
Students should mention the variation in way the profiled smokers reacted physiologically to smoking. For instance, some people were able to quit easily and some are unable to quit, or some people felt a buzz or felt relaxed while others did not report these feelings related with smoking.

D. Could your research topic/factor have a genetic component?
   1. Ask each group to complete RPP-4 of the Research Project Pages: Genetic Component to Research Project. They should discuss their responses as a group, but each student should complete their own sheet.

Peer Review
Have each student pair up with someone from a different research team to discuss if there are possible genetic components to their research topics. They should each share their research topic and explanation for why there might be a genetic component. They should ask each other if the explanations make sense and if there are alternative ways genetics could be involved.

Homework: If it is appropriate for the level of your class, ask students to complete Student Sheet 3.4 for homework.

a. A typical neuron
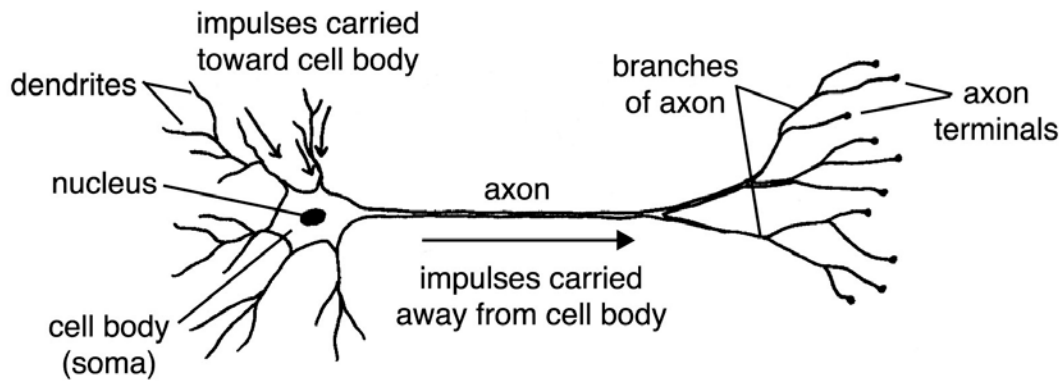
b. Close-up of a neural synapse
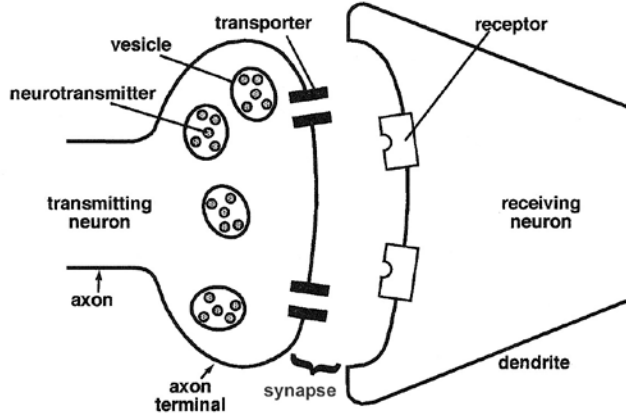
c. Neurotransmission
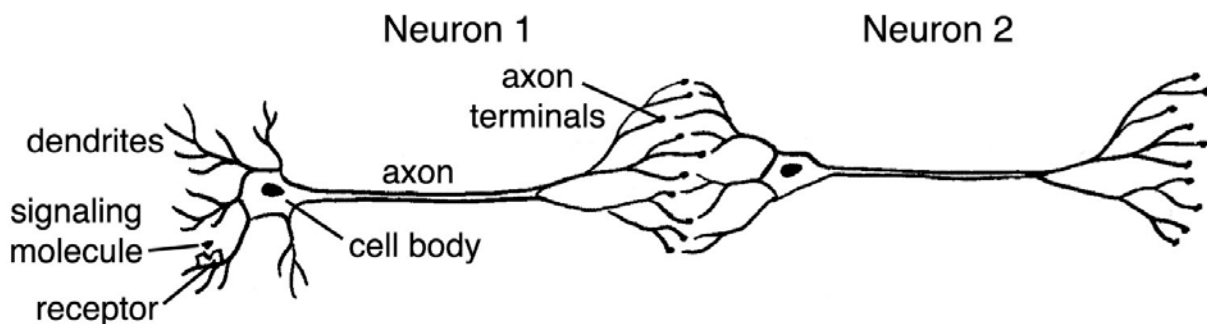
Diagram 3.1: Neurons and neurotransmission

DRD2 gene

Direction of transcription
of *DRD2* gene

Promoter   Transcribed region

-141C insertion/deletion
Insertion/deletion of C in
promoter 141 bases before the
start of transcription

C957T
C to T at nucleotide position 957 is a
"silent" mutation (no change in amino
acid sequence of protein)

DDC Gene

Promoter   Transcribed region

G to A in Intron 9
G allele may be part of an intronic
splicing enhancer involved in
regulation of splicing

Diagram 3.2: Candidate Genes in Smoking Behavior Study

35

Possible Responses to:
STUDENT SHEET 3.3: **Genetic Exposures in the Smoking Behavior Study**

As a class, look at Question 103 in the Hypothesis Testing view of the smoking behavior database. Under "3. Identify answers corresponding to your exposure," there is a list of four possible answers for the C957T region of the *DRD2* gene. Work with your class to answer the questions below.

1. How many possible genotypes (combinations of genes) are there for this gene region?
   There are three possible genotypes, C/C, C/T, and T/T.

2. Why does each genotype have two letters associated with it (C/C, T/T, C/T)?
   People have two copies of each gene, one inherited from their mother and one from their father.

3. *DRD2* is on the long arm of chromosome 11, shown by the arrow at the left in the figure below. Draw the genotypes on the chromosome pairs, as shown in the example at the left. Why are there two chromosomes for each genotype?

| C/C | T/T | C/T |

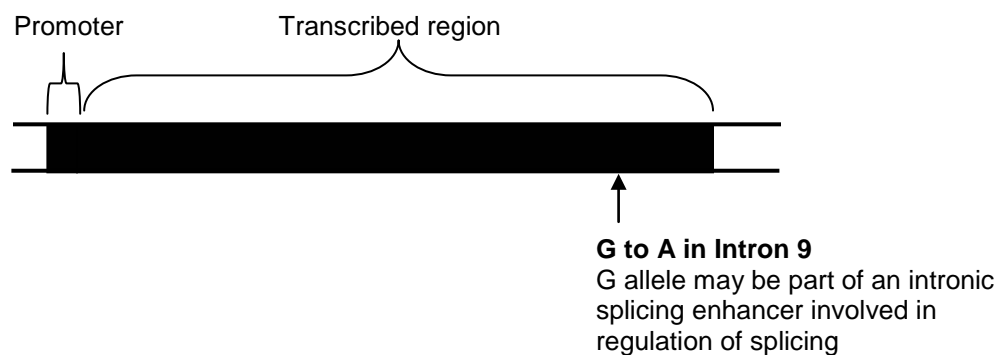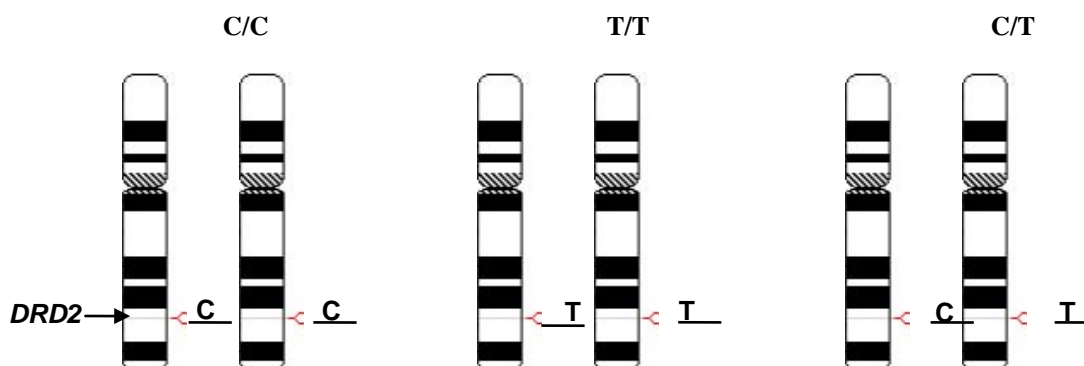*There are two chromosomes for each genotype because we have two copies of each chromosome.*

4. *DRD2* codes for the dopamine receptor. The T allele of the C957T gene region results in fewer receptors than the C allele. Draw the number of receptors you expect for each genotype in the diagrams below. One receptor is shown in each figure.
   C/C would have the most receptors, T/T the fewest, and C/T an intermediate amounts
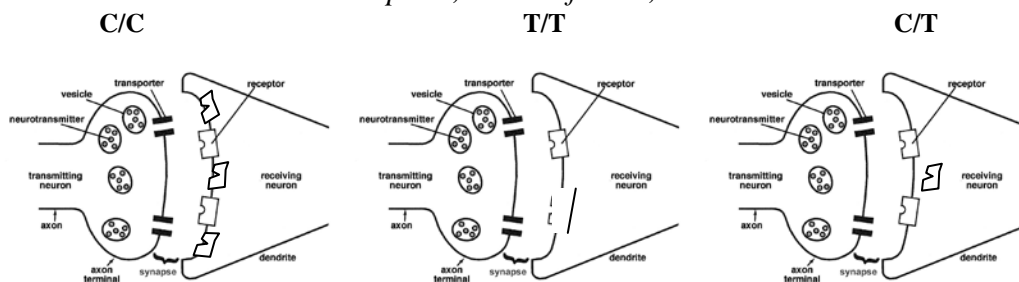
| C/C | T/T | C/T |

5. You just learned that the *DRD2* gene codes for the dopamine receptor. What is the function of this protein? How might making less of this protein (compared to "normal")

affect a person's smoking behavior? (Hint: Think about the giant neuron and the effect of nicotine on dopamine release).

The dopamine receptor is located on the post synaptic neuron, and it binds dopamine that is released into the synaptic cleft. This has an effect on neurotransmission on that neuron. Making less of the dopamine receptor may result in a person who is a smoker smoking more cigarettes than someone who has more receptors in order to feel the effect of nicotine.
Note: Students may also say that people who make less of the dopamine receptor are less likely to become smokers because they may not be able to stimulate their reward pathway. There is no right answer to this question—students should provide a reasonable justification for their response.

6. What protein does the *DDC* gene code for? What is the function of this protein? How might making less of this protein (compared to "normal") affect a person's smoking behavior?
DDC codes for dopa decarboxylase. This enzyme is part of the dopamine synthesis pathway. Making less dopa decarboxylase may result in a person having less dopamine in their neurons. People who make less dopamine may smoke more than people with a normal amount so that they can stimulate their reward pathway.
Note: Students may also say that people who make less DDC are less likely to become smokers because they don't get much stimulation of their reward pathway from smoking. There is no single correct answer to this question—as in question 5; the important thing is that students provide a reasonable justification for their response.

7. Several of the questions in the questionnaire may have a physiological/genetic component to them, even though there is not a specific gene mentioned. Identify at least three questions that may be asking about something that is partially or largely genetic, and discuss why you think this for each question.
Possible responses include:
- *Q18. Each part of this question refers to a physiological sensation as a result of smoking, and these sensations would be determined mainly by genetic factors.*
- *Q26 and 28. Having parents or siblings who smoke may also have a genetic component because we share many of the same gene alleles with our parents and siblings.*
- *Q67. Frequency of drinking alcohol. There may be a physiological aspect to whether a person drinks or smokes, and these may be related.*
- *Q68-70. Satisfaction with life, stress, and self esteem as a teen may have both physiological and environmental components.*

Teacher Background
Making a Model of a Neuron
(from Eric Chudler, Neuroscience for Kids, http://faculty.washington.edu/chudler/chmodel.html)

1.1   As shown below, the model consists of two buckets taped together at their open ends to form the cell body. Short ropes protrude from one end to represent dendrites, and a long rope at the other end is the axon. A single bucket at the end of the long rope represents the axon terminal, and a pool float on this rope is used to simulate the electrical impulse (action potential). Colored balls are neurotransmitters. Three different colors should be used: balls of color 1 = Neurotransmitter 1; balls of color 2 = Neurotransmitter 2; balls of color 3 = nicotine.
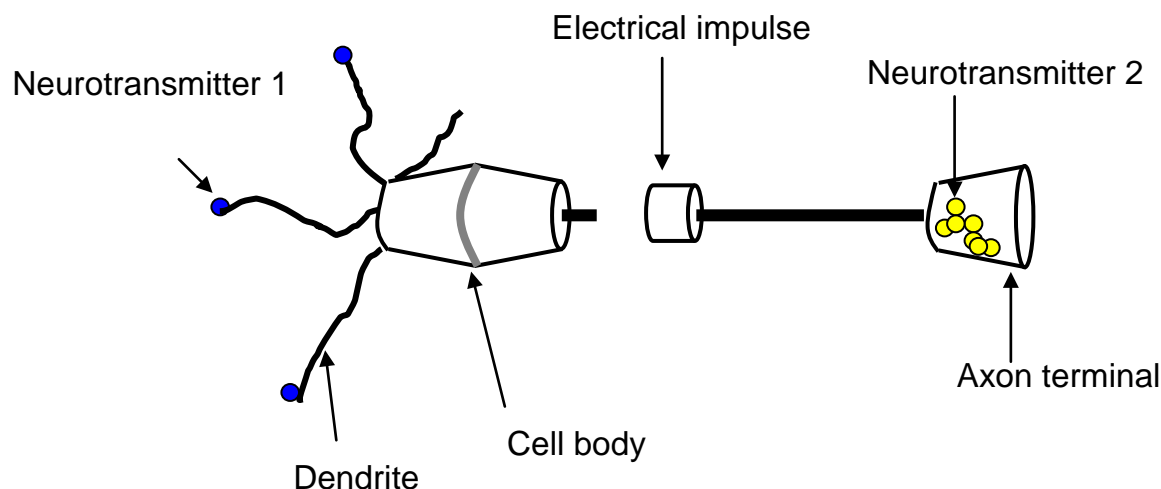
How to use the model:

1. Have students hold the following parts of the model: Student 1: Neurotransmitter 1 (plastic balls of one color); Students 2-4: one of the dendrites (thin ropes hanging from cell body); Student 5 cell body (double bucket); Student 6: electrical impulse (crab float); Student 7: axon terminal (end bucket containing Neurotransmitter 2, colored balls of a second color. Make sure student holds bucket on its sides, not its end, and pulls it so the rope is taut); .Students 8-10: ready to catch Neurotransmitter 2 (wear safety glasses).

2. Have the person holding molecules of neurotransmitter 1 TOSS the plastic balls to the people who are dendrites. The "dendrite people" try to catch the plastic balls. *This models the release of neurotransmitters and the binding of neurotransmitters to receptors on dendrites.*

3. When three plastic balls are caught by dendrites, the person holding the pool float propels it forcefully down the axon. *This simulates the conduction of an electrical impulse down the axon.*

4. The pool float should speed down the axon toward the axon terminal where it will slam into the container. This should cause the plastic balls to fly out of the bucket. *This simulates the release of neurotransmitter from the axon terminals.*
5. Have the students at the end of the axon terminal try to catch the balls as they fly out of the bucket. *This simulates neurotransmitter binding to the receiving neuron.*

CAUTION: The pool float will travel very fast! Make sure that the person holding the axon terminal keeps his or her fingers and hands AWAY from the end where the pool float will hit.

Nicotine, Neurotransmission, and the Reward Pathway

The Reward Pathway

Certain behaviors that are important for survival (e.g. eating, drinking, and sexual activity) are reinforced by the body through the reward pathway (Figure 3.1). These behaviors stimulate a specialized set of nerve cells in the brain to create the sensation of pleasure. Pleasurable feelings encourage repetition of the behavior. Three important areas of the reward pathway are the **ventral tegmental area** (VTA), located just above the brain stem, the **nucleus accumbens** (NA), a part of the emotional center of the brain, and the **prefrontal cortex**, which is important for thinking. To understand how the reward pathway works, we need to discuss the process of neurotransmission.
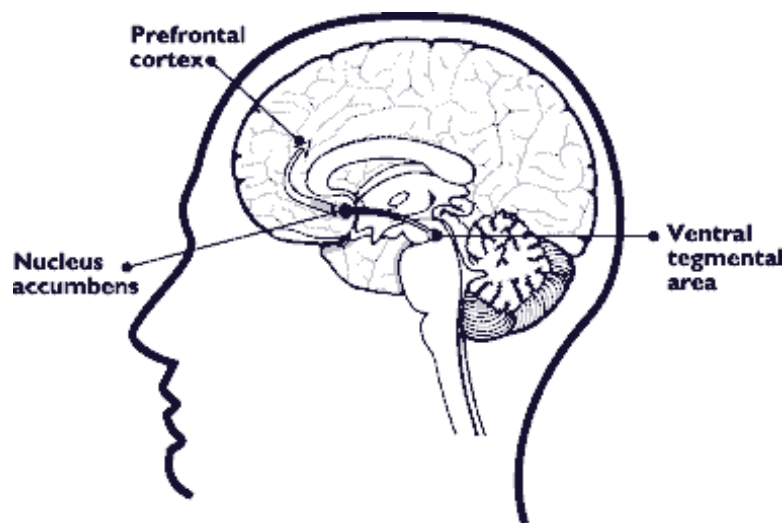
Figure 3.1. The reward pathway in the brain

Neurotransmission

There are billions of nerve cells, or **neurons**, in the brain. As shown in Figure 3.2, each neuron is made up of three parts: the central **cell body**, which directs the activity of the cell; **dendrites**, the short fibers extending from the cell body that receive messages from other neurons; and the **axon**, the single, long fiber that extends from the one side of the cell body and transmits messages to other neurons or other tissues, like muscle.
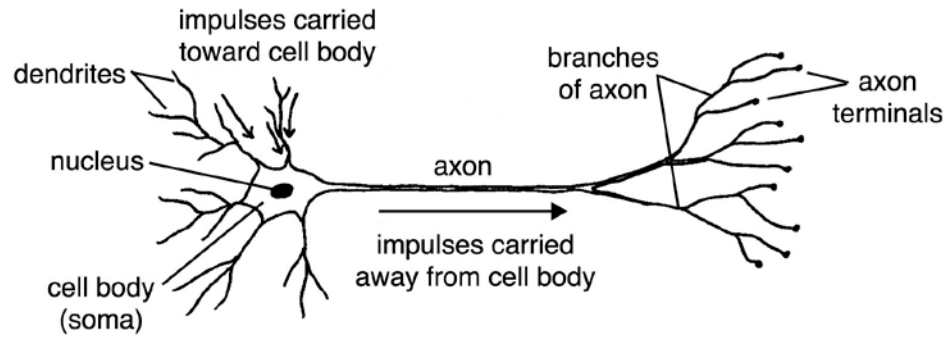
Figure 3.2. A typical neuron

To send messages, neurons use two types of signals, electrical and chemical. Electrical impulses are used to transmit messages from the cell body down the length of the axon to the axon terminal. Chemical messages are used to communicate with other cells, either neurons or muscle cells.

Neurotransmission is the process of transferring a message from the axon of one nerve cell, across a small space or **synapse**, to the dendrites of a nearby neuron. Messages between cells are transmitted by a chemical substance called a **neurotransmitter**. When a neuron is activated, it sends an electrical impulse down the length of its axon to the axon terminal. As shown in Figure 3.3, the axon terminals of one neuron, called the pre-synaptic neuron, lie close to the dendrites of another neuron, the post-synaptic neuron. (There are usually many other neurons, although only one is shown in this figure.) When the electrical impulse reaches the axon terminals, it results in the release of neurotransmitter into the synaptic space. The neurotransmitter molecules bind to receptors on the dendrites of nearby neurons, resulting in stimulation or inhibition of an electrical impulse in that cell.
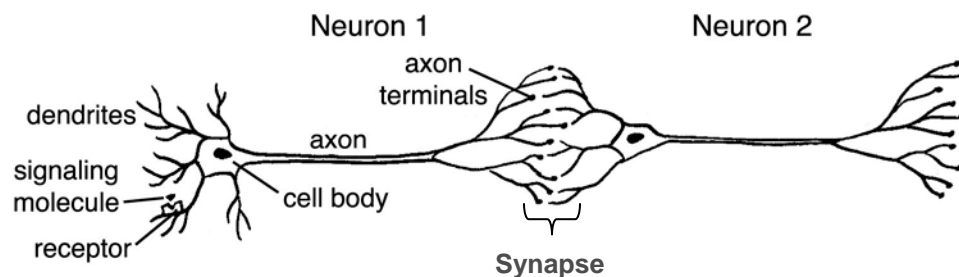
Figure 3.3. A neural synapse

A close-up of the synapse is shown in Figure 3.4. Normally, neurotransmitter is stored in membrane-bound sacks called vesicles in the terminals of the axon. When an electrical impulse reaches the axon terminals, the vesicles fuse with the cell membrane, releasing the neurotransmitter into the synapse. The neurotransmitter crosses the synapse and binds to its specific receptor on the dendrites of a nearby nerve cell. The binding of the neurotransmitter results in either stimulation or inhibition of an electrical impulse in the receiving cell. After being released into the synapse, the neurotransmitter is quickly inactivated. Some neurotransmitters

are broken down by enzymes; others are reabsorbed by the axon that released them through special transporter molecules located on the cell membrane.
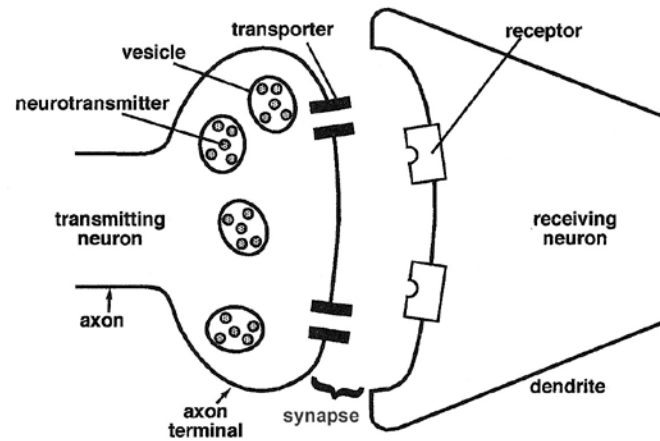
Figure 3.4. Close-up of a neural synapse

Details of the Reward Pathway
The reward pathway consists of a few thousand neurons that have their cell bodies in the VTA and their axon terminals in the NA or prefrontal cortex. These neurons use the neurotransmitter, **dopamine** so they are called **dopaminergic** neurons. Stimulation of dopaminergic neurons in the VTA results in the release of dopamine at axon termini in the NA and prefrontal cortex. Some of this dopamine binds to **dopamine receptors** on the dendrites of neurons on the other side of the synapse. Dopamine binding results in a series of changes in the receiving neuron, including the stimulation or inhibition of an electrical impulse. Dopamine does not remain in the synapse for long—it is quickly reabsorbed by the cell that released it by a dopamine-specific pump called the **dopamine transporter**.

1.2 **How Does Nicotine Interact with Neurotransmission and the Reward Pathway?** Drugs of addiction have a reinforcing effect because they interact with the reward pathway. When cigarette smoke is inhaled, the nicotine in the smoke is absorbed into blood vessels in the lungs and reaches the brain within 10 seconds of the first puff. Once in the brain, nicotine binds to receptors located on the cell bodies of neurons in the VTA area, as well as the terminals of these neurons, which are in the NA and the prefrontal cortex. Normally, these receptors bind the neurotransmitter acetylcholine. Nicotine is similar in structure to acetylcholine, so it is able to bind to the acetylcholine receptor. To simplify this discussion, these receptors will be called nicotine receptors, but the normal function of these receptors is to bind acetylcholine.

The binding of either nicotine or acetylcholine to the receptor results in the brief opening of a pore in the receptor. This opening allows cations to move into the neuron. This influx of cations triggers an electrical impulse to move down the nerve axon.

At the terminals of the neuron, dopamine is released into the synapse. It binds to dopamine receptors on nearby nerve cells in the nucleus accumbens and the prefrontal cortex (Figure 3.5). The binding of dopamine to dopamine receptors in these regions of the brain is associated with

41

the rewarding effect of smoking often reported by smokers, such as relaxation or an increased ability to focus.

Transmitting neuron   Receiving neuron

1. Nicotine binds to receptors on transmitting neuron in the VTA
2. Electrical impulse travels along axon away from cell body to terminals in nucleus accumbens and prefrontal cortex
3. Vesicles of dopamine fuse with cell membrane
4. Dopamine is released into synapse
5. Dopamine binds to receptor on receiving neuron
6. Increase or decrease in electric impulse along receiving neuron
7. Dopamine transporter takes up dopamine from synapse

Figure 3.5. Effect of nicotine on neurons in the reward pathway

Metabolism of Nicotine

As discussed earlier, nicotine binds acetylcholine receptors on neurons in the brain. However, unlike acetylcholine, nicotine is not broken down in the brain, but instead is processed in the liver and excreted in the urine.

The liver carries out many important functions, including clearing toxic substances from the blood. It produces a class of enzymes called the **cytochrome P450**s that help to break down toxins by making them more water soluble so that the body can excrete them in the urine (see Figure 3.6). The cytochrome P450 enzyme CYP2A6 converts nicotine to a substance called **cotinine**, which is modified by other enzymes and then excreted in the urine. About 70% of the nicotine in a person's bloodstream is converted to cotinine by CYP2A6. The rest of the nicotine is treated by other enzyme pathways before being excreted. What would happen if a person did not make the enzyme CYP2A6?
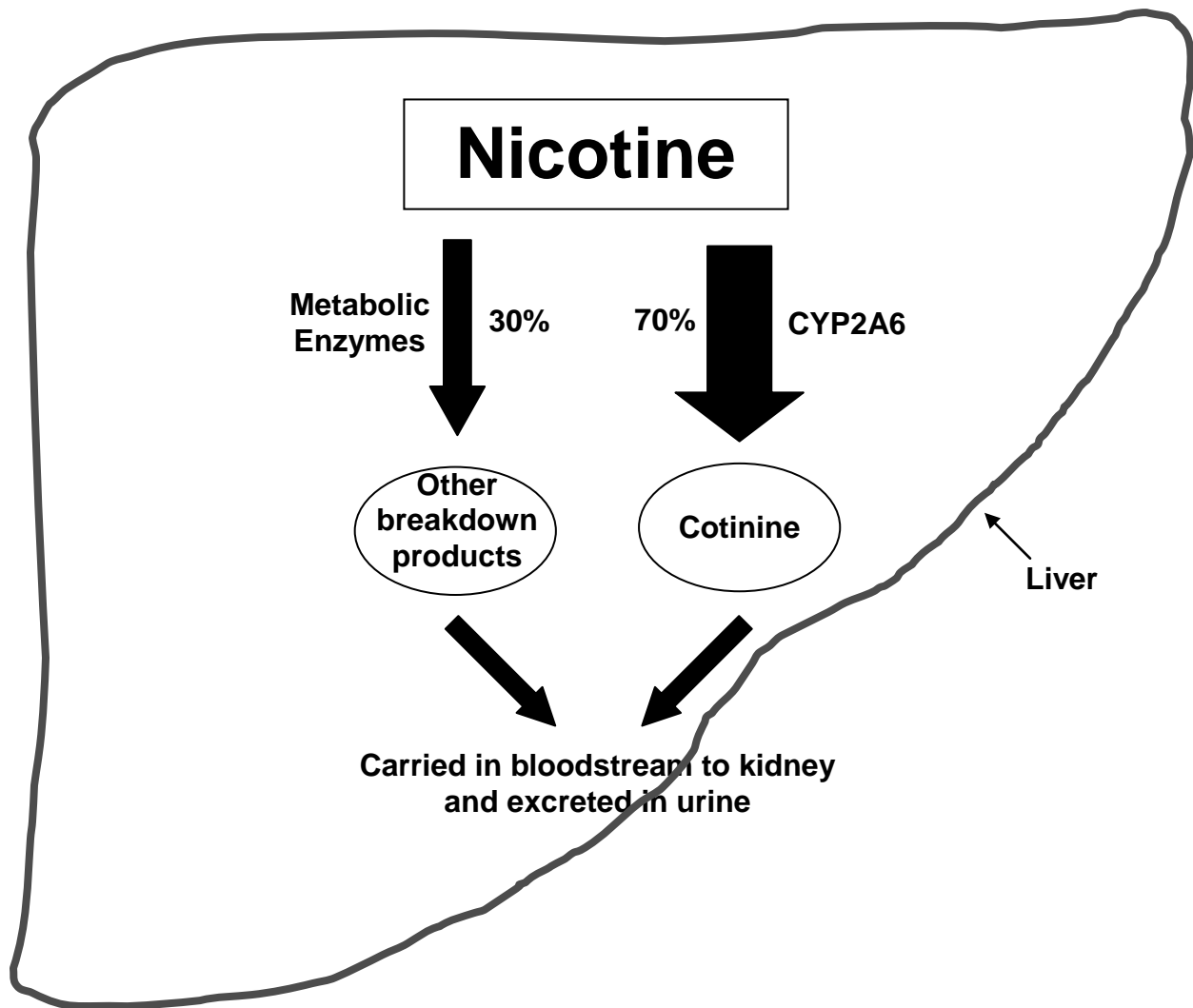
Figure 3.6. Nicotine metabolism

Choosing Candidate Genes

In a candidate gene study like the smoking behavior research study, scientists focus on one or a few genes that are likely to be associated with the particular condition they are investigating. They make their selection based on what is known about the biology of the condition under study and previous research on the same or related conditions.

Then they need to identify regions of genetic variation in the gene that may affect the protein produced from that gene and thus have an effect on the condition under study. There are several factors that need to be considered in choosing useful alleles:

1. **Frequency of the allele.** Complex traits like smoking behavior are influenced by several genetic and environmental factors, and the contribution of each factor is usually small. In order to see an effect from a particular polymorphism, scientists choose one that is fairly common, at least 5% in the study population, so that the polymorphism will be detected in a relatively small study group.

43

2. **Distribution of allele in different racial groups.** If a study includes subjects from different racial groups, then the alleles being tested should be common in all the populations involved in the study. The allele frequencies sometimes vary among different groups, and this can lead to erroneous results if the races of cases and controls are not well matched.
3. **Whether the allele has a biological effect.** Scientists are interested in genetic variations that have an effect on the function or amount of the protein made from that gene.
4. **The allele may be a marker for *another* allele that has a biological effect.** Some alleles do not have an effect on protein production, but they are co- inherited with another allele, not yet identified, that does affect protein production.

Many of the genes that code for proteins involved in the reward pathway have been studied in connection with addictive behaviors such as smoking, alcoholism, and compulsive gambling, as well as some mental illnesses like schizophrenia. Several candidate genes were considered for the smoking study, including genes for the nicotine receptor, for proteins involved in the synthesis, breakdown, and re-uptake of dopamine, the dopamine receptor, and the enzyme involved in nicotine breakdown. The genes that were ultimately included in the study are *DRD2*, which codes for the dopamine receptor 2, and *DDC*, which codes for dopa decarboxylase, an enzyme involved in dopamine synthesis (shown in Table 3.1).

Two regions of the *DRD2* gene are included in this study. The first is the -141C insertion/deletion (-141C Ins/Del), which occurs within the promoter of the gene. The deletion allele results in reduced expression of the *DRD2* gene. The -141C Ins/Del is correlated with DRD2 protein density and with addictive behaviors like alcoholism in some populations.

The second locus is called C957T. This is a "silent" or "synonymous" mutation, meaning that the T form of this gene does not affect the amino acid sequence of the receptor. A series of elegant experiments has shown that the T allele results in lower mRNA stability and decreased translation of the mRNA. Why would a silent mutation have these effects? The single nucleotide change may make the RNA transcribed from this gene more susceptible to RNA degradation. Having less of this mRNA may account for the lower level of translation. The single base change in the mRNA codes for the same amino acid, but it changes one codon, which means that a different tRNA is needed to deliver the correct amino acid. This alternate tRNA may be less plentiful or not bind to the mRNA as well as the usual one, affecting the rate of translation. Studies that measured the effect of dopamine on expression of the *DRD2* gene showed that the T allele results in a lower increase in gene expression in response to dopamine than the C allele does, making this is an interesting candidate gene to include in the study.

The *DDC* allele under study is a G to A change in an intronic (untranslated) region. In one study the A allele has been shown to be associated with nicotine addiction in African and European Americans. A possible explanation for this difference is that the G allele is part of an intronic splicing enhancer involved in regulation of splicing.

Genes included in the Smoking Behavior Research Study

Table 3.1: Gene regions included in study

Allele, gene, and location	Description	Frequency	Biological effect	Reported associations with addictive behaviors or mental disorders
-141C Ins/Del In *DRD2* gene On chromosome 11	Insertion/deletion of C in promoter 141 bases before the start of transcription	**Frequency of deletion**[1] European Americans: 39% African Americans: 21% Japanese: 11%	In tissue culture cells, the -141 deletion resulted in lower transcription compared to the insertion.	• In one cessation study, smokers homozygous for insertion (Ins/Ins) responded better to bupropion and counseling than those with one or two deletions; smokers homozygous for deletion (Del/Del) responded better to nicotine replacement than smokers with one or two copies of insertion. • In a study of Mexican Americans, insertion was associated with alcoholism
C957T In *DRD2* gene On chromosome 11	C to T at nucleotide position 957 is a "silent" mutation (no change in amino acid sequence of protein)	**Frequency of T**[2] European Americans: 57% Hispanic: 35% Pacific Rim: 11% African Americans: 9-11% African:6.5% Asian: 6.2%	Gene with T results in decreased mRNA translation and stability, and decreased effect of dopamine to stimulate transcription of the **DRD2** gene.	• In a cessation study, T/T smokers more likely to be abstinent after nicotine replacement therapy than C/C or C/T smokers. • C/C associated with schizophrenia.
DDC G/A in Intron 9	G to A in Intron 9	**Frequency of G allele**[1] Asian: 45.5%[2] Asian: 46.3%[2] European: 55.6% Sub-Saharan African: 48.2%	G allele may be part of an intronic splicing enhancer involved in regulation of splicing	• Significant association of A allele with nicotine addiction in African and European Americans

[1]Frequencies from J. Gelenter, Kranzler, H., Cubells, J.F., Ichinose, H., and Nagatsu, T. (1998). *DRD2 Allele Frequencies and Linkage Disequilibria, Including the -141C Ins/Del Promoter Polymorphism, in European-American, African-American, and Japanese Subjects*. Genomics 51, pp 21-26.
[2]Frequencies from Single Nucleotide Polymorphism Database from National Center for biotechnology Information (http://www.ncbi.nlm.nih.gov/SN

Resources

National Institute on Drug Abuse. Slide Teaching Packets
http://www.nida.nih.gov/pubs/teaching/default.html

Many concepts about addiction, neurotransmission, and the reward pathway are nicely illustrated in downloadable powerpoint slides available from the National Institute on Drug Abuse. You may wish to use some of these slides as a teaching aid.

Genetic Science Learning Center. Tour of the Basics

http://learn.genetics.utah.edu/units/basics/tour/

Explore the following topics using simple animations: What is DNA? What is a Gene? What is a Chromosome? What is a protein? What is Heredity? What is a Trait?

Genetic Science Learning Center. The New Science of Addiction: Genetics and the Brain
http://learn.genetics.utah.edu/content/addiction/

Explore graphics, animations, and interactive online activities about the reward pathway.

References

Arinami, T., Hishiguro, H., Onaivi, E. S. 2000. Polymorphisms in genes involved in neurotransmission in relation to smoking. *European Journal of Pharmacology, 410*, 215–226.

Balfour, D. J. K. 1994. Neural mechanisms underlying nicotine dependence. *Addiction, 89,* 1419-1423.

BSCS, Videodiscovery. 2001. *The brain: Understanding neurobiology through the study of addiction*. [NIH Curriculum Supplement Series Grades 9–12]. Rockville, MD: National Institute on Drug Abuse and National Institutes of Health.
Dani, J. A., and Heinemann, S. 1996. Molecular and cellular aspects of nicotine abuse. *Neuron*, *16*, 905–908.

Genetic Science Learning Center. May 27, 2008. *The New Science of Addiction: Genetics and the Brain.* http://learn.genetics.utah.edu/content/addiction/

Kandel, E.R., Schartz, J.H., and Jessell, T.M. 2000. *Principles of Neural Science*, Fourth Edition. McGraw-Hill, New York.

Munn, M. and Martinez, K. (2004) *Genetic Variation and Nicotine Addiction: investigating How Genes Influence Smoking Behavior*.
http://hshgp.genome.washington.edu/teacher_resources/StarNet_Student.pdf

National Institute on Drug Abuse. 1998. NIDA Research report series: Nicotine addiction (NIH Publication Number 01-4342). Rockville, MD

NIDA for teens: The Science Behind Drug Abuse. 2004. Retrieved March 2004 from http://teens.drugabuse.gov/glossary.asp

Oscarson, M. 2001. Genetic polymorphisms in the cytochrome P450 2A6 (*CYP2A6*) gene: Implications for interindividual differences in nicotine metabolism. *Drug Metabolism and Disposition, 29*, 91–95.

Tyndale, R. F., Sellers, E. M. 2001. Variable Cyp2A6-mediated nicotine metabolism alters smoking behavior and risk. *Drug Metabolism and Disposition*, *29*, *–548-552*.

Lesson 4. How can we study genetic and environmental influences on smoking behavior?

Overview
Students learn that the goal of epidemiology is to identify probable causes and prevent disease or other health problems. They also learn that it is often difficult to conduct experimental studies in epidemiology due to ethical and practical reasons. Students compare and contrast observational studies to experimental design studies. They learn that a case control study is one approach used to identify an association between a particular factor and health outcome in human subjects. Students learn the difference between association and causation and develop a conceptual understanding of a framework for causal reasoning used by epidemiologists to make causal inference from observational studies. Finally, students learn about key concepts in case control studies (cases, controls, exposures, outcome, and matching)

Learning Objectives

Objective	Evidence of Learning
Students will understand the components of the case control study design, including cases, controls, exposures, and outcome.	Students are able to identify the cases, controls, exposures, and outcome for the smoking behavior study and other case control studies they identify through an online search (homework).
Students will understand the difference between association and causality. Students will develop a conceptual understanding of the criteria for causality.	Students are able to distinguish the difference between association and causation by discussing examples of association but not causation. Students will understand epidemiologists may make causal inference from association by evaluating case control study results along with the criteria for causality.
Students will understand the importance of matching cases and controls, and they will learn about sources of bias in a case control study.	Students are able to articulate the potential consequences of not matching case and control subjects in the context of the smoking behavior study, such as drawing inaccurate conclusions about the smoking behavior of a specific group.

Lesson Preparation
Time
70-80 minutes

Materials
For each student: Student Sheets 4.1, 4.2, 4.3, Research Project Page-4

Getting Ready
1. Photocopy Student Sheets 4.1, 4.2, and 4.3
2. Set up your classroom computer and projector, and make sure you can access the database
   https://gsoutreach.gs.washington.edu/database2/

Glossary: Introduce and define these terms as you encounter them in the lesson:

Epidemiology: *The study of the distribution and causes of disease and other health problems in groups of people. It is also the application of the information learned from such studies to the control of the health problems.*
Case control study: *An observational study that tries to identify factors associated with a particular condition by comparing case subjects, who have the condition, with control subjects, who do not.*
Association: *An apparent (not just by chance) relationship between exposure and outcome. Does not necessarily mean the exposure causes the outcome.*
Causation*: *An exposure which increases (or decreases) the risk of someone developing an outcome.*
Study sample: *The people who participate in the study drawn from the study population.*
Study population: *The people who are able to participate in the study (and the specific characteristics used to select them, such as age range, sex, ethnicity and race)*
Case: *A study participant who has the condition of interest for the study*
Control: *A study participant who does not have the condition of interest for the study*
Outcome: *The condition that is being studied (for example, having a particular disease or set of symptoms)*
Exposure: *A factor that may be associated with developing the outcome*
Matching: *The process of selecting case and control subjects so that they are similar in certain characteristics and thus are drawn from the same study sample*
Observational study: *A study in which variables are not deliberately manipulated. Samples or subjects are observed without trying to change their environment*
Retrospective study: *A study that looks backward in time to identify factors that might be associated with a particular outcome*

* The meaning of "cause", "causation", or "causal" in the curriculum includes both direct and indirect "cause". For example, "the psychological and sociological forces that give rise to and sustain smoking behavior are just as much as causes of lung cancer as smoking itself, if it is inferred that in the absence of these forces some persons would not become or remain smokers, and thus would not develop lung cancer" (Koespell & Weiss, 2003: 189). The goal of epidemiology is to "infer" probable cause and "prevent" it, NOT "prove" it. Students may think that making a claim "A causes B" requires them to "prove" it for EVERY case. This is incorrect. Smoking is a multifactorial trait, and there can be multiple "causes" or multiple pathways that contribute to smoking.

Presenting the Activity

A. Conducting Research on Human Subjects

   1.Present students with the following scenario:

> The Beta Fraternity House held a party on Friday night. By 4 pm on Saturday afternoon, 17 people who had been at the party had called Doug, the frat house manager, complaining of headaches and severe nausea. None of the frat brothers who were absent from the party reported feeling sick. What should Doug do to figure out what was making people sick?

   2. Ask students to propose how they would determine the cause of the sickness.

Students will probably suggest doing a comparison of the people who got sick and those who didn't. Encourage them to suggest what factors might have caused the sickness, as well as how they would go about gathering information. Students will probably recognize that they need to look for one or more factors that are more common among the people who got sick than those who didn't get sick.

3. Tell students that **epidemiology** is a field of science that investigates disease in populations and tries to identify probable causes. There are several kinds of studies frequently conducted in epidemiology. For an investigation like this one, in which an outbreak of a sickness has occurred, epidemiologists would often use a **case control study**.

4. Continue with the scenario by reading the paragraph below.

> Doug reported the outbreak to Student Health Services. The public health team there designed a case control study to try to find out what had made the students sick. Their study design consisted of giving a questionnaire and a health exam to as many students as they could locate who had been to the party, including those who got sick and those who didn't. Both groups of students in the study—those that got sick and those that didn't—included males and females in the age range of 18 to 22. The questionnaire asked students about what foods they had eaten, what beverages they had consumed, and how much alcohol they drank. It also asked them to list the people they had been in contact with during the party. During the health exam, the nurse measured the students' temperature, blood pressure, and other vital signs, as well as asking them how they felt. The students who took the questionnaire were asked to report back to Student Health Services one week after the party for a follow-up health exam. Results from the study were analyzed to see whether there were any differences between the students who got sick and those who didn't.

   5.Give students the flowchart from Student Sheet 4.1, and complete the flowchart as a class, using the above example to demonstrate the characteristics of a case control study, as shown in the completed flowchart provided at the end of the Teacher Pages. Be sure to define the following terms, which students can review in the glossary in Student Sheet 4.1: epidemiology, cases, controls, study sample, outcome, exposures

Discussion Question and Evidence of Student Understanding

Ask students whether they can suggest other examples where a case control study design might be an effective approach for determining what might have caused the outcome being investigated. (*They may suggest specific cancers or outbreaks of food poisoning or infection. They should recognize that the outcome has already occurred, and the study would be looking backward in time to analyze what factors might have contributed to the outcome.*)

You may want to draw the analogy between a case control study and solving a crime—in both, the outcome has occurred, and the investigator is faced with the challenge of determining what might have caused it.

Extension and Evidence of Student Learning

Sometimes students are confused by the differences in design and terminology between case control studies and the classical experimental study usually presented as the "gold standard" in high school science classes. The activity presented in Student Sheet 4.2 is intended to help students understand the similarities and differences. This activity describes two research studies, the case control study just discussed and a classical experimental study. Students analyze the two studies and show the similarities and differences between the two by filling in a table.

As students complete Table 4.1, they should be able to see the parallels between experimental and case control studies. For example, they should recognize that: 1) matching cases and controls is analogous to having controlled variables in a classical experiment; 2) an exposure is similar to a manipulated variable; and 3) the outcome of a case control study is analogous to the responding variable in the experimental study.

They should also note the differences between the two designs, including: **1**) an experimental study is prospective (looks forward in time), while a case control study is retrospective (looks backward in time); **2**) The responding variable is measured toward the end of the classical experiment, while the analogous outcome is present in case subjects at the start of a case control study; **3**) the researcher deliberately chooses the manipulated variable in an experimental study, whereas the exposures have occurred independent of the researcher in a case control study; and **4**) in a carefully controlled experimental study, we can usually say that the manipulated variable **causes** the responding variable, while in a case control study, we say that the exposure is **associated with** the outcome. Students will learn and apply the criteria for causation in Lesson 5.

Ethical Consideration

Ask your students to discuss the following question: "When investigating whether a particular factor causes an outcome like a sickness in humans, why do researchers often use a case control study instead of an experimental study?"

Students should recognize that it is unethical to deliberately expose human subjects to a factor that might make them sick. You might ask students to suggest or research examples where experimental protocols were conducted on human subjects that lead to harm to the subjects.

For more information on conducting discussions of bioethics in your classrooms, refer to *An Ethics Primer*, a definitive resource for teaching this topic (Chowning, J. & Fraser, P. (2006). *An Ethics Primer*. Seattle WA: Northwest Association of Biomedical Research. http://nwabr.org/curriculum/ethics-primer).

B. Causality vs. Association

1. Remind students that observational studies, such as a case control study, can only show an association. Explain that there is a difference between **causality** and **association**. For example, people with red hair often have freckles, so they are definitely associated. But red hair does not necessarily cause freckles. Ask students if they can think of any other examples where two things are associated but one doesn't cause the other. Epidemiologists detect associations through observational studies, but need to go beyond the results of one observational study to infer causality. Explain to students that epidemiologists define a cause as an exposure which increases the risk of someone developing an outcome

2. Tell students that when two things (A and B) are associated, there are three possible explanations for the association. Use Figure 4.2 as you explain the causal diagram pathway. Tell students that for an exposure to cause an outcome, the directional arrow must point to the outcome and originate from the exposure.

Possibility 1: A causes B
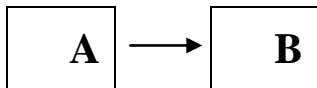
Possibility 2: B causes A
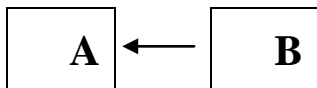
Possibility 3: A third factor, C, causes both A and B. (Neither A or B causes the other; they appear associated only because C causes them both.) A third factor (C) that is associated with both the exposure and outcome, leading to their apparent association, is called a **confounding factor**.

3. Refer to the Criteria for Causality in Figure 4.3 with your class. In lesson 5, students will work with the Criteria for Causality in an example question from the smoking behavior database and will apply the criteria to their research hypothesis in lesson 6.

Teacher Note

When analyzing case-control studies, epidemiologists need to apply a number of criteria to a detected association to help determine whether the exposure is likely to hav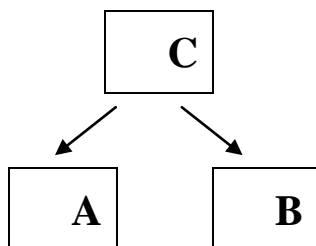e increased the risk of developing the outcome. The more criteria that are satisfied, the more confident one can be that the association is causal, however the process is somewhat subjective with consensus built after many peer-reviewed studies. Six commonly used criteria are given in Figure 4.3.

C. Student sheets: Using the questionnaire to test a research hypothesis

1. Ask students to get out their questionnaires, and give them Student Sheet 4.3.
2. Bring up the database on your projector – Step 1.1 Subject-classification and Matching Questions.
3. Explain to students that the different questions in the research questionnaire have specific purposes. Below is a table that shows the specific purposes of the questions. As you carry out a short review of these, ask the students to color code their questionnaires by circling question numbers, using the color code in the table that follows. The database is also color coded in this fashion.

Purpose of question type	Question numbers	Color
Determining whether subject is a case or a control	1 – 4, 101[1], 102	Red
Determining whether smokers qualify to be cases (see reading on Fagerstrom Test for Nicotine Dependence)	5 - 11	Blue
Matching cases and controls for age, sex, ethnicity, race,	74, 75, 76, 78	Green
Learning other demographic information about subject	73, 77, 79, 84, 85[1]	Purple[2]
Learning about cases or controls, but not both	23 - 25	Purple[2]
Learning about multiple responses to a question (These questions not useful for some types of analysis)	44, 57, 81	Purple[2]

[1]These questions are not included in database

[2] These questions are for Hypothesis Generation (Lesson 7) only because they are multiple answer questions and thus cannot be used for odds ratio calculations.

4. Explain to students that completely matching cases and controls is like having controlled variables in an experiment. In order to be able to say something about the cause of the outcome, we want make sure people in the groups are similar in age, sex, ethnicity and race.
5. Select Question 75 in Step 1.1, *What is your sex?*
   a. As you look at Question 75, ask students what they notice about the distribution of males and females in the study.

> **Teacher Note**
> As you look at data, point out the following features:
> - The data for each question are presented in a table. Each row in the table is a different response to the question.
> - Responses for cases and controls are given as a number.
> - Under the table is a graph of the data using a different color for cases and controls.

6. Bring up question 74, *What is your age in years?*

   a. Ask students if they notice differences in age between cases and control. Does this make sense?

   It is important for students to understand that there are shortcomings to these data, which they will have to consider when they do their data analysis. In Lesson 5, they will learn about different kinds of study bias, and they should consider how the level of matching of cases and controls might contribute to bias.

7. Now select Question *78, Which of these groups would you say best represents your race? (Check only one.)*
   a. Ask students what they notice about the differences between cases and controls. How well do they think the group of cases is matched with the group of controls?
   b. How well do the people in this study represent the people in their community?
   c. Do students think this will be a problem when generalizing about smoking behavior?

8. Have students spend 3 minutes making sure the questions they identified for their group research study are available for analysis. They should also pay attention to the color codes of the questions—those intended for defining cases and controls or for matching cases and controls are not useful for hypothesis testing.

D. Research project pages: The smoking behavior research study

1. Make sure students have a copy of Research Project Pages-4. They can fill in information about the study as you discuss it with them.
2. Explain to the students that they're going to learn more about the design of the smoking behavior research study through a discussion and by looking at the data in Step 1.1 Subject-classification and Matching Questions.
3. Show students overhead 4.1, the diagram of the stages of smoking that shows who the cases and controls are. They saw a similar diagram in Lesson 2.
4. Ask students to fill in the data on Research Project Pages-4 or give students the Key and discuss the differences between cases and controls for these questions.

Homework Assignment:
A quick internet search will show students how common case control studies are.
Ask them to go to http://scholar.google.com/ and search on the term "a case control study."
Provide them with a copy of Student Sheet 4.1, and ask them to complete this sheet for one of the studies that they find in their search.

Teacher Background

Facts about the Smoking Behavior Research Study
1. It is a **case control study**. This is a type of study in which case subjects, who have the particular characteristic or condition under investigation, are compared to control subjects, who do not.
2. The characteristic or condition under investigation is called the **outcome**. In our study, the outcome is whether or not people become smokers after trying smoking.
3. We are interested in learning what factors are more or less likely to occur in the case group compared to the control group. These factors are called **exposures**. The exposures in our study are many environmental factors, as well as a few genetic factors, that might influence whether someone became a smoker.
4. **Cases** are the subjects who that have the outcome. In this study, they are people who became smokers after initiating smoking.
5. **Controls** are the subjects that don't have the outcome. They are the people who didn't become smokers after initiating smoking.
6. A case control study is a **retrospective study**, meaning that we look back in time to see what might have happened to the cases and not the controls (or visa versa) to result in the outcome. The outcome occurred before the start of the study.
7. A case control study is also an **observational study** (as opposed to an experimental study). We do not deliberately manipulate any variables in this study. Instead, we select two groups of subjects, one with and one without the outcome, and we observe what exposures are associated with one or the other group.

Characteristics of the Study population

Characteristics of all subjects:
- male or female
- 25 to 54 years old
- from the ethnic and racial groups found in the Puget Sound area.

Characteristics of cases:
- Have smoked more than 100 cigarettes in their entire life (Question 1)
- Have smoked more than 30 days straight at some time in their life (Question 2)
- Last smoked within the last thirty days
- May currently use tobacco or nicotine products other than cigarettes

•Scored 4 or more on the Fagerstrom Test for Nicotine Dependence (FTND) (Questions 6-11, see below)

Characteristics of controls:
  •Have smoked 11-100 cigarettes in their entire life (Question 1)
  •Have not smoked more than 30 days straight at any time in their life (Question 2)
  •Do not currently use tobacco or nicotine products other than cigarettes (Question 3)
  •Scored 0 on the FTND

Fagerstrom Test for Nicotine Dependence
The Fagerstrom Test for Nicotine Dependence (FTND) is perhaps the most widely used test to assess nicotine dependence. This questionnaire is surprisingly short, consisting of just six questions. Responses to this multiple-choice test are assigned a numerical value, and at the end of the test, these values are added. Lower scores indicate lower dependence on nicotine, and higher scores indicate higher dependence. In our study, cases must have a score of 4 or more on the FTND. Controls are current non-smokers who initiated smoking but never became persistent smokers.

---

1. How soon after you wake up do you smoke your first cigarette?
  - After 60 minutes (0)
  - 31-60 minutes (1)
  - 6-30 minutes (2)
  - Within 5 minutes (3)

2. Do you find it difficult to refrain from smoking in places where it is forbidden?
  - No (0)
  - Yes (1)

3. Which cigarette would you hate most to give up?
  - The first in the morning (1)
  - Any other (0)

4. How many cigarettes per day do you smoke?
  - 10 or less (0)
  - 11-20 (1)
  - 21-30 (2)
  - 31 or more (3)

5. Do you smoke more frequently during the first hours after awakening than during the rest of the day?
  - No (0)
  - Yes (1)

6. Do you smoke even if you are so ill that you are in bed most of the day?
  - No (0)
  - Yes (1)

Figure 4.5.Fagerstrom Test for Nicotine Dependence

Why are the Cases and Controls in Our Study Not Completely Matched?
Two challenging aspects of conducting a case control study are 1) recruiting research subjects who meet the criteria for case or control and 2) matching cases and controls. On the face of it, finding 150 cases and 150 controls who meet the criteria described above seems trivial, especially since we are drawing on the population of the Puget Sound region—several million

people! Here are some of the reasons that make recruitment for this and other studies so challenging.

Recruitment is voluntary. This study was advertised using printed flyers posted in public places, through online ads like Craigslist, through a public service announcement played on NPR, and through newspaper ads. People interested in participating had to phone the study coordinator, be screened by telephone, and if they qualified, come to the University of Washington for consenting, completing the questionnaire, and giving a blood sample. That takes a lot of effort!

Not everyone who calls is qualified. Only about 1 in 25 of the people who called actually qualified for the study.

Some people participate because of the $30 gift card. Although the amount of compensation given to participants in this study was comparable to similar studies, the $30 gift card was incentive to participate for some people. Several of the case subjects mentioned that they had learned about the study and the $30 gift card at the homeless shelter they frequented. This group was almost entirely smokers, and they tended to be in the older age range, skewing the average age for cases.

Some people participate for altruistic reasons. Several people expressed delight that they could contribute to a study that involved high school students in doing research on why people smoke. In general, these people tended to be younger and were controls. Thus, the control group tended to be a younger crowd, while the case group tended to be older.

Some groups of people are distrustful of researchers and research studies. Historically, many underrepresented minority groups have been treated unethically when they have participated in studies, which has led to a general distrust of all research studies among people in these groups. This may account for difficulties in finding controls for some groups.

Figure 4.1: What factors might influence whether people become regular smokers after trying smoking?
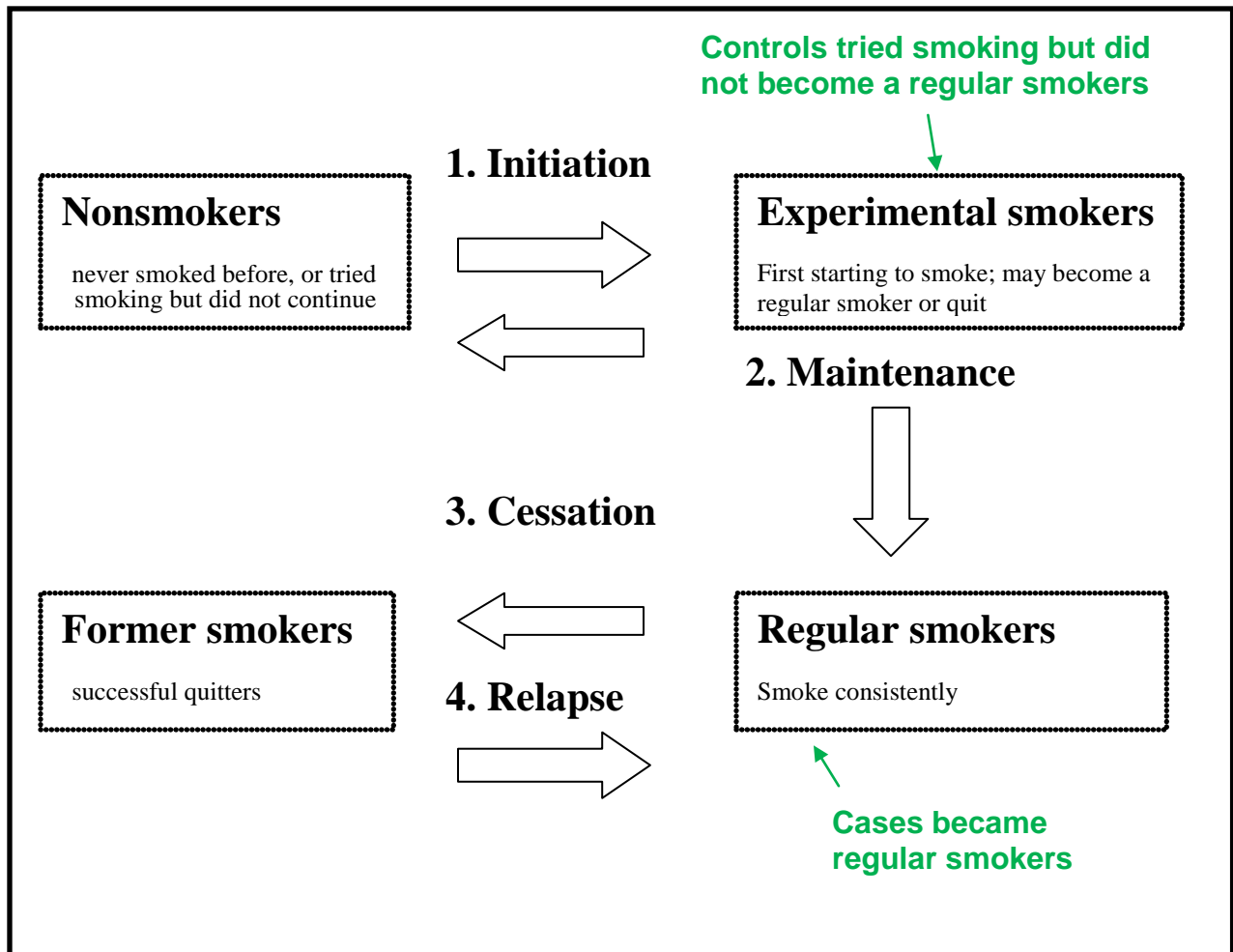
Figure 4.2: Causal Diagram

1. A causes B

A → B

2. B causes A

A ← B
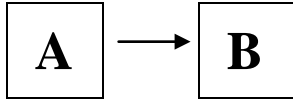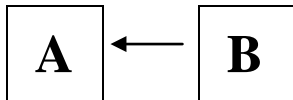
3. A third factor, C, causes both A and B. The third factor (C) is called a confounding factor.

C

A   B
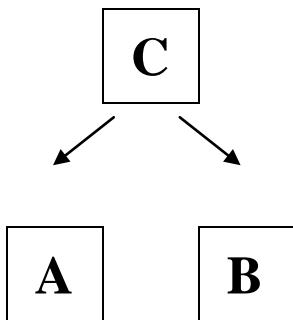
Figure 4.3: Criteria for Causality

Criteria for causality
Strength of association. A strong association between the exposure and outcome is demonstrated using statistical methods. (The larger the odds ratio, the stronger the association.)
Dose-response relationship. An increased dose of the exposure is associated with a greater risk of having the outcome. (For example, having more than 2 passengers is associated with a greater likelihood of getting into a car accident than having only 1 passenger.) <u>May not be applicable to the smoking behavior database study</u>.
Temporal sequence. The exposure must occur <u>before</u> the outcome. (Sometimes in case control studies, this can be difficult to verify.)
Consistent with other studies. The result should be mostly consistent with what is already known in the field. (If it is not, there is always the possibility that you have discovered something new and unexpected, but there is also the possibility that your study design or assumptions were in some way flawed.)
Biological plausibility. The result should make biological sense. (For example, it makes sense that having passengers in the car would cause one to get into a car accident.)
Lack of confounder or significant bias. *Can the association be explained by another factor (a confounder)?* Is there a factor or bias that explains the association?

Possible answers to STUDENT SHEET 4.1: **Case Control Flowsheet**

A case control study starts with an <u>event</u> or a <u>condition</u> that has occurred, for which the researcher wants to find the cause.

STEP 1: Identify the event or condition you wish to study. This is called the <u>outcome</u>.
We want to know: What caused the outcome?

Outcome (Condition under investigation) *Sickness at the Beta Frat House*

STEP 2: Identify exposures- factors that may have caused the outcome
Exposures are defined based on the researcher's hypothesis for what caused the outcome

What are some exposures that might have caused the outcome of this study? *One or more of the foods served at the party* *Too much alcohol* *Contact with a person at the party who was sick with a viral or bacterial infection*

STEP 3: Experimental design: Compare 2 groups of people
Researchers compare people who have the outcome with people who don't have the outcome

Cases (Have outcome) *The people who were at the party and got sick*	⟷	Controls (Don't have outcome) *The people who were at the party but did not get sick*
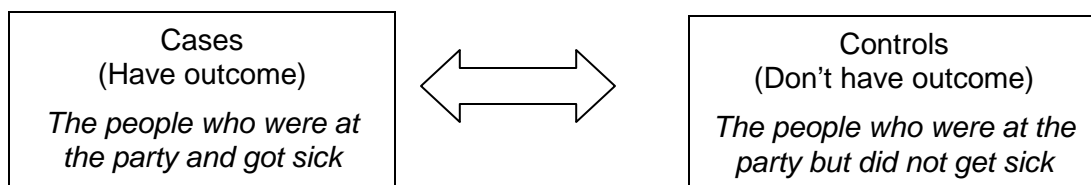
STEP 4: Measure exposures in both groups (cases and controls)

Some ways of measuring are: *Give a written survey to people who had been at the party* *Give a physical health exam*

STEP 5: Analyze incidence of exposures in cases compared to controls
The researcher asks: Is there a difference in the extent that cases and controls are exposed to a particular factor?

Possible responses to STUDENT SHEET 4.2: Comparing Experimental Studies and Case Control Studies. Answer the questions in the table.

Table 4.1. Experimental study vs. case control study

	Case Control study (students)	Experimental study (mice)
Kind of Study	Observational, retrospective (looking backward in time)	Experimental, prospective (looking forward in time)
Study design	Participants already have the condition or outcome being studied. The research looks backward in time to identify factors that might have caused that condition.	Participants are assigned to a treatment (the manipulated variable) and followed to see if they develop the condition or outcome.
Variables	The two groups of subjects are **matched** so that they are the same in as many features as possible. *How are the subjects matched?* *Everyone in the study was at the party. People in both groups are similar in age and include both sexes. The only difference in the two groups is that cases got sick after the party and controls didn't get sick.*	**Controlled variables** are the variables that are kept the same among all samples or subjects *What are the controlled variables?* *All mice in this study are Strain X lab mice. All are injected with saliva, and after injection, all are treated the same.*
	The **exposure** is a factor <u>observed</u> by researcher to be different between the two groups of subjects. *What are the exposures?* *Exposures include: foods they ate, beverages they drank, how much alcohol they drank, people they contacted*	The **manipulated (independent) variable** is the factor that is <u>deliberately changed</u> by the researcher. *What is the manipulated variable?* *Half the mice are injected with the saliva from the infected person, and the other half with saliva from the uninfected person.*
	The **outcome** is the factor that defines how the two groups of subjects <u>differ</u> and is <u>chosen</u> by the researcher at the start of the study. *What is the outcome?* *The outcome is getting sick after attending the party.*	The **responding (dependent) variable** is the factor that <u>changes</u> in response to the manipulated variable and is <u>measured</u> by the researcher. *What is the responding variable?* *Researchers monitored the mice to see whether they got sick or died.*
What experiment shows	The exposure is <u>associated with</u> the outcome. *Were any of the exposures associated with outcome in Study 1?* *There is an <u>association</u> between having close contact with Sally and getting sick.*	The manipulated variable <u>causes</u> the responding variable. *Did the manipulated variable cause the responding variable in Study 2?* *The infective agent in the saliva from the infected person <u>caused</u> the mice to get sick.*

Possible responses to RESEARCH PROJECT PAGES-4:
Lesson 4. The Smoking Behavior Case Control Study

What is the **research question** of the study?
What genetic and environmental factors might lead to someone becoming a regular smoker?

What are the **outcome**, **cases**, and **controls** of the study?
Outcome: Being a regular smoker
Cases: Regular Smokers
Controls: People who have tried smoking and quit

What is the **study population** (check how the **study sample** was determined)?
25 to 54 year old male or females from the ethnic and racial groups found in the Puget Sound area

What are the characteristics of the **study sample**? Answer this question by filling the actual numbers in the boxes with bold borders

Intended Characteristic	Category	Cases	Controls
Equivalent ratio of men and women	Men:	89	73
	Women:	73	62
Equivalent ratio across age	25-29:	12	34
	30-34:	15	28
	35-39:	24	20
	40-44:	29	19
	45-49:	41	12
	50-54:	41	18
Diverse racial backgrounds	American Indian / Alaskan Native:	5	3
	Asian:	4	9
	Black / African American:	51	15
	Native Hawaiian / Other Pacific: Islander	1	0
	White	88	100

Answer the following questions

1. Do you find any bias in the **study sample**? How do you know there is a bias?

There is a bias when there is imbalance between cases and controls. There are more male cases than male controls and more female cases than female controls, older cases and younger controls, more Black/African American cases than controls.

2. Why do you think there might have been problems properly matching cases and controls?

In a study involving human subjects, there are always complications in recruiting and retaining subjects. For example, in this study, participants received $30 to a grocery store for participating in the study. Many participants came from a center that supports the homeless, and these subjects tended to be cases. Also, the study was done at a university where many college students are willing to participate in studies, and that contributes to the ages of the controls. Often, after initial study recruitment is completed, the researchers must go back and try to recruit people from specific groups to try to achieve a balance between cases and controls.

3. How might the problems matching cases and controls (as far as gender, age, and racial group) affect the results of your investigation? Refer to your specific exposure(s).

Depending on the exposures students are studying, this may or may not be an issue. Incomplete matching could lead to confounding if the unmatched variable is causally-related to the exposure. For exposures related to education, the matching problems with ages of cases and controls will impact the results because people who are older may not have received compulsory smoking education but people from 25-34 most certainly did.

Lesson 5. Analysis of data in case control studies: the odds ratio

Overview
Students learn about the odds ratio, the statistic used by epidemiologists to analyze data in a case control study. They apply their knowledge by calculating the odds ratio for a case control study example. They also learn about the error measurement tool used for odds ratios, the 95% confidence interval. The class visits the database to see how to calculate OR and 95% CI using the "Graphical Data for Q 19" step. Students then learn how to apply the criteria for causality by working through example in the smoking behavior database. Finally, in an optional activity, students learn about sources of errors in case control studies.

Learning Objectives

Objective	Evidence of Learning
Students will understand odds, odds ratios, and the 2x2 table as they apply to case control studies.	Students are able to articulate the outcome and exposure, populate a 2x2 table, and calculate odds and odds ratios through the car passengers study (Student Sheet 5.1).
Students will understand the 95% confidence interval and the purpose of sampling.	After reviewing Figure 5.2 and analyzing the results from question 19, students are able to discuss whether the result occurred by chance and if the exposure may be causal.
Students will learn how to apply the criteria for causality.	Students are able to articulate the causal criteria and apply them to question 19 in the database.
Students will learn about sources of errors in case control studies (optional activity).	Students determine what error is present in four example studies (Student Sheet 5.6)

Lesson Preparation
Time
100 minutes

Materials
Figures 5.1-5.4
Student Sheets 5.1 and 5.2 for each student
Optional: Student Sheets 5.3-5.6

Getting Ready
1. Photocopy Student Sheets 5.1, 5.2, and <u>optional</u> activities 5.3-5.6 if needed.
2. Prepare Figures 5.1- 5.5 to use on a document projector or overhead projector.
3. Make sure you can access the database on your class computer and project it for the class to see

Presenting the Lesson

A. Interview with an Epidemiologist
This video clip introduces the concepts taught in this lesson.

1. Show students the epidemiologist video clip with Jeffrey Stanaway (video 3/3).

2. Discuss questions in class from Student Sheet 5.1 or ask students to complete the worksheet.

3. Hold a brief discussion to compare student responses. Students should understand that epidemiologists use databases frequently to try to connect exposures and outcomes, or find associations. Scientists must go beyond the results of a case control study to imply causality.

B. Reviewing the case control study and odds ratio

1. Ask students to review what information they currently have about the study, and make a list on the board. Here are some things they should include:
    - That it is a case control study (compares case and control subjects)
    - Cases are regular smokers and controls are nonsmokers who tried smoking
    - The database has both environmental and genetic information about the subjects based on the questionnaire and genotyping from the blood samples of subjects.
2. Access the database, go to "Step 1.2: Graphical Data for Q 19": Did you believe that smoking cigarettes could be harmful to your health?"
    At this stage you are accessing the database using Step 1.2: Graphical Data for Q 19 so you can show students the "raw data" for Question 19 and elicit their ideas about how you might describe the data mathematically. The first time students work with the database, they'll use Hypothesis Testing so they can't see the raw data before setting up their odds ratio calculations.

Discussion Question and Evidence of Student Understanding

Ask the students how they would interpret the data in Question 19. What do the data tell us?

The data show that many of the cases (smokers) did not believe that smoking was harmful to their health, while very few controls did. A large number of cases and most of the controls did believe that smoking was harmful to their health.

How do we know?

Looking at the data, it seems there are many more smokers who answered "no" to this question than nonsmokers and more nonsmokers who answered "yes" compared to smokers. If students are having trouble noticing these differences, you can point them to the graph which visually shows this difference.

How can we express the results in numerical terms?

Students should suggest ways to interpret the data mathematically, such as ratios or probabilities.

---

3. Tell students that they are going to learn the calculation used by epidemiologists to analyze data in a case control study, called the odds ratio.

C. What is odds?

Although this segment is presented as a discussion, you could also distribute coins and die to your students and have them do it actively.

1. Ask students: *Do you know what "odds" means?*

2. Tell students that **the odds compares the likelihood (not probability; see teacher note below for the detail) of something occurring to the likelihood of something not occurring** and is written as:

$$\frac{\text{number of times an event occurs}}{\text{number of times that event does not occur}}$$

3. Ask students: *If you flip a coin, what are the odds that you will get heads?*
    Explain as follows:
    *   There are 2 possible outcomes: heads and tails
    *   1 possible outcome is heads: 1 possible outcome is not heads (tails)
    The odds are read as a ratio. In this situation, the odds are said to be "1 to 1" or 1/1 that you will flip a heads.

4. Ask students: *What are the odds of rolling a three on a die?*
    Explain as follows:
    *   6 possible outcomes when rolling a die: 1, 2, 3, 4, 5, 6
    *   1 possible outcome is 3, and 5 possible outcomes are not 3.
    *   Odds are said to be 1:5 ("1 to 5") or 1/5 that you will roll a three. Note that this is different from saying "1 in 5."

    Ask class some other simple odds questions.

What are the odds that you will roll an odd number on the die?
*   3 possible outcomes are odd (1, 3, 5): 3 possible outcomes are not odd (2, 4, 6)
*   Odds are 3:3 or 1, that you will roll an odd number.

You have a sack of 10 tootsie rolls, 20 Jolly Ranchers, and 10 gumballs. What are the odds that you will stick your hand in the bag and pull out a tootsie roll?
*   10 possible outcomes are tootsie: 30 possible outcomes are not tootsie
*   Odds are 1:3 that you will choose a tootsie roll.

Teacher Note
Comments on comparison of odds and probability. Students may have learned about odds when studying probability in math. How beneficial a comparison to probability is depends on the students' background on figuring probabilities. If they have a lot of experience with this, but no experience with odds, they may try to equate the two and become confused. You may want to determine student math background prior to doing this activity. If they've had low exposure to figuring probabilities, you may not need to mention it now, but begin teaching odds directly, without a direct comparison.

Odds is distinct from probability, which is the chance that an event will occur, expressed as the number of favorable outcomes divided by the total number of outcomes (e.g. a 1 in 6 chance/probability that you will roll a 1 on a die.) Using the same language, odds in favor of an event occurring is the number of favorable outcomes divided by the number of unfavorable outcomes (e.g. the odds in favor of rolling a 1 on a die are 1:5).
Probability = (number of favorable events) / (total number of outcomes)
Odds = (number of favorable events) / (number of unfavorable outcomes)
In the dice example, the probability or odds or rolling a 1 on a die:
Probability = 1 / 6
Odds = 1 / 5
Total number of events = 6 = rolling a 1, 2, 3, 4, 5, 6.
Unfavorable number of events = 5 = rolling a 2, 3, 4, 5, or 6.

Trivia: We often speak of the probability or chance that an event will happen. But there are at least two places where odds are used instead of probability. Where are they? In gambling and epidemiology.

D. What is an Odds Ratio?

1. Pass out Student Sheet 5.2, *The 2 x 2 Table and the Car Passenger Case Control Study*. Read out loud the two paragraphs describing the Car Passenger study.

2. Tell students that in epidemiological research, a table like the one on Student Sheet 5.1 is called a 2 X 2 table and is used to organize the data obtained in a case control study. It is called a 2 X 2 table because there are 4 squares arranged in a 2 X 2 array. All of the data gathered will fall into one of the 4 squares. Point out the 4 squares on Figure 5.1 (outline them in red, for example).

3. Ask students what the outcome is, and who the cases and controls in the study are. On the Figure label the Case and Control columns "drivers who got in an accident" and "drivers who did not get in an accident."

4. Ask students what the exposure is. Explain that an exposure can be any factor (genetic, environmental, policy, or social factor) that you believe causes the outcome. Label the Exposed and Not Exposed rows "had 1 or more passengers" and "did not have passengers."

   The exposure is 1 or more passengers (rather than no passengers) and is entered on the **top** line of the 2 X 2 table.

5. Give students a few minutes to fill out the table with the numbers from the Car Passenger Study and to answer questions 1-6 on Student Sheet 5.2.

6. Go over the answers and make sure students understand how to fill out the table.

7. Explain to students that in order to analyze their data, they need to make a calculation called the odds ratio. This requires that they calculate the odds for cases and for controls. Help them to answer questions 7 and 8 in Student Sheet 5.2.

8. Use the following explanation as well as Figure 5.1 to help students answer questions 9-12.

Odds Ratio Explanation
To determine whether cases are more likely to have been exposed than controls, epidemiologists perform a calculation called the **odds ratio**. For example, in the Car Passenger study, they would compare the odds that cases (drivers in car accidents) were exposed (had 1 or more passengers) to the odds that controls (drivers not in car accidents) were exposed. They compare the two odds by making a ratio of them.

If cases and controls had equal odds of being exposed, then the odds ratio would be 1.

In other words, if having 1 or more passengers was not associated with getting in a car accident, then:

$$\frac{\text{odds of a case having one or more passenger}}{\text{odds of a control having one or more passenger}} = 1$$

If the odds ratio turned out to be quite a bit more than 1, then you would have demonstrated an *association* **between having 1 or more passengers and car accidents.** The larger the odds ratio is, the greater the strength of association. Whether the association is statistically significant or could be due to chance alone, is a question that students will look at more closely a little later.

Conversely, if the odds ratio turned out to be quite a bit less than 1, then you would have demonstrated that drivers in car accidents are less likely to have 1 or more passengers than drivers not in car accidents.

Significance: If the sample size is fairly big and the odds ratio is 2 or 3 or 4, then the association is very likely significant (this is not a hard and fast rule). But if the odds ratio is only 1.4, can we conclude it isn't statistically significant? How different is 1.4 from 1 (no association)? Could an odds ratio of 1.4 be due just to chance? If the sample size was very large, then it could very well be that an odds ratio of 1.4 is a significant association. But if the sample size is small, an odds ratio of 1.4 would be much more likely to be due to chance alone.

Try this odds ratio explanation with your students
Students may need to have odds ratio explained to them in several ways. Try this explanation for additional clarity. You can use this explanation when presenting Overhead 5.1.

If the odds ratio is 1, then cases and controls are equally likely to have been exposed. This shows there is no association between the outcome and exposure.

If the odds ratio is 6, then cases are 6 times more likely to have been exposed than controls. This demonstrates there is an association between outcome and exposure.

If you calculate the odds ratio and find it greater than 1, you can always make the statement:
"If the odds ratio is X, then cases are X times more likely to have been exposed than controls."

The bigger your sample size is, the more likely it is that the association you have detected with your odds ratio calculation is *real*, that is, statistically significant and not due to chance alone.

It is also possible to have an odds ratio that is less than 1. When this occurs, it shows that the exposure is less likely to be associated with cases than controls.

How to use Figure 5.1
You may want to use this figure to show students that the 4 squares in the 2 X 2 table can be labeled a, b, c, and d to help keep track of the numbers in their calculations. Epidemiologists do this too. That means that the odds of a case being exposed would be a/c and the odds of a control being exposed would be b/d. The odds ratio is nothing more than the ratio between two odds. Therefore, the odds ratio would be a/c / b/d. This is shown on the bottom of the overhead transparency. What this odds ratio means can be stated in sentence form, "Cases are a/c / b/d more likely to have been exposed than controls."

How to use Figure 5.2
This figure shows how a sample is not necessarily representative of a whole population. Even though researchers try very hard to select a sample that is representative, this is very hard to do. Consider the following points as you discuss the figure with the students:
- Is the sample a good representative of the whole population?
  Students should notice that there are relatively more red stars and yellow circles in the sample than in the whole population.
- Will the odds ratio calculated for the sample be the same as the "true odds ratio" for the whole population?
  Students should recognize that the sample OR is an estimate of the population OR and may be different.
- Challenge the students to consider what to do about the difficulty of getting a sample that is representative of a whole population.
  Students may suggest doing additional sampling or trying to select cases and controls that are more similar for the characteristics that are being used for matching. They may also suggest that you need a way to estimate the error.
- Use this conversation to lead into a discussion of the 95% confidence interval.

E. Error bars for odds ratio: 95% confidence interval

1. Use Figure 5.3 to discuss the 95% confidence interval.
2. Part A of Figure 5.3 compares traditional error bars used in many types of measurements with 95% confidence intervals.
3. Explain to students that the confidence interval is the range of values that is believed to contain the true OR with 95% confidence. By "True OR" we mean the OR for the entire population, not just our study sample. The 95% confidence interval means if the study was repeated 100 times, the confidence interval would contain the "True OR" 95 times. However, the interval would also NOT contain the true OR 5 times out of 100. Unfortunately, we do not know if our calculated 95% confidence interval contains the "True OR", and therefore we must conclude that "we are 95% confident that our confidence interval contains the True OR".

4. The formula for the 95% confidence interval is given. Tell students that they don't need to remember it because the database calculates it automatically.
5. Use the number line in Part B to point out that the 95% CI tells about the statistical significance of the odds ratio: if the 95% CI contains the value of 1 then there is no association, but if it does not contain the value of 1 then the OR is statistically significant, meaning that the OR did not occur as a result of chance. Like the OR, the 95% CI can be greater than 1, less than 1, or could include 1.

F. Calculating odds ratio and 95% confidence interval using the database
1. Project the database on the class screen and select Hypothesis Testing. Select Question 19, "During your experimental smoking phase, did you believe that smoking cigarettes could be harmful to your health?"
2. Ask students to suggest a specific hypothesis that could be tested using this question, and type it into the appropriate text box.

   Possible specific hypothesis: *Smokers are more likely to have thought smoking wasn't harmful to their health than nonsmokers."*

3. Ask students how they would define "exposed" and "not exposed" using responses to this question, and enter their responses into the text boxes for exposed and not exposed. Exposed: believing smoking is not harmful to your health (response b); Not exposed: believing smoking is harmful to your health (response a)
4. Ask students whether they think response c should be used to define exposed or not exposed. They will probably say that it should not be used. Make sure that students understand that the response, "Don't know/not sure" <u>should not</u> be used to define exposed or not exposed.
5. Show students how to select "exposed" for response b and "not exposed" for response a. They should leave response c as "neither."
6. Show students that they can select one of three populations -- everyone, males only, or females only. Tell students that unless their hypothesis is specific for either males or females, they should select "everyone" to obtain the greatest sample size to make the correct statistical conclusion.
7. Hit "Get odds ratio" to calculate the OR for this question.
8. The screen will show a section called, "Report your results and interpretation." Be sure to point out the following features:
   - The original questionnaire question and the responses for exposed and not exposed are given.
   - Everything they typed into the text boxes is also given. This allows students to double check that they calculated the odds ratio they intended to do.
   - The 2 x 2 table is given, and the sample size, odds ratio, and 95% CI are calculated.
9. Guide students in providing responses to Task a) and b). They should be able to state the following:
   Task a) The odds ratio was 9.03. This means that smokers were 13.11 times as likely as nonsmokers to have not believed that smoking was harmful to their health during their experimental smoking period.
   Task b) The 95% confidence interval is [4.38, 18.62]. It does not include the value 1, so the association is significant (likely did not occur by chance).

G. Applying the Criteria for Causality

Task c) asks about causality. Remind students that just because there is an association between believing that smoking is not harmful to one's health and becoming a smoker, we can't infer that this <u>causes</u> smoking. Epidemiologists apply several criteria to determine whether the exposure might increase the risk of the outcome. Review the Criteria for Causality found below task c) or on Figure 5.4 with your class. Then work with the class to apply the criteria to the question, "Does not believing smoking is harmful to your health during your experimental smoking phase cause smoking?" The following are some possible responses:

> **Strength of association:** The odds ratio of 9.03 is very high, indicating a strong association and the 95% confidence interval does not contain 1.00.
> **Dose-response relationship:** This is not applicable to this question.
> **Temporal sequence:** People had their beliefs about whether or not smoking was harmful to their health during their experimental smoking stage, which occurred before they became regular smokers.
> **Consistent with other studies:** Students can use Google Scholar (http://scholar.google.com/) to look for other studies that addressed this issue. Students should only read the abstract of research papers.
> **Biological plausibility:** It makes sense that if a person does not believe smoking can be harmful to their health they will be more likely to continue smoking.
> **Lack of confounder or significant bias:** Information about the health risks of smoking became more prevalent in the 1970s and later, so people who were in their experimental smoking phase during the 1960s may not have been aware of the health risks. The age distribution of cases in this study is older than controls, so it is possible that many cases became regular smokers at a time when the harmful health effects of smoking were not well publicized. The different age distribution for the case and control groups is an example of selection bias. The incomplete matching by age may <u>partly</u> contribute to the high OR for this question.

Teacher Note
Students may find this discussion frustrating because there is not a definitive answer to the question, "Does not believing smoking is harmful to your health during your experimental smoking phase cause smoking?" Encourage students to share their ideas and provide justification for their thinking.

H. Sources of error in case control studies and other extensions (optional)

Optional Extensions

1. **Analyzing data from a fictional case control study.** Student sheet 5.3 can be used as an additional example of a case control study.
   Analyzing data from real case control studies. Student sheet 5.4 allows students to analyze data from an actual case control study. We've chosen the classic study by Doll & Hill that provided some of the first scientific evidence that smoking causes lung cancer.

2. **Finding case control studies.** Students may also look for case control studies online by doing a simple Google search on case control study and a disease or condition of interest. For example, a Google search on "case control study leprosy" turns up studies examining how effective leprosy vaccines are. Allow students to follow their interests by choosing their own topics. Seeing the many studies they can find in this simple way will help demonstrate to students how widely used and valuable the case control study really is.

3. **Sources of Error.** Student Sheet 5.6 is a reading containing further information on distinguishing association from causality. It also includes information on types of study errors that can bias results and lead to apparent associations. This information may be even more useful when students are preparing to analyze data in the database in Lesson 6.

References for real case control studies
Doll, R. & Hill, A. B. (1950) Smoking and carcinoma of the lung: preliminary report. British Medical Journal 221, 739-748.

McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2007). The contribution of passengers versus mobile phone use to motor vehicle crashes resulting in hospital attendance by the driver. Accident Analysis and Prevention 39, 1170-1176.

http://www.nature.com/milestones/milecancer/full/milecancer08.html
Nature Milestones Cancer. Milestone 8. 1950 Smoking and Cancer: Smoking gun. by Ezzie Hutchinson, 2006.

CDC Excite curriculum on Cigarette Smoking and Lung Cancer
http://www.cdc.gov/excite/classroom/smoking_q.pdf
In depth exploration of Doll and Hill's classic case control and cohort studies on smoking and lung cancer.

Figure 5.1: The 2 X 2 Table

	Cases	Controls
Exposed	*a*	*b*
Not Exposed	*c*	*d*
Total		

Odds that a case was exposed:

$$\frac{\text{number of cases exposed}}{\text{number of cases not exposed}} = \frac{a}{c}$$

Odds that a control was exposed:

$$\frac{\text{number of controls exposed}}{\text{number of controls not exposed}} = \frac{b}{d}$$
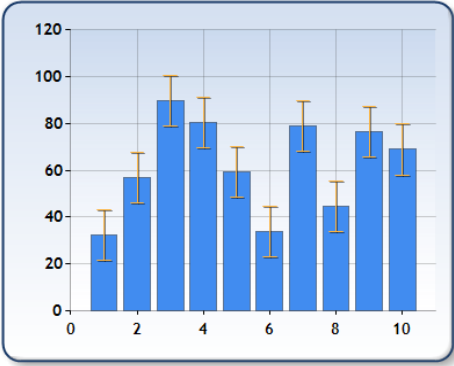
Odds ratio: $\dfrac{a/c}{b/d}$

"Cases are $\dfrac{a/c}{b/d}$ more likely to have been exposed than controls"

Figure 5.2: Sampling a population

Population
(All the people in the Puget Sound area who qualify for the study)

✧ = case
○ = control

Colors show range in characteristics used to match cases and controls (sex, age, race/ethnicity)

True population odds ratio
(The OR if you included every case and control in the Puget Sound area who qualifies for the study)

Sample
(The part of the population that was included in the study)

Sample odds ratio
(Odds ratio calculated for the portion of the populated sampled in the study)

Figure 5.3: 95% Confidence interval=error bar for odds ratios

A. Comparing error bars and 95% confidence intervals

Traditional error bars	95% Confidence interval
	A range of values that is believed to contain the true population OR with 95% confidence Expressed as: 95% CI = [lower bound, upper bound]
Often determined experimentally	Calculated: $Lower\ Bound = OR \times e^{-1.96 \times \left(\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right)}$ $Upper\ Bound = OR \times e^{1.96 \times \left(\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right)}$

B. How to interpret the 95% Confidence interval
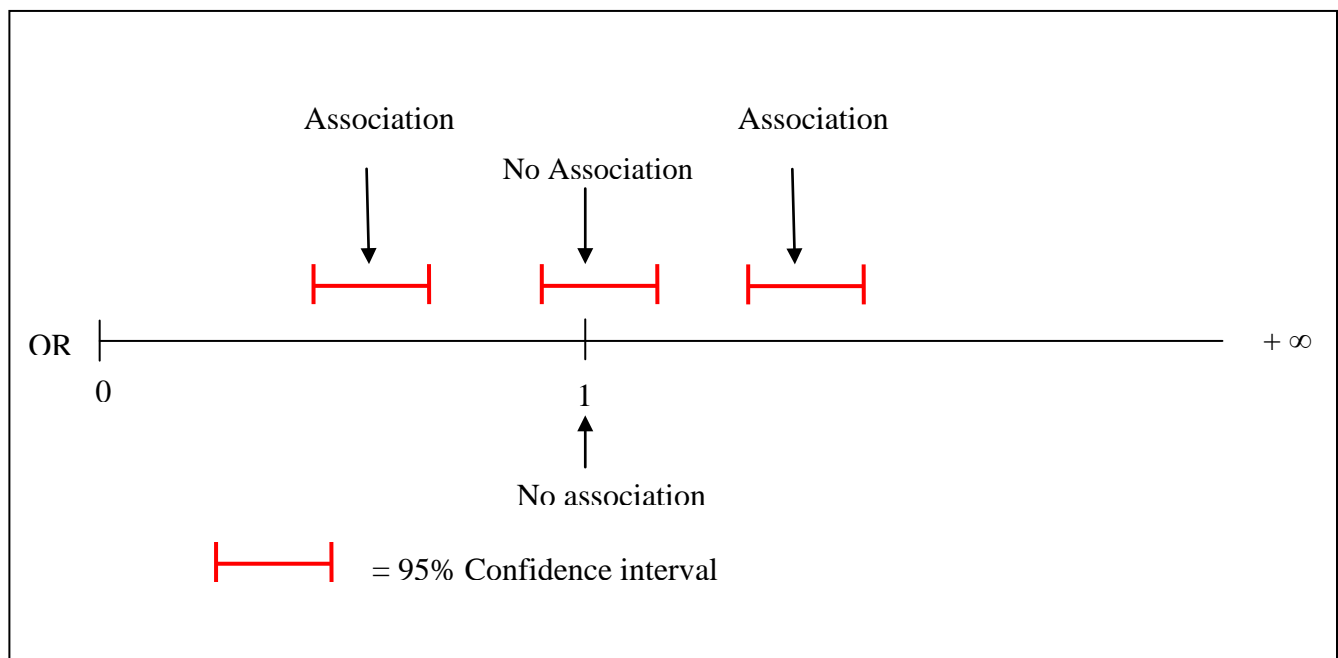
Figure 5.4: Criteria for Causality

Criteria for causality
Strength of association. A strong association between the exposure and outcome is demonstrated using statistical methods. (The larger the odds ratio, the stronger the association.)
Dose-response relationship. An increased dose of the exposure is associated with a greater risk of having the outcome. (For example, having more than 2 passengers is associated with a greater likelihood of getting into a car accident than having only 1 passenger.) <u>May not be applicable to the smoking behavior database study</u>.
Temporal sequence. The exposure must occur <u>before</u> the outcome. (Sometimes in case control studies, this can be difficult to verify.)
Consistent with other studies. The result should be mostly consistent with what is already known in the field. (If it is not, there is always the possibility that you have discovered something new and unexpected, but there is also the possibility that your study design or assumptions were in some way flawed.)
Biological plausibility. The result should make biological sense. (For example, it makes sense that having passengers in the car would cause one to get into a car accident.)
Lack of confounder or significant bias. *Can the association be explained by another factor (a confounder)?* Is there a factor or bias that explains the association?

Key to Student Sheet 5.1. **Epidemiologist Clip Guide**

https://gsoutreach.gs.washington.edu/exda-pd/lesson-5-this-is-an-example-lesson/
Play Jeff Stanaway 3/3

Contents: Four question responses from epidemiologist interview.

Learning objectives: Epidemiologists use databases to connect exposures and outcomes.
Epidemiologists need more than just the results of a case-control study to imply causality.
Understand the 95% confidence interval.

Questions:
1. How often do epidemiologists use databases?
   Almost every day.

2. What do epidemiologists try to connect in their research?
   Epidemiologists use databases to connect exposures and outcomes.

3. Name four things that epidemiologists do to imply causality in case control studies?
   Through biological plausibility, timing or temporality, multiple studies, dealing with bias and confounders through good study design.

Key to STUDENT SHEET 5.2. **The 2x2 Table and the Car Passenger Case Control Study**

The contribution of passengers to car accidents—a case control study (2007)

Whether the presence of passengers in an automobile contributes to car accidents is an important safety issue and has implication for public policy and law. A study conducted in Perth, Western Australia in 2003-2004 assessed the contribution of passengers to accidents resulting in non-fatal injuries. The presence of passengers in cars involved in injury-causing accidents and cars not involved in accidents were compared. The study included drivers aged 17 and older. Controls were matched to cases by location (recruited from nearby service stations), time of day, day of the week, and road and driving conditions.

Among 274 drivers having injury car accidents, 73 had one or more passengers in the car. The control group consisted of 1096 drivers not involved in car accidents, and among these drivers, 190 had one or more passengers in the car.

1. Fill out the Case Control Study Design form below for the car passenger study

---

Case Control Study Design

Research Question: *Does having one or more passengers in a car increase the likelihood of having an injury accident?*

Outcome: *Having an injury car accident*

Exposure: *Having one or more passengers in the car*

Study Population: *Drivers in Perth, Western Australia in 2003-2004 in areas where injury car accidents had occurred*

Study Size: *1370*

Cases: Drivers who had injury accidents

Controls: *Drivers who did not have injury accidents and who were driving in the same locations and at the same times as the cases*

Key to STUDENT SHEET 5.2: **Car Passenger Case Control Study** *(continued)*

2. On the table below, label who the cases were and who the controls were.

3. What is the exposure (the factor you think may have caused the outcome)? Label the "Exposed" and "Not Exposed" rows.

4. Fill in the shaded boxes with the appropriate numbers from the study.

	Cases Drivers having injury accidents	**Controls** Drivers not having injury accidents
Exposed	73	190
Not Exposed	201	906

Total	274	1096

5. What are the odds that a **case** was carrying 1 or more passengers? Show your work. Remember that odds is:

number of times an event occurs (i.e. cases were carrying 1 or more passengers)
number of times an event does not occur (i.e. cases were not carrying passengers)

$= 73/201 = 0.363$

6. What are the odds that a **control** was carrying 1 or more passengers? Show your work. Remember that odds is:

number of times an event occurs (i.e. controls were carrying 1 or more passengers)
number of times that event does not occur (i.e. controls were not carrying passengers)

$= 190/906 = 0.210$

7. The odds ratio (OR) is a comparison between two ratios. What is the odds ratio for this study? Show your work.

OR = Odds the cases have the exposure (odds that cases were carrying 1 or more passengers)
    Odds the controls have the exposure (odds that controls were carrying 1 or more passengers)

= 0.363/0.210 = 1.73

8. What does the odds ratio tell us? Use the language in the following statement to say what your odds ratio means.

If the odds ratio is X, then cases are X times more likely to be exposed than controls.

Drivers having injury accidents are 1.7 X more likely to have had one or more passengers than drivers not having injury accidents.

9. What would an odds ratio of 1 tell you?

An odds ratio of 1 would tell you that there is not an association between having passengers in the car and injury accidents.

10. What laws or public policies can you think of that may have been influenced by studies like this one?

Answers will vary. In Washington State, new drivers cannot have any passengers except family members for the first 6 months after they get their drivers license.

Key to STUDENT SHEET 5.3. **Mr. Limon's History Class**

1. What is Mr. Limon's research hypothesis?

   Students who fall asleep in class don't get enough sleep the night before.

2. Why do you think Mr. Limon did not require students to put their names on the questionnaires?

   If he required students to identify themselves, he might not get truthful responses because students might be afraid that it they admitted falling asleep, it would affect their grade. If this happened, the study would give inaccurate results because a number of students who identified themselves as controls (awake group) should actually belong to the cases (asleep) group. Studies that rely on the accounts of individuals are subject to this problem.

3. Why did Mr. Limon give the questionnaire to all 4 of his World History classes instead of just Jake's class?

   To increase the sample size. This could help his results be statistically significant.

4. Who are the cases and who are the controls? Label the table below.

   Cases = asleep; controls = awake

5. What is the exposure (the factor you think may have caused the outcome)? Label the "Exposed" and "Not Exposed" rows.

   Exposed = Slept <7 hours the night before; Not Exposed = Slept >7 hours the night before

6. Fill in the shaded boxes with the appropriate numbers.

	Cases *(asleep)*	**Controls** *(awake)*
Exposed Slept <7 hours the night before	*20*	*39*
Not Exposed Slept >7 hours the night before	*8*	*53*
Total	*28*	*92*

Key to STUDENT SHEET 5.3. **Mr. Limon's History Class** *(continued)*

7. What are the odds that cases slept less than 7 hours the night before? Show your work. Remember that odds is:

      number of times an event occurs (i.e. cases slept less than 7 hours)
         number of times an event does not occur (i.e. cases slept 7 or more hours)

    $=$ *20/8 = 2.5*

8. What are the odds that controls slept less than 7 hours the night before? Show your work.

       number of times an event occurs (i.e. controls slept less than 7 hours)
         number of times that event does not occur (i.e. controls slept 7 or more hours)

    $=$ *39/53 = 0.736*

9. The odds ratio (OR) is a comparison between two ratios. What is the odds ratio for this case control study? Show your work.

    OR = Odds the cases have the exposure (odds that cases slept less than 7 hours)
         Odds the controls have the exposure (odds that controls slept less than 7 hours)
    $=$ *2.5/0.736 = 3.40*

10. What does the odds ratio tell us? Use the language in the following statement to say what your odds ratio means.

    If the odds ratio is X, then cases are X times more likely to be exposed than controls.

    Students who fell asleep in class are 3.40 X as likely to have had less than 7 hours sleep the night before than students who didn't fall asleep in class.

11. Do these results support Mr. Limon's hypothesis? Give evidence from your calculations to support your answer.

    Yes, there is an association between falling asleep in class and not getting enough sleep the night before. The odds ratio is quite a bit greater than 1, which suggests that the association may be significant (not due to chance alone).

12. What would an odds ratio of 1 tell you?

    It would tell you that students who fell asleep in class and those who stayed awake had the same odds of having slept less than 7 hours the night before, and there is no association of amount of sleep and whether students fall asleep in class.

Key to STUDENT SHEET 5.4: **A real case control study**

The occurrence of lung cancer increased rapidly in the first half of the twentieth century. Why? There were several ideas. During this period, in western nations, manufactured cigarettes were plentiful and easy to obtain. Could increased smoking account for the increase in lung cancer cases? A classic case control study carried out in 1947 by English epidemiologists Sir Richard Doll and Tony Bradford Hill examined the relationship between smoking and lung cancer. Doll and Hill compared the smoking habits of hospitalized London lung cancer patients to the smoking habits of patients hospitalized for other causes. All the patients were men and under the age of 75. Here are their results:

	Cases (lung cancer)	**Controls** (no lung cancer)
Cigarette Smokers	1350	1296
Non-smokers	7	61
Total	1357	1357

1. Fill out the Case Control Study Design below for Doll and Hill's study.

Case Control Study Design

Research Question: *Is smoking associated with lung cancer? or Does smoking cause lung cancer?*

Outcome: *Lung cancer*

Exposure: *Smoking cigarettes*

Study Population: *Hospitalized patients in London, male, under 75 years of age*

Study Size: *2714*

Cases: *Hospitalized lung cancer patients*

Controls: *People hospitalized for other reasons besides lung cancer, male, under 75 years*

Key to STUDENT SHEET 5.4, **A Real Case Control Study** *(continued)*

2. Using the data provided for this study, calculate the percentage of cases who smoked and the percentage of controls who smoked.
   %cases who smoked = 1350/1357 = 99.5%
   %controls who smoked = 1296/1357 = 95.5%

3. What can you conclude from these percentages and from comparing them?
   The percentage of people smoking is very high in both groups, and it is quite similar in both groups. It is hard to conclude by comparing these two figures that smoking is higher in one group or the other.

4. What are the odds that a case will be a smoker?

   1350/7 = 192.9

5. What are the odds that a control will be a smoker?

   1296/61 = 21.2

6. What is the odds ratio for this study? Show your work.

   192.9/21.2 = 9.1

7. What does this odds ratio tell you about smoking and lung cancer? Give evidence from your calculations to support your answer.
   People with lung cancer are more likely to be smokers than people who don't have lung cancer. The odds ratio tells us that lung cancer patients are 9.1 times more likely to be smokers than patients without lung cancer.

Key to STUDENT SHEET 5.6. **Association is Not Always Causality**

Test Your Knowledge
Which of the following, made-up studies is an example of each type of error?

Type of error	Study number
a. Random error	1
b. Selection bias	4
c. Information bias	2
d. Confounding	3

Lesson 6. How can we test a hypothesis using the smoking behavior database?

Overview
Students will use the database to investigate specific hypotheses related to their team's overarching hypothesis. They will interpret information and artifacts gathered throughout the unit to complete their team research investigation. Students are given the opportunity to reflect on what they know and seek more information to strengthen their knowledge and final product.

Learning objectives

Objective	Evidence of Learning
Students will demonstrate that they can propose and test specific hypotheses related to their overarching hypothesis	Students develop specific hypotheses that are based on preliminary evidence from smoker profiles and other resources, justify why they proposed the hypotheses they did, and perform an odds ratio calculation
Students will evaluate whether an association is statistically significant	Students demonstrate their ability to interpret the meaning of the odds ratio and 95% confidence interval for specific queries they make
Students will apply the criteria for causality to determine whether an exposure might have caused the outcome	Students discuss whether an observed association between an exposure and the outcome is consistent with one or more of the criteria for causality
Students will evaluate whether their evidence supports or refutes each specific hypothesis	Students evaluate evidence from different sources and discuss whether this evidence supports or refutes each specific hypothesis
Students will evaluate whether their results support or refute their overarching hypothesis.	Students will use logical reasoning to relate their specific hypotheses with their overarching hypothesis.

Lesson Preparation
Time
120-150 minutes

Materials
for each student
Research Project Pages packet (students already have these)
Directions for Research Project Pages

Getting Ready
1. Make a class set of Directions for Research Project Pages.
2. Students will need one computer per pair of students for this lesson. Make sure all student computers are running Firefox 3.5 or higher and can access the database. If possible, bookmark the database on all computers.
Database URL: https://gsoutreach.gs.washington.edu/database2/
3. It is no longer possible to save your queries on the database. You may want to instruct your students to use "Print Screen" to capture images of their database query and save them as a word or PDF document.

Presenting the Lesson

A. Comparison of Hypothesis Testing and Hypothesis Generation
1. Figure 6.1 compares two types of experiments that students can do using the smoking behavior database.
2. Use this table to highlight the similarities and differences between these two approaches.
3. Students may ask why they are limited to doing just 4 queries in the hypothesis testing activity. Remind them of the jelly bean cartoon they saw in Lesson 2 (if you ask enough questions, you're bound to get a positive association by chance alone).

B. Hypothesis Testing
1. At the end of Lesson 2, students chose a research topic, formulated a hypothesis related to their topic, and identified 4-5 questions from the questionnaire that will allow them to test their hypothesis. Have each group report their research topic and hypothesis, and record them for the whole class.

2. Ask students to work with their teams to fill in the first sections of RPP-7-10 with their research topic, their overarching hypothesis, the four questions they plan to analyze, and the specific hypothesis for each question. Additionally, students should fill in the responses they will use to define exposed and not-exposed for each question before working with the smoking behavior database and indicate the study population they will use (Everyone, Males, or Females).

Teacher Notes

It is important that students decide how they will define exposed and not-exposed for each question and write them in the Research Project Pages before they proceed to "Odds ratio calculation".

Here are some general tips for defining exposed and not-exposed:

• Use most or all responses to the question except "don't know/not sure to define exposed or not-exposed

• Unless the hypothesis is specific to men alone or women alone, select "everyone" for study population

• Make sure that responses for exposed and not-exposed make sense. For example, with a question like #13: "How old were you when you first smoked an entire cigarette?" responses for exposed should include consecutive ages, as should responses for not-exposed.

---

3. Ask each student team to break into pairs. Each pair will share a computer and analyze half of the group's questions.

4. Have each student pair navigate on their computer to:
   https://gsoutreach.gs.washington.edu/database2/

5. Instruct students on how to access their account using the account number you gave them or that they set up (see Appendix for instructions).

6. Ask students to click on "Hypothesis Testing". Note that putting the mouse cursor over the blue question icon gives helpful hints.

7. When students are ready to begin odds ratio calculations, you may want to lead them through the process of calculating and interpreting an odds ratio for one or more questions, as you did for Question 19 in Lesson 5. Students may need additional help defining exposed versus not-exposed, stating what the odds ratio and 95% CI mean, and applying the criteria for causality.

8. Have students complete the rest of RPP-7-10 using the Hypothesis Testing step in the database. You may want to give them the option of entering their data interpretation into Tasks a, b, and c of the database and then saving and printing their results (instead of writing their responses in the bottom half of RPP-7-10).

9. Monitor students' progress and ability to interpret data in relation to their overall research topic and hypothesis.

Evidence of student understanding

The odds ratio means the following: *Cases are (odds ratio) times as likely as controls to have had the exposure.*

Make sure that students are able to phrase this statement in the context of the queries they make.

Note: Make sure that students DO NOT interpret the odds ratio as meaning that people with the exposure are (odds ratio) times as likely to be smokers as people who don't have the exposure. This interpretation is <u>not</u> correct.

If the 95% confidence interval does not include the value 1, students should be able to say, "Because the 95% CI does not contain 1, the association is statistically significant." However, if the 95% confidence interval does include the value 1, they should state that the association is not statistically significant.

Note: If the 95% CI = [1.2, 5.5], it does not include the value 1 (even though 1.2 is very close to 1), the association is significant.

The "reverse" button switches the definitions of exposed and not-exposed. When exposed and not-exposed are reversed, the odds ratio will be the reciprocal of the value calculated originally. For example, in Lesson 5 you calculated the OR for question 19, "During your experimental smoking phase, did you believe that smoking cigarettes could be harmful to your health?" and used response b, no, for exposed and response a, yes, for not exposed. The odds ratio was 13.11 and the 95% CI = [5.36, 32.04]. If you reversed exposed and not-exposed, the OR becomes 1/13.11 = 0.08 and the 95% CI is [0.03, 0.2]. Students may find it difficult to interpret an odds ratio that is less than one, and if so, they can use the reverse button to get an OR greater than one.

Just because there is an association between an exposure and an outcome does not necessarily mean that the exposure **causes** the outcome. Students will need to apply the criteria for causation to provide evidence that an exposure increases or decreases someone's risk of becoming a regular smoker. An exposure that results in an odds ratio greater than one is called a **risk factor**, because it can put the person at risk for becoming a smoker. An

---

exposure that results in an odds ratio less than one is a **protective factor** because it can
protect the person from becoming a smoker. Refer back to this discussion in Lesson 5.

10. Once students complete the odds ratio analysis and causality criteria checklist (in
Directions for Research Project Pages and popup on database) for each of their questions,
they should meet with their team to share and interpret results.

C. Mapping Activity and Drawing Conclusions

1. Using the Mapping Activity in RPP-11, each group should fill in their overarching
hypothesis in the oval. In each of the four squares, they should write the number for each
of the four questionnaire questions and the specific hypothesis for that question.

2. Students should then draw the relationship between each question and becoming a regular
smoker, using the key that is provided. Their decision for each question should be based
on the results recorded in RPP-7-10 (or the pages they printed from their database
queries).

3. The questions in RPP-12, Drawing Conclusions, are intended to guide students'
syntheses of their results. For Question 1, students should address how their results for
each of the four questionnaire questions support their overarching hypothesis. Their four
questions may not all support their overarching hypothesis to the same extent, so they
should provide justification for their observations. In Question 2, they should discuss the
two specific hypotheses that best support their claim and why. In Question 3, they should
discuss the strength of their evidence supporting their overarching hypothesis and
whether other factors, such as confounding factors or selection bias might have a
significant impact on their results. Question 4 asks about the broader implications for
their findings on human health and the community. They may consider discussing how
their results might impact public policy.

Evidence of Student Understanding

Students should be able to discuss how their evidence from each of their four specific hypotheses
supports or refutes their overarching hypothesis. The strength of association for the four specific
questions may be different, and if so, they should discuss why they are different.

For each specific hypothesis, students should be able to apply the criteria for causality. Not all of
the criteria may apply. In addition, there may be contradiction among the different criteria. For
example, there may be good evidence for a causal relationship, except their results may not be
consistent with other studies. If so, students should weigh their results against previous findings
and discuss why there might be differences (e.g. different experimental design or study
population among the studies they are comparing). As a second example, they may think that
there is good evidence for causality, but also be concerned about the possibility of a confounder
or significant bias. In this case they would have to weigh the influence of the confounder or bias
on their overall results.

Students should demonstrate an understanding of their findings on human health and their
community. They may want to suggest how their results might be used to influence public

---

policy. For example, if their research shows that smokers are more likely to have had parents who smoke, then they may propose that family doctors counsel parents who smoke about the risk of their children also becoming smokers.

4.   Make sure that students have completed RRP-7-12 prior to continuing with Lesson 7. They will need these pages to complete their final project.

Figure 6.1: Comparison of Hypothesis Testing and Hypothesis Generation

	Part I: Hypothesis Testing	**Part II: Hypothesis Generation**
Overview	-Hypothesis needs to be determined before looking at database -Only analyze up to four questions related to hypothesis	-Analyze many questions and exposures to generate new hypotheses -Results from Part I: Hypothesis Testing can lead to new hypotheses
Propose Hypotheses	-Develop hypothesis based on intuition/observations/smoker profiles/past research	-Develop hypotheses from analyzing many questions from the database
Design and Perform Study	-Analyze up to four queries related to hypothesis to test if an association exists -Define exposed and not-exposed before analyzing data (calculating odds ratio and looking at data)	-Analyze many queries and exposed/not-exposed combinations to develop multiple hypotheses -Change exposed/not-exposed multiple times
Conduct Odds Ratio and Confidence Intervals Calculations	-Perform odds ratio and 95% confidence interval calculations for each pre-determined question	-Learn what happens to odds ratio when exposed/unexposed combinations are changed.
Analyze Data and Make Conclusions	-Analyze the odds ratio and 95% confidence interval from each question to determine if there is an association -Apply the criteria for causality to determine the likelihood that an exposure causes someone to become a smoker -Determine if odds ratio and 95% confidence interval calculations support your hypothesis	-Analyze all query results to propose future studies and develop new hypotheses.

Lesson 7. Hypothesis generation and proposing a new case control study

Overview
Students will explore the Smoking Behavior database to analyze many queries and generate a new hypothesis, which they will integrate into a new hypothetical case control study.

Learning Objectives

Objective	Evidence of Learning
Students will understand how to query the Smoking Behavior database to look for patterns among case and control subjects to generate hypotheses for future studies	Students demonstrate that they can 1) interpret data in the "Hypothesis Generation" view of the database by recognizing differences between cases and controls; 2) define exposed and not-exposed to achieve the strongest association; and 3) generate new hypotheses based on the patterns they observe in the data

Lesson Preparation
Time
100 minutes

Materials
for each student
Research Project Pages packet (students already have these)
Directions for Research Project Pages (students already have these)

Getting Ready
1. Make copies of Figure 6.1, 7.1 and 7.2 for the overhead projector or the document camera.
2. Make sure you can project the database in your class.
3. You will need one computer per pair of students for this lesson.

Presenting the Lesson

A. Reviewing Hypothesis Testing and Hypothesis Generation
  1. Use Figure 6.1 to review the similarities and differences between hypothesis testing and hypothesis generation.
     a. Remind students that when they were testing their specific hypotheses, they were limited to using four questions only (due to the problem of multiple comparisons), and they had to define their specific hypotheses, exposed and not-exposed, and the study population before looking at the data.
     b. Explain that in hypothesis generation these constraints don't apply because they aren't going to make any scientific claims based on their queries. Instead, they are going to look for trends in the data and propose a new hypothesis that they would want to test in a future experiment. They will also provide some information about the design of the future experiment.

B. Modeling How to Explore the Data
   Work as a class to complete Student Sheet 7.1.

   1. Project the database on the screen for your students to view it:
      https://gsoutreach.gs.washington.edu/database2/.

   2. Select Step 1.4 Hypothesis Generation, and then Question 12, "How old were you when
      you first tried a cigarette, even a single puff?"

   3. Guide the class in selecting how they will define exposed and not-exposed for Queries 1,
      2, and 3. Ask students to record the data from the three queries on Student Sheet 7.1. You
      may want to use Figure 7.1 to help students think about how to get the highest odds ratio.

   4. Have students work individually or in pairs to answer Questions 1 and 2 at the bottom of
      Student Sheet 7.1. Then hold a class discussion about the trends they observe as they
      change the exposed/not-exposed and study population. Students should notice that having
      more cases than controls who are exposed, and more controls than cases who are not-
      exposed leads to a high odds ratio. They should also notice that when the study size
      decreases by selecting males or females for the study population, the 95% confidence
      interval becomes broader, making the results less accurate.

   5. Figure 7.2 shows some tips for getting the largest possible odds ratio for any one
      question. Students may need some help in understanding these tips and linking them to
      the observations they made in Student Sheet 7.1.

Hints for Guiding Students through Student Sheet 7.1
Point out the similarities and differences between the Hypothesis Generation view of the
database and the view in Hypothesis Testing.

   ● In both views, you define exposed and not-exposed by clicking beside each
     response.
   ● In Hypothesis Generation, the data are shown as a table and a graph. You can use
     this information to define exposed and not-exposed to give the highest possible
     odds ratio.

As you work through Student Sheet 7.1, students may need some help in defining exposed and
not-exposed. Figure 7.1 will provide some guidance in using the graph to select the optimal
responses for exposed and not-exposed.

After students have completed and discussed the questions, you may want to show them Figure
7.2, which describes a systematic way to get the highest possible odds ratio.

B. **Using the Database to Generate a New Hypothesis, and Proposing a New Case Control Study**

This activity guides research teams in considering research topics and questions they might want to explore in order to generate a new hypothesis and propose a new case control study.

1. Ask students to work in their teams to complete RPP-13, Planning Your Hypothesis Generation Research. After completing this sheet, students should form pairs and divide the questions they plan to query between the two pairs. Point out that several questions, with multiple response answers, cannot be used to calculate an odds ratio. For these questions (purple questions from Student Sheet 4.3) students should graphically interpret the data.

2. Direct the student pairs to complete their queries using the database and record their data in RPP-14, following the instructions in Directions for RPP. Students will need an additional copy of RPP-14 if they do more than four queries.

3. When teams have completed their queries, they should get together with their teammates to complete RRP-15-16, which guides them to generate a new hypothesis and propose a new case control study in which to test it.


Tips for guiding hypothesis generation and proposing new research studies

Students frequently express frustration that they can't investigate certain research topics using the Smoking Behavior database, either because the data aren't there or because of constraints on how the data can be used. This lesson provides students with the opportunity to mine the database for associations that they may not have discovered as part of hypothesis testing. In addition, it allows them to propose a hypothetical case control study of their own design.

As students complete Table 1 in RPP-14, they should apply their understanding for how to maximize the odds ratio through the way they define exposed and not-exposed, as they learned from completing Student Sheet 7.1. By posing two or more queries for a single question, they should be able to identify patterns in the data that will enable them to generate a new hypothesis for a future study.

In RPP-15-16, students work in their teams to decide which of their queries revealed the most interesting results and might lead to a future study. For example, they may notice a dramatic difference when they compare males and females for a particular query, which may prompt them to design a future study that focuses on this exposure in one sex or the other. Or their queries on related questions may show contradictory results that pique their curiosity to learn more about this research topic. This activity provides a lot of latitude in student responses, but they should use their data throughout to justify their proposal. Students should record the details of their proposed research study in the Case Control Study Design template.

Figure 7.1. Using graphs to look for trends in the data
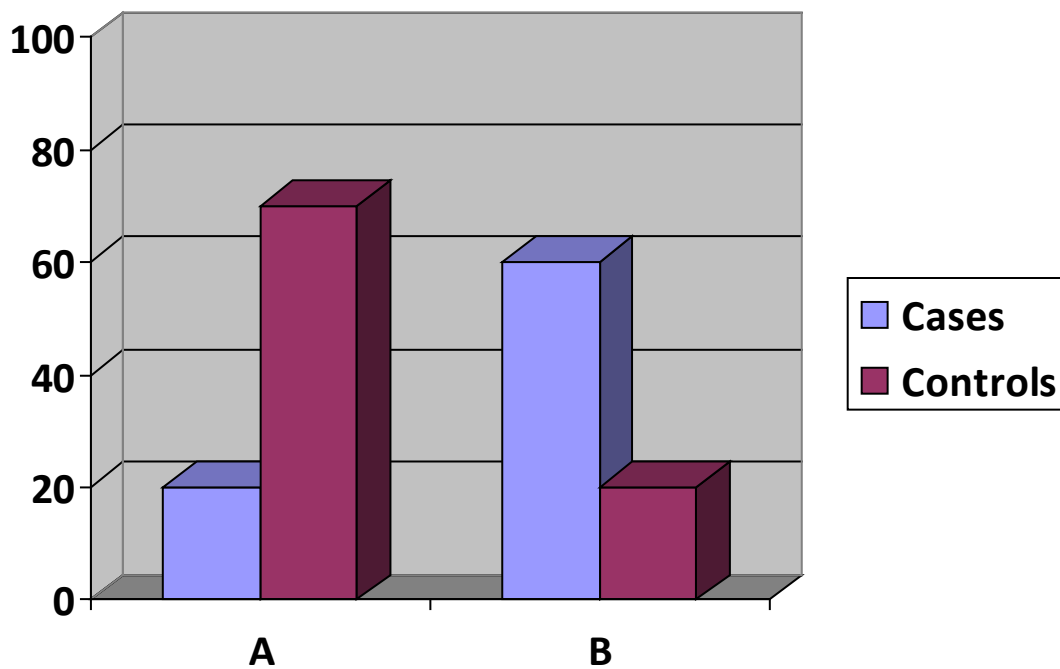
Assume Response B is Exposed and Response A is Not-exposed:

● Large differences between cases and controls for Responses A and B mean that an association is very likely (OR ≠1). In this example, the odds ratio will be >1.

● Small differences between cases and controls for Responses A and B mean that an association is unlikely (OR =1).

Figure 7.2. Using the 2x2 table to look for trends in the data

Remember that an odds ratio is calculated from a 2x2 table:

The 2 X 2 Table

	Cases	Controls
Exposed	a	b
Not-Exposed	c	d

Odds ratio: $\dfrac{a/c}{b/d} = \dfrac{ad}{bc}$

- **The odds ratio will be greater than one when a x d > b x c.**
  In other words, the odds ratio increases when the counts in "a" and "d" cells (exposed cases and not-exposed controls) increase or when the counts in the "b" and "c" cells (not-exposed cases and exposed controls) decrease.

For large odds ratios:

- **Case counts need to exceed control counts for each exposure**

- **Control counts need to exceed case counts for each non-exposure**

- **Try to use all study sample subjects by using all the responses to the question except don't know/not sure to define exposed and not-exposed, and by selecting everyone for the study population (unless your hypothesis is specific to men alone or women alone).** This will result in smaller confidence intervals, which means that your odds ratio will be closer to the "true odds ratio" for the entire population.

Overview
Students will synthesize what they have learned from their database research by preparing a poster, PowerPoint presentation, or other final product and presenting it to their peers.

Learning Objectives

Objective	Evidence of Learning
Students will analyze their results from hypothesis testing to create a final presentation.	Students demonstrate their understanding of their hypothesis testing research by developing a presentation using a format you assign (PowerPoint, poster, or other).
Students will take part in scientific argumentation by critiquing the presentations of their peers and responding to the questions and comments of others during their own presentations.	Students are able to respond to questions asked by their peers and provide reasonable justifications for their analyses and conclusions. Students demonstrate their understanding of their peers' research by asking relevant, thoughtful questions. They demonstrate their ability to take part in reasonable discourse by posing and responding to questions respectfully and with a focus on providing deeper scientific understanding.

Final Project Preparation
Time
50-100 minutes for preparing presentations and 50-100 minutes for presenting to class

Materials
Materials will vary, depending on the kind of project you and your students choose to do
Final project template
PowerPoint template (optional)
Final Presentation Rubric, Team Research Presentation Notes, Presentation Self and Peer Evaluation (see examples from Judy Shaw provided at the end of this lesson)

Getting Ready
Make available the materials will need for preparing their final projects. This may include poster paper and colored markers for a poster presentation or having computers available so students can prepare PowerPoint presentations. A PowerPoint template is provided to guide students' presentation and ensure that they include their research (and not just images that they download from the Web!).

Presenting the Lesson
1. Describe the final project format and requirements to your students. Some format ideas are provided in Final Project Ideas.

2. Provide student teams with the final project template, which will help them to organize the information, evidence, and conclusions they have gathered throughout the unit. Also give them the Final Presentation Rubric so they know what the expectations are for the final project. You may also want to give them the PowerPoint template.

3. Provide guidance as needed while students prepare their projects.

   <u>Project presentations:</u>

4. The format for presentations will vary, depending on your preference and the type of presentations. Some ideas are provided in the next section.

5. Provide students with a copy of the Presentation Notes and Self & Peer Evaluation sheets to complete during the presentations.

6. Discuss your expectations for asking questions and providing critique to their peers during the presentations.

Final Project Ideas

The final project should help students synthesize and communicate what they have learned during the Exploring Databases curriculum. Students should use the content of the Research Project Pages as their guide. Below are descriptions of a few options for final projects to help guide you in asking students to share their work. Please note that all projects require students to include all information from the **Final Project Template.**

PowerPoint presentation: Students use the attached template to create a PowerPoint presentation to display their research findings and suggestions for future research. Students respond to questions from other students and the teacher to defend their research.

Poster: Students create a poster (large poster or using two manila folders to make a small stand-up poster) with all the components of the attached abstract. Here are a few options for having students present their posters: (a) Students do a gallery walk of posters with half the students presenting their work as the other half of students walk around and discuss and evaluate the posters using a rubric; (b) Students participate in an "expo" open to the school community (peers, teachers, administrators, members of the community), in which they present their posters and defend their research to people who attend the expo.

Pamphlet: Students design a pamphlet for the public explaining the genetic and environmental influences (or lack thereof) on smoking behavior. Students present this pamphlet to the class and defend their research.

Write-Up: Students use the template provided to write an abstract describing their research. You can have students peer-review the abstracts based on the rubric and they can make a decision about whether or not they believe claim is supported with sufficient evidence and reasoning.

Take-home questions: You can choose to have your students respond to questions in a take-home test format. You choose the focus of the questions, but they could include questions about design of a case control study, neurobiology and addictive behaviors, the effect of genetic and environmental factors on smoking behavior, causality and association in their own work, etc.

Exploring Databases Final Project Template

Reporting and Interpreting Hypothesis Testing Results

Title and Authors (RPP-2):
The title should describe the work to the reader. Include all group members' names.

Abstract:
The abstract is a one or two paragraph summary of the entire article, giving the main features and results of the work described more completely in the report.

Introduction:
The introduction has four parts: 1) A summary of your group's initial thoughts about smoking behavior (RPP-1), 2) Your research question (RPP-2), 3) Background context – where does this question fit with what is known? (RPP-2-3), and 3) Your overarching hypothesis (RPP-2).

Methods (for all projects)
We used the smoking behavior database to calculate the odds ratio and 95% confidence interval for four questions related to our hypothesis.

Results
Summarize results from RPP-7-10 for each of the four questions. Questions to answer:

    a) Provide the odds ratio and 95% confidence interval for each of the four questions.

    b) Use the odds ratio in a sentence that describes what it means.

    c) Is there an association between the exposure and becoming a regular
       smoker? How do you know?

    d) Do you think this exposure causes people to become a regular smoker? Apply
       the criteria for causality to support your answer.

Discussion/Relevance (Include your mapping activity – RPP-11, and the conclusions you discussed in RPP-12)

A. What do your results show about your overarching hypothesis?

B. The evidence that most strongly supports our claim is:
        1. Discuss 1 piece of evidence and why it supports your claim
        2. Discuss 1 piece of evidence and why it supports your claim

C. How strong is the evidence that your hypothesis is supported?

D. What are the implications of your research for health? For your community?

Example of an Abstract for an Epidemiology Paper

This example follows the format for a structured abstract that might be included in an epidemiology paper. It contains many of the sections suggested in the Final Project Template. Although this abstract is not identical in format to the Template, it demonstrates how to present the results that determine whether or not there is an association (odds ratios and confidence intervals), and then to apply the criteria for causality to argue that the two strongest associations are causal. Note that one of the four questions did not support the hypothesis.

Affect of Parental Apathy toward Smoking on Whether a Person Becomes a Smoker
By Mark Gallivan

Abstract

Background
In 2008, a case-control study collected genetic and environmental data from 300 adult smokers (cases) and non-smokers (controls) to determine what factors might influence whether people become regular smokers after trying smoking.

Purpose
Our study group hypothesized that people who had parents that were apathetic to other person's smoking were more likely to become smokers.

Methods
We used the smoking behavior database to calculate the odds ratio and 95% confidence interval for four questions related to our hypothesis.

Results
Regular smokers were 1.36 (95% CI: 0.63, 2.91) times more likely to have had parents/guardians that asked nearby smokers to stop smoking compared to non-smokers. Regular smokers were 13.7 (95% CI: 5.69, 32.99) times more likely to have had parents that allowed them to smoke inside their home than non-smokers. Regular smokers were 4.21 (95% CI: 2.46, 7.19) times more likely to have had parents that allowed visitors and guests to smoke inside their home compared to non-smokers. Compared to non-smokers, regular smokers were 7.61 (95% CI: 4.06, 14.28) times more likely to have been from a family that asked to be seated in a smoking section or had no preference rather than a non-smoking section or choosing non-smoking restaurants.

Conclusion/Relevance
The smoking behavior database provided strong evidence that having a parent who was apathetic to other's smoking is associated with becoming a regular smoker. According to the 95% confidence intervals, three of our four research questions were determined to be meaningful to the entire population. There is evidence that having a parent who was apathetic to other's smoking may cause someone to become a regular smoker. The strength of association (odds ratio) was particularly strong for having a parent who let the study participant (13.7) or guest (4.17) smoke inside their house. Having a parent who was apathetic to other's smoking occurred before the participants became a regular smoker. Our study's findings are consistent with other studies which have shown that adolescent smoking behavior is largely affected by parental attitudes towards smoking. It is highly plausible that having a parent who does not discourage other's smoking behavior will increase the likelihood that their child maintains smoking. Another factor cannot explain this strong association, nor do I believe that cases would recall their parent's attitudes on smoking significantly different from controls. In conclusion, smoking prevention interventions need to target parental attitudes towards smoking.

UW Smoking Behavior Project
Exploring Databases Team Research
Final Presentation Rubric

Research Team Name: _____

Research Team Members: _____

_____

The following rubric is for the team's presentation.

A. Team is well Prepared   2   4   6   8   10

    All materials ready
    Completed on time

B. Introduction   2   4   6   8   10

    Research topic stated clearly and overall hypothesis
    Hypothesis presented with clear connection

C. Body – Clear and organized   10   20   30   40   50

    Data presented   Questions used from data base
    Odds Ratio
    95% Confidence interval addressed   Implications
    Evidence Presented clearly   New Hypotheses
    Clear Association between a particular factor   How new hypothesis was developed
        and becoming a smoker or nonsmoker
    Hypothesis was proven correct or wrong with'
        clear evidence

D. Visual Aid Used   2   4   6   8   10

__ Power Point __ Poster __ White board __ Doc Camera __Other:_____

- Creative
- Colorful   • Clearly organized
- Accurate Information   • Fits well with presentation

E. Conclusion – Complete and compelling   2   4   6   8   10

    Interesting connection to hypothesis
    Was the overall question answered in presentation

F. Overall evaluation   2   4   6   8   10

    Presentation was complete, well presented, interesting

Comments:

TOTAL SCORE: _____/100 pts

UW Smoking Behavior Project – Exploring Databases Team Research

Presentation Notes

Please fill in the following table for each Research Team's during presentations.

No.	Research Team Members	Research Topic	Hypothesis	Evidence/Findings
1				
2				
3				
4				

No.	Research Team Members	Research Topic	Hypothesis	Evidence/Findings
5				
6				
7				
8				

UW Smoking Behavior Project - Presentation
Self & Peer Evaluation

Name:	My Research Team Members/Team Name:
Self evaluation Evaluate your group's presentation (1-10)	% of work distribution among team members (note below - example- 4members - Pete - 25%; Sam - 25%; Mary - 25%; Anne - 25%)
Peer Evaluation (1-10 pts)	Comments/What you learned from the group's presentation
1	
2	
3	
4	
5	
6	
7	
8	

Appendix A
Student Sheets
Exploring Databases

Name:_____ Date: _____ Period: _____

STUDENT SHEET 1.1: **Conducting Research on Smoking Behavior**

1. How did the scientists in the interviews become interested in science and research?

2. What personal connections do these scientists have with their work?

3. Using your experience and any information from the interviews, describe what you think happens in science. What do scientists do? How do they do experiments? Where do scientists do science?

4. How are the ways that these scientists do work similar or different from how you do science in school or at home?

5. What are some ways that you see science being applied in your everyday life?

6. How do scientists work together to solve problems and answer new questions?

7. What surprised you about these scientists? About the kinds of science they talk about?

STUDENT SHEET 2.1: **Why do some people become smokers and others do not?**

As you read and discuss the smoker profiles, answer these questions.

1. What patterns do you see across the profiles? Did people start smoking around the same age? When? What other similar do things you notice?

2. What factors influence people:
   a. to begin smoking?

   b. to continue smoking after they have started?

3. Why do they have difficulty quitting?

4. What other information might you want to know from these people?

5. What aspects, factors related to smoking behavior would you want to investigate further?

6. What statements or claims can you make that address the question: "Why do some people become smokers and others do not?" Provide evidence from the profiles for those claims. For example, someone who has a parent who smokes is more likely to be a smoker.

STUDENT SHEET 2.2 **Observations about smoking behavior**

WHY PEOPLE TRY SMOKING

WHY PEOPLE CONTINUE TO SMOKE AFTER TRYING IT

WHY PEOPLE HAVE DIFFICULTY QUITTING SMOKING

RELATED TERMS

STUDENT SHEET 2.3. **Reading: The Stages of Smoking**
Highlight or record 3-5 points that are important to your group's topic.

As you saw in the smoker profiles, people's smoking behavior often developed along a similar path. This shared path consists of several phases. First comes the **initiation** stage of smoking, when people experiment with cigarettes, usually as teens. As people smoke more regularly and frequently, they pass into the **maintenance** phase. Smokers in this stage are usually addicted to nicotine. Later, smokers may try to quit smoking, known as **cessation**. Unfortunately, many people exit cessation and **relapse** back into the maintenance phase, but some do successfully quit.

Not all smokers will go through all stages. Many people experiment with cigarettes but don't pass on to the maintenance phase to become regular smokers. Other people become smokers but are able to quit smoking the first time they try and so never experience relapse. The stages and the flow between them are shown in the figure below. A person's environment and physiology/genes can influence smoking behavior at each stage.

Stages of Smoking

EXPLORING DATABASES   APRIL 2013
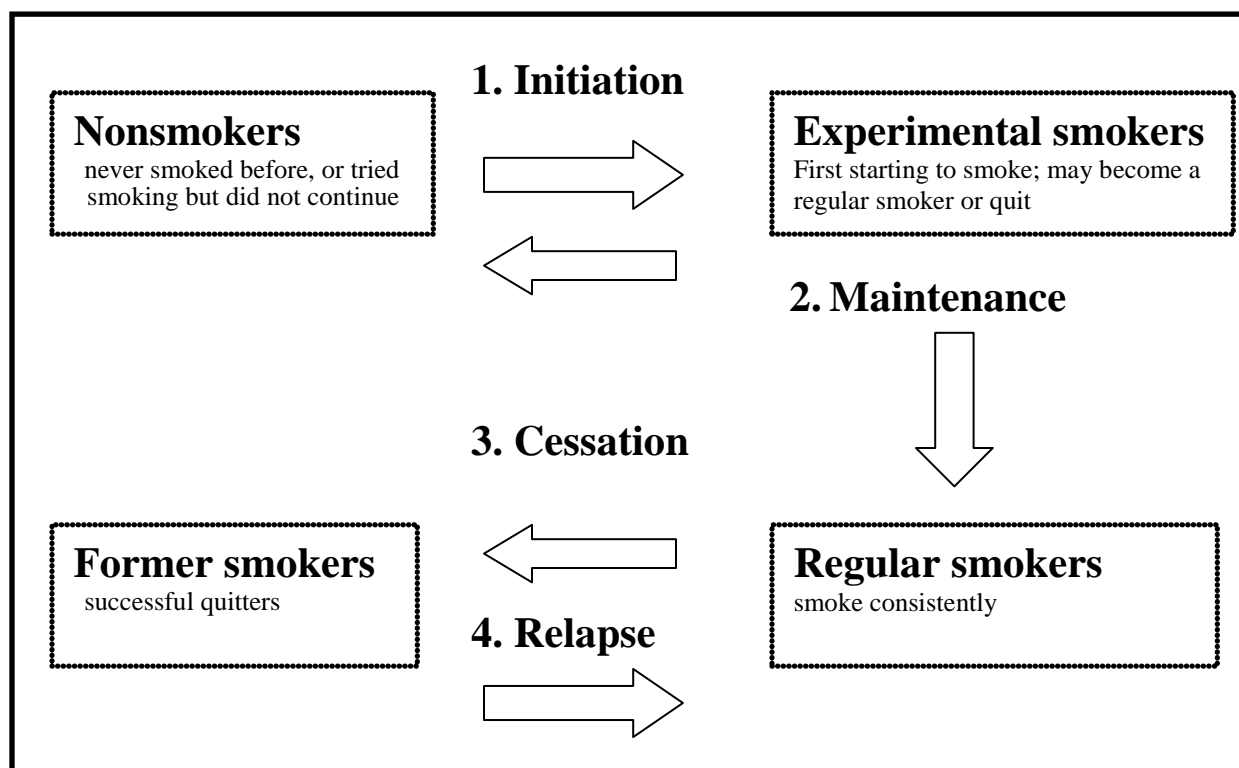
Initiation

During the initiation stage of smoking, people first experiment with nicotine but have not become regular smokers. Studies have shown that over half of the population tries at least one cigarette sometime in their lives. As you saw from the profiles, most smokers initiate smoking in their teens. Researchers agree that the window of initiation opens around age 8 and closes at about 20. Statistics show that if children reach the age of 18 without smoking, their chance of remaining nonsmokers is around 90%.

The initiation phase often involves deliberating about smoking, becoming aware about the social aspects of smoking (both good and bad), and sampling a few cigarettes. At this point, differences in how people taste cigarettes may play a role in whether people continue smoking or not. Some people are less sensitive to bitter tastes than others and may be more likely to enjoy the somewhat bitter taste of cigarettes. A single gene can determine this difference! At this early stage, many people will not continue to smoke. They tried it, didn't like it, or got little behavioral reinforcement (for example, their friends didn't want to be around them when they smoked), and so they quit. Other people experience a growing desire to keep smoking and will gradually increase the number of cigarettes they smoke.

Environmental factors during adolescence contribute to smoking initiation. For example, both peer groups and school environment strongly influence whether teens try cigarettes and whether they become regular smokers. Recent studies have also identified even earlier influences on early smoking behavior. In a study of high school seniors, it was found, not unexpectedly, that nonsmokers were more likely to have had parents who had never smoked. In families where parents do smoke, if one parent quits by the time the child is eight or nine, the child's chance of becoming a smoker drops by 25%. If both parents quit, the child's chance of smoking decreases by 40%. So even in these pre-adolescent years, factors in children's environment, such as parental smoking, can influence their future smoking behavior.

Maintenance

Maintenance is the stage describing established smoking behavior—persistent smoking beyond the experimental initiation phase. This stage is marked by complex physiological changes in the body that cause addiction and nicotine cravings.

Moving from initiation to maintenance is strongly influenced by a smoker's environment. Among young people, it has been found that those who had at least two smoker friends were more than six times as likely to move from initiation to maintenance. This study also found that teenagers with strong ties to their family or school were less likely to enter maintenance. And as previously mentioned, having smoker parents greatly influences whether teens initiate smoking, and whether initiators go on to become regular smokers. Between one third and one half of all people who try smoking will move on to become regular users.

As reflected in the smoker profiles, maintenance behavior varies from person to person. Some people smoke regularly, but only a few cigarettes a day. Others go for years smoking a pack or more per day. Why the difference? Part of the answer may lie in a person's genes.

Studies have shown a strong genetic contribution to maintenance behavior. Although we don't completely understand how a person's genes can influence their nicotine use and susceptibility to addiction, some genes that are likely to be involved have been identified and are being studied.

Two types of genes that may be involved are genes that regulate nicotine metabolism and genes that function in the brain's reward pathway. Variation in genes from the first group can lead to differences in how long nicotine stays in a person's blood, what substance the nicotine is converted to, and how quickly it has a biochemical effect. The brain's reward pathway is a complex biochemical pathway involving the neurotransmitter dopamine that leads to rewarding feelings of pleasure. Genetic variation in this pathway can change how potently people feel the effects of nicotine, how often they need to smoke to maintain those effects, and how many cigarettes they need to get the "reward" feeling. Some of these genetic differences might also influence smoking behavior during the initiation stage.

Cessation
The process of smoking cessation generally follows a pattern of deliberation, planning how to stay off cigarettes, decreasing consumption, and efforts to stop smoking. Decisions on why to quit vary from personal health concerns to powerful social incentives such as heavy taxes on cigarettes and the disapproval of friends. Although some smokers manage to quit "cold turkey," many gradually decrease the number of cigarettes they smoke each day, using aids like nicotine patches or gum, hypnosis, counseling, or anti-depressants to help themselves "kick the habit." Even with these tools, many smokers find it incredibly difficult to beat the powerful addictive properties of nicotine and stop smoking permanently.

Many of the same environmental factors that promote initiation and maintenance can hamper a person's effort to quit. In a peer group where smoking is common, the constant exposure to cigarette smoke and offers of cigarettes can challenge the strongest of wills. The feelings of pleasure and calm reported by many smokers begin to fade soon after "the last cigarette." For many people, that withdrawal can be too uncomfortable to maintain.

Relapse
Despite the health benefits of quitting smoking, nicotine addiction often proves too much to defeat. Many people struggle with their cravings and desire to smoke and move through cycles of cessation and relapse several times. According to the World Health Organization, success rates in programs designed to help young people quit smoking are usually low, around 13%.

Summary
As you have read, both environmental and genetic factors influence smoking behavior. In some stages, such as initiation, environmental factors appear to play the most important role. In other stages, such as maintenance and cessation, both environmental and genetic factors are important influences. The specific contributions of environmental and genetic factors haven't been completely defined yet, but many studies are being directed at this question, both in the laboratory and in epidemiological studies of human populations. Research findings are expected to aid in preventing nicotine addiction and in helping addicted smokers quit.

STUDENT SHEET 3.1: **Tracing Nicotine's Path through the Body**

Think about where nicotine might go when it enters the body. On the diagram of the human body, trace the pathway that you think nicotine might follow from the time the person inhales it until it leaves the body.
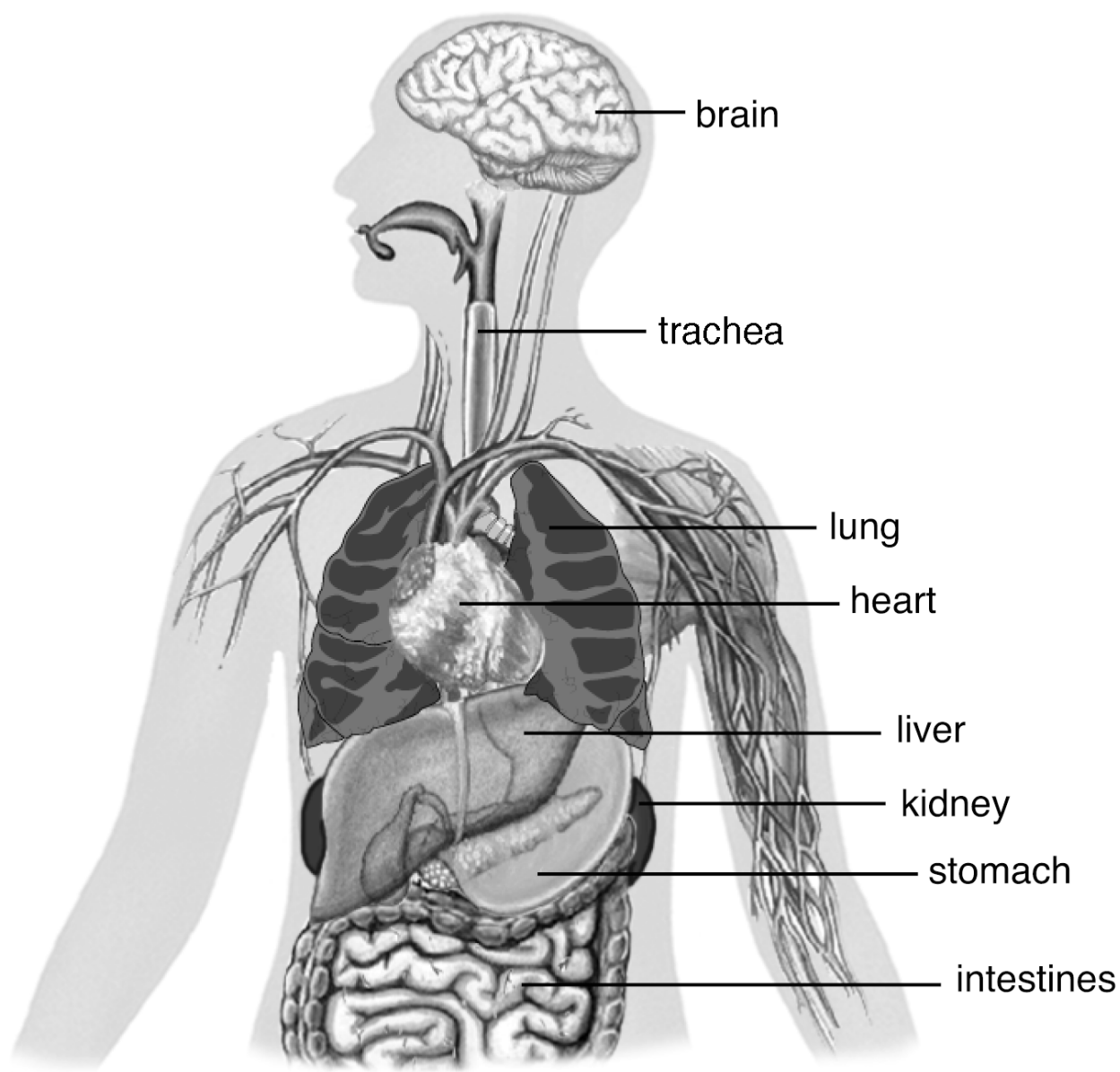
STUDENT SHEET 3.2: **Neurons, Neurotransmission, and the Giant Neuron**

During the class discussion of neurons and neurotransmission, take notes on diagrams a, b and c. Then label the parts of the giant neuron in part d.

a. **A typical neuron**
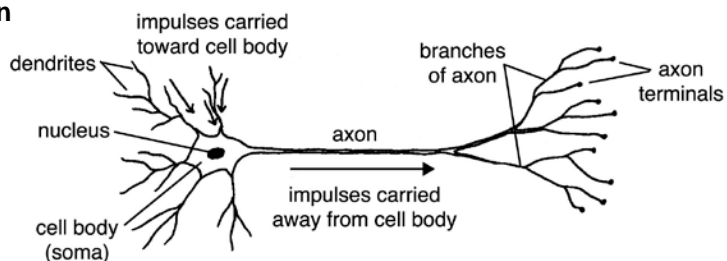
b. **Neurotransmission**
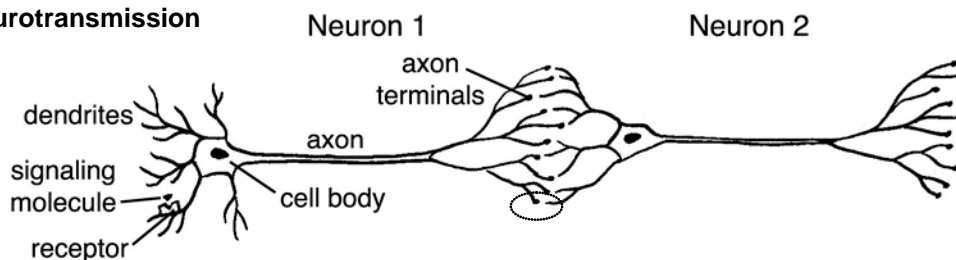
c. **Close-up of neural synapse**
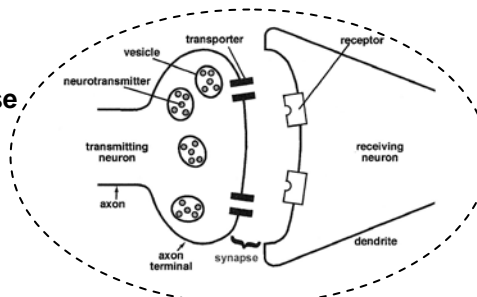
d. **Sketch of the giant neuron**
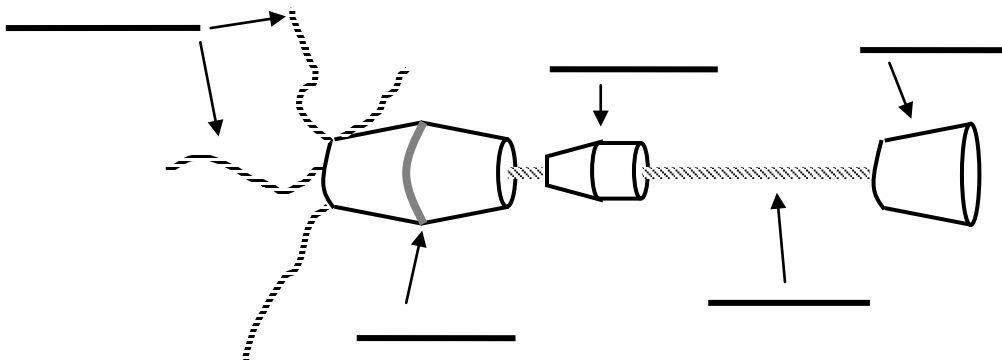
STUDENT SHEET 3.3: **Genetic Exposures in the Smoking Behavior Study**

As a class, look at Question 103 in the Hypothesis Testing view of the smoking behavior database. Under "3. Identify answers corresponding to your exposure," there is a list of four possible answers for the C957T region of the *DRD2* gene. Work with your class to answer the questions below.

1. How many possible genotypes (combinations of genes) are there for this gene region?

2. Why does each genotype have two letters associated with it (C/C, T/T, C/T)?

3. *DRD2* is on the long arm of chromosome 11, shown by the arrow at the left in the figure below. Draw the genotypes on the chromosome pairs, as shown in the example at the left. Why are there two chromosomes for each genotype?
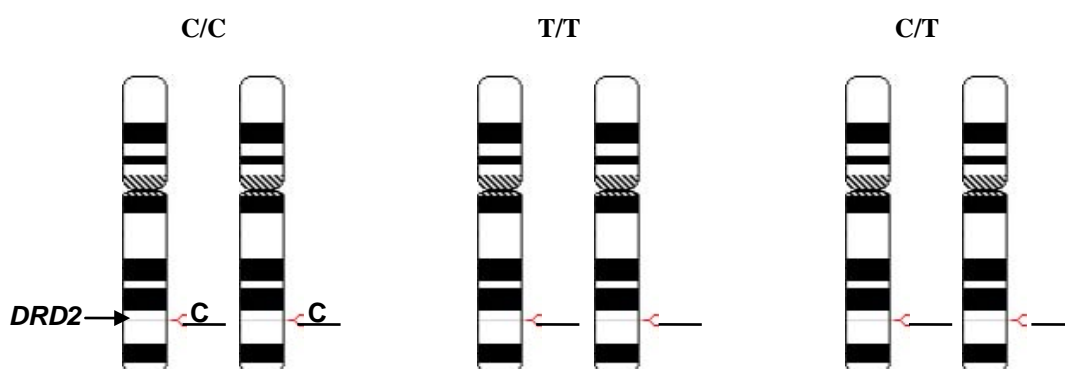
4. *DRD2* codes for the dopamine receptor. The T allele of the C957T gene region results in fewer receptors than the C allele. Draw the number of receptors you expect for each genotype in the diagrams below. One receptor is shown in each figure.
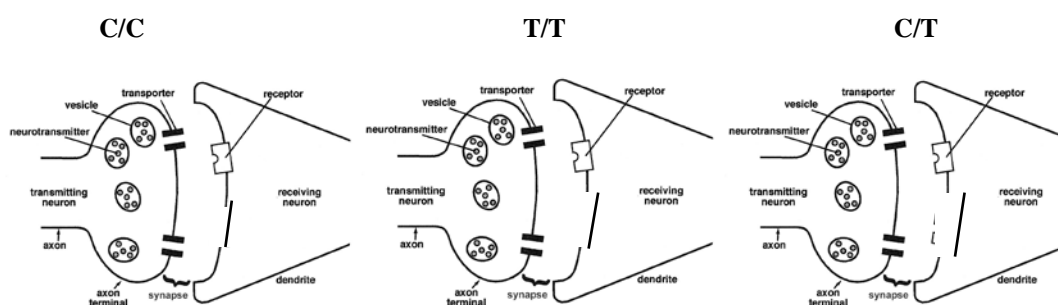
STUDENT SHEET 3.3, Continued

5. You just learned that the *DRD2* gene codes for the dopamine receptor. What is the function of this protein? How might making less of this protein (compared to "normal") affect a person's smoking behavior? (Hint: Think about the giant neuron and the effect of nicotine on dopamine release).

6. What protein does the *DDC* gene code for? What is the function of this protein? How might making less of this protein (compared to "normal") affect a person's smoking behavior?

7. Several of the questions in the questionnaire may have a physiological/genetic component to them, even though there is not a specific gene mentioned. Identify at least three questions that may be asking about something that is partially or largely genetic, and discuss why you think this for each question.

Transmitting neuron

Receiving neuron

a. Electrical impulse travels along axon away from cell body to terminals in nucleus accumbens and prefrontal cortex
b. Dopamine transporter takes up dopamine from synapse
c. Vesicles of dopamine fuse with cell membrane
d. Nicotine binds to receptors on transmitting neuron in the VTA
e. Dopamine binds to receptor on receiving neuron
f. Increase or decrease in electric impulse along receiving neuron
g. Dopamine is released into synapse
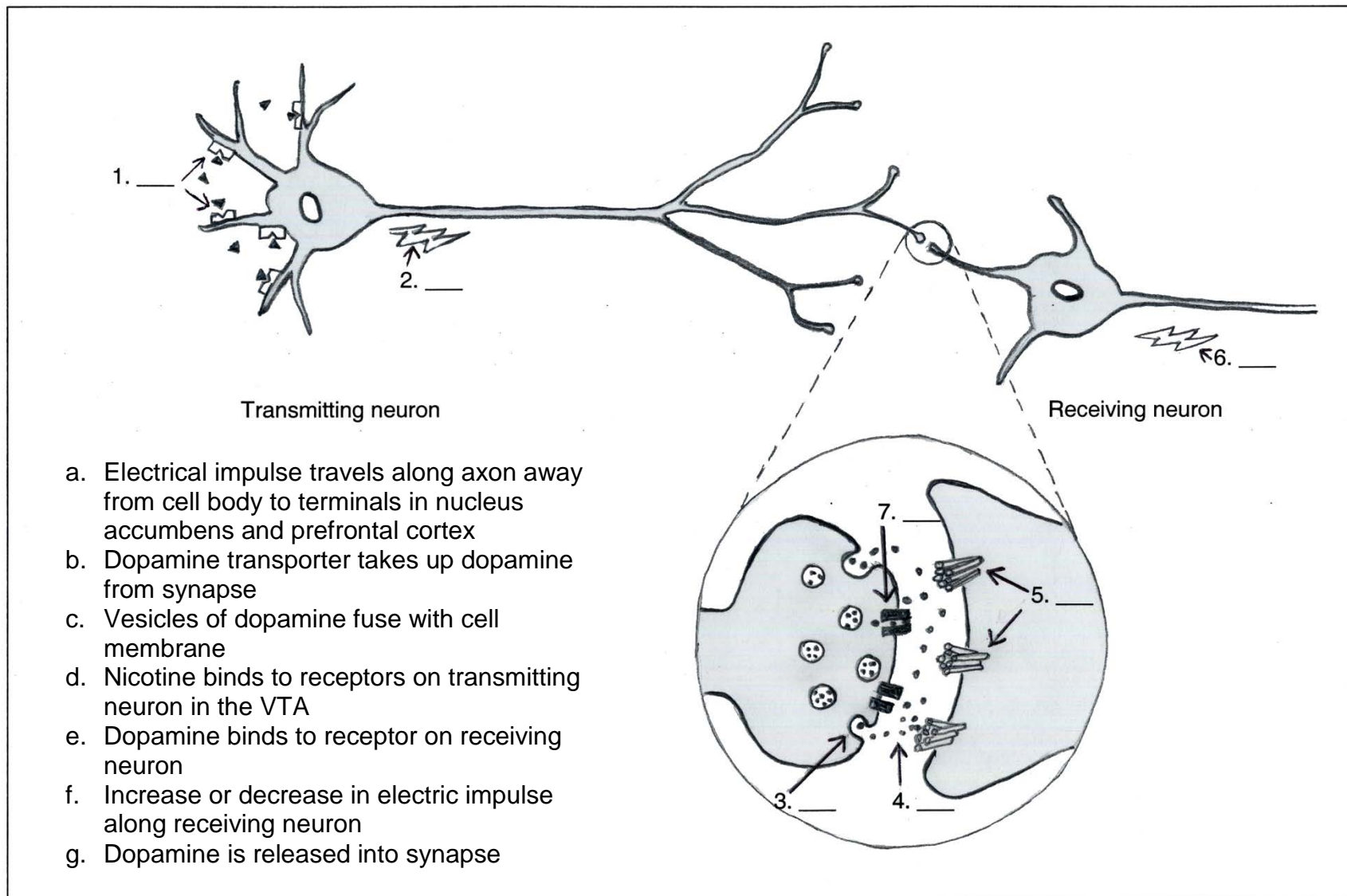
Student Sheet 3.4 (optional) Match the descriptions steps of neurotransmission (a through g) with steps 1 through 7 in the diagram. Circle the proteins that are genotyped in the smoking behavior research study.

STUDENT SHEET 4.1 **The Case Control Flow Sheet**

A case control study starts with an <u>event</u> or a <u>condition</u> that has occurred, for which the researcher wants to find the cause.

STEP 1: Identify the event or condition you wish to study. This is called the <u>outcome</u>.
We want to know: What caused the outcome?

```
                         Outcome
               (Condition under investigation)



```

STEP 2: Identify exposures- factors that may have caused the outcome Exposures are defined based on the researcher's hypothesis for what caused the outcome

```
   What are some exposures that might have caused the outcome of this study?



```

STEP 3: Experimental design: Compare 2 groups of people
Researchers compare people who have the outcome with people who don't have the outcome

```
      Cases                              Controls
   (Have outcome)      <---->         (Don't have outcome)
```
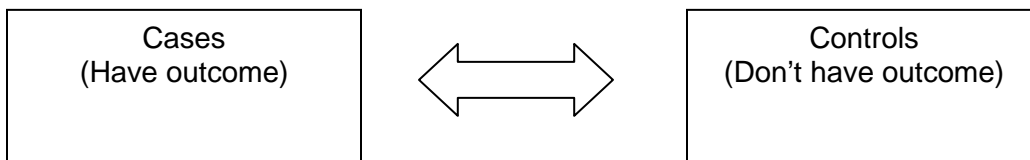
STEP 4: Measure exposures in both groups (cases and controls)

```
                   Some ways of measuring are:


```

STEP 5: Analyze incidence of exposures in cases compared to controls
The researcher asks: Is there a difference in the extent that cases and controls are exposed to a particular factor?

STUDENT SHEET 4.1, continued

Glossary

Epidemiology: *The study of the distribution and causes of disease and other health problems in groups of people. It is also the application of the information learned from such studies to the control of the health problems.*

Case control study: *A study that tries to identify factors associated with a particular condition by comparing case subjects, who have the condition, with control subjects, who do not.*

Study population: *The people who participate in the study (and the specific characteristics used to select them, such as age range, sex, ethnicity and race)*

Case: *A study participant who has the condition of interest for the study*

Control: *A study participant who does not have the condition of interest for the study*

Outcome: *The condition that is being studies (for example, having a particular disease or set of symptoms)*

Exposure: *A factor that may be associated with developing the outcome*

Matching: *The process of selecting case and control subjects so that they are similar in certain characteristics and thus are drawn from the same study population*

Observational study: *A study in which variables are not deliberately manipulated. Samples or subjects are observed without trying to change their environment*

Retrospective study: *A study that looks backward in time to identify factors that might be associated with a particular outcome*

STUDENT SHEET 4.2. **Comparing Experimental Studies and Case Control Studies**

Use the information in each of the two research studies described below to complete the blanks in Table 4.1.

Scenario
The Beta Fraternity House held a party on Friday night. By 4 pm on Saturday afternoon, 17 people who had been at the party had called Doug, the frat house manager, complaining of headaches and severe nausea. None of the frat brothers who were absent from the party reported feeling sick. What should Doug do to figure out what was making people sick?

Research Study 1 (case control study): Doug reported the outbreak to Student Health Services. The public health team there designed a case control study to try to find out what had made the students sick. Their study design consisted of giving a questionnaire and a health exam to as many students as they could locate who had been to the party, including those who got sick and those who didn't. Both groups of students in the study—those that got sick and those that didn't—included males and females in the age range of 18 to 22. The questionnaire asked students about what foods they had eaten, what beverages they had consumed, and how much alcohol they drank. It also asked them to list the people they had been in contact with during the party. During the health exam, the nurse measured the students' temperature, blood pressure, and other vital signs, as well as asking them how they felt. The students who took the questionnaire were asked to report back to Student Health Services one week after the party for a follow-up health exam. Results from the study were analyzed to see whether there were any differences between the students who got sick and those who didn't. They learned that 95% of the students who were sick had hugged and kissed Sally, who had just returned from a long trip. Those who did not hug and kiss Sally maintained their health. Sally did not feel good at the party, but she was excited about seeing her friends so went to it anyway.

Research Study 2 (experimental study): People stricken with the "Beta Bash" sickness were severely ill for 5-8 days, and many developed a rash. Public health researchers determined that all of the sick students had hugged and kissed Sally, who had just returned from doing research on desert mice in New Mexico. They hypothesized that Sally had contracted a viral or bacterial infection from the mice she was studying. To test whether the infectious agent could also make mice sick, they injected 20 Strain X lab mice with Sally's saliva. Another group of 20 mice were injected with saliva from a student who did not get the sickness. The two groups of mice were cared for in separate facilities and observed for the next four weeks. Researchers monitored the two groups of mice to see whether they got sick or died. Of the 20 mice injected with Sally's saliva, 17 were weak and lost weight, and three died. All of the control mice injected with saliva from the healthy student remained healthy through the study.

Table 4.1. Experimental study vs. case control study. Answer the questions in the table.

	Case Control study (students)	Experimental study (mice)
Kind of Study	Observational, retrospective (looking backward in time)	Experimental, prospective (looking forward in time)
Experimental design	Participants already have the condition or outcome being studied. The research looks backward in time to identify factors that might have caused that condition.	Participants are assigned to a treatment (the manipulated variable) and followed to see if they develop the condition or outcome.
Variables	The two groups of subjects are **matched** so that they are the same in as many features as possible. ***How are the subjects matched?***	**Controlled variables** are the variables that are kept the same among all samples or subjects ***What are the controlled variables?***
	The **exposure** is a factor <u>observed</u> by researcher to be different between the two groups of subjects. ***What are the exposures?***	The **manipulated (independent) variable** is the factor that is <u>deliberately changed</u> by the researcher. ***What is the manipulated variable?***
	The **outcome** is the factor that defines how the two groups of subjects <u>differ</u> and is <u>chosen</u> by the researcher at the start of the study. ***What is the outcome?***	The **responding (dependent) variable** is the factor that <u>changes</u> in response to the manipulated variable and is <u>measured</u> by the researcher. ***What is the responding variable?***
What experiment shows	The exposure is <u>associated with</u> the outcome. ***Were any of the exposures associated with outcome in Study 1?***	The manipulated variable <u>causes</u> the responding variable. ***Did the manipulated variable cause the responding variable in Study 2?***

STUDENT SHEET 4.3: **Identifying the purpose of questions in the questionnaire.**

Purpose of question type	Question numbers	Color
Determining whether subject is a case or a control	1 – 4, 101, 102	Red
Determining whether smokers qualify to be cases (see reading on Fagerstrom Test for Nicotine Dependence)	5 - 11	Blue
Matching cases and controls for age, sex, ethnicity, race,	74, 75, 76, 78	Green
Learning other demographic information about subject	73, 77, 79, 84, 85	Purple
Learning about cases or controls, but not both	23 - 25	Purple
Learning about multiple responses to a question (These questions not useful for some types of analysis)	44, 57, 81	Purple

Homework Assignment:
A quick internet search shows how common case control studies are. Prove it to yourself with the following activity:

1. Go to http://scholar.google.com/ and search on the term "a case control study."
2. Select one of the studies from the search.
3. Complete Student Sheet 4.1, the case control flow sheet for your selected study.

Student Sheet 5.1. **Epidemiologist Clip Guide**

https://gsoutreach.gs.washington.edu/exda-pd/lesson-5-this-is-an-example-lesson/

Jeff Stanaway, 3/3

Questions:
1. How often do epidemiologists use databases?

2. What do epidemiologists try to connect in their research?

3. Name four things that epidemiologists do to imply causality in case control studies?

STUDENT SHEET 5.2. **2x2 Table and the Car Passenger Case Control Study**

The contribution of passengers to car accidents—a case control study (2007)

Whether the presence of passengers in an automobile contributes to car accidents is an important safety issue and has implication for public policy and law. A study conducted in Perth, Western Australia in 2003-2004 assessed the contribution of passengers to accidents resulting in non-fatal injuries. The presence of passengers in cars involved in injury-causing accidents and cars not involved in accidents were compared. The study included drivers aged 17 and older. Controls were matched to cases by location (recruited from nearby service stations), time of day, day of the week, and road and driving conditions.

Among 274 drivers having injury car accidents, 73 had one or more passengers in the car. The control group consisted of 1096 drivers not involved in car accidents, and among these drivers, 190 had one or more passengers in the car.

1. Fill out the Case Control Study Design form below for the car passenger study

<div>

Case Control Study Design

Research Question:

Outcome:

Exposure:

Study Population:

Study Size:

Cases:

Controls:

</div>

STUDENT SHEET 5.2: **Car Passenger Case Control Study** *(continued)*

2. On the table below, label who the cases were and who the controls were.

3. What is the exposure (the factor you think may have caused the outcome)? Label the "Exposed" and "Not Exposed" rows.

4. Fill in the shaded boxes with the appropriate numbers from the study.

	Cases	**Controls**
Exposed		
Not Exposed		

Total		

5. What are the odds that a **case** was carrying 1 or more passengers? Show your work. Remember that odds is:

 number of times an event occurs (i.e. cases were carrying 1 or more passengers)
   number of times an event does not occur (i.e. cases were not carrying passengers)

       =

6. What are the odds that a **control** was carrying 1 or more passengers? Show your work. Remember that odds is:

 number of times an event occurs (i.e. controls were carrying 1 or more passengers)
number of times that event does not occur (i.e. controls were not carrying passengers)

       =

7. The odds ratio (OR) is a comparison between two ratios. What is the odds ratio for this study? Show your work.

OR = <u>Odds the cases have the exposure (odds that cases were carrying 1 or more passengers)</u>
        Odds the controls have the exposure (odds that controls were carrying 1 or more passengers)

=

8. What does the odds ratio tell us? Use the language in the following statement to say what your odds ratio means.

   If the odds ratio is X, then cases are X times more likely to be exposed than controls.

9. What would an odds ratio of 1 tell you?

10. What laws or public policies can you think of that may have been influenced by studies like this one?

STUDENT SHEET 5.3. **Mr. Limon's History Class**

Mr. Limon notices that each day a number of students fall asleep in his World History class. He thinks students are falling asleep because they don't get enough sleep at night. He decides to do a case control study to test his idea. He will ask students who fall asleep in class and those who don't whether or not they slept for 7 hours or more the night before and then compare these two figures.

The next day Mr. Limon passes out anonymous questionnaires to his four World History classes and collects the following data. Out of 120 students, 28 reported falling asleep in class yesterday. Of these students, 20 had slept less than 7 hours the night before. Among students who had stayed awake, 39 had slept less than 7 hours.

1. What is Mr. Limon's research hypothesis?

2. Why do you think Mr. Limon did not require students to put their names on the questionnaires?

3. Why did Mr. Limon give the questionnaire to all four of his World History classes instead of just Jake's class?

4. Who are the cases and who are the controls? Label the table below.

5. What is the exposure (the factor you think may have caused the outcome)? Label the "Exposed" and "Not Exposed" rows.

6. Fill in the shaded boxes with the appropriate numbers.

	Cases	**Controls**
Exposed		
Not Exposed		
Total		

STUDENT SHEET 5.3. **Mr. Limon's History Class** *(continued)*

7. What are the odds that cases slept less than 7 hours the night before? Show your work. Remember that odds is:

    number of times an event occurs (i.e. cases slept less than 7 hours)_____
    number of times an event does not occur (i.e. cases slept 7 or more hours)

    =

8. What are the odds that controls slept less than 7 hours the night before? Show your work.

    number of times an event occurs (i.e. controls slept less than 7 hours)_____
    number of times that event does not occur (i.e. controls slept 7 or more hours)

    =

9. The odds ratio (OR) is a comparison between two ratios. What is the odds ratio for this case control study? Show your work.

OR = Odds the cases have the exposure (odds that cases slept less than 7 hours)
        Odds the controls have the exposure (odds that controls slept less than 7 hours)

=

10. What does the odds ratio tell us? Use the language in the following statement to say what your odds ratio means.

If the odds ratio is X, then cases are X times more likely to be exposed than controls.

11. Do these results support Mr. Limon's hypothesis? Give evidence from your calculations to support your answer.

12. What would an odds ratio of 1 tell you?

EXPLORING DATABASES   APRIL 2013

STUDENT SHEET 5.4. **A Real Case Control Study**

Smoking and Lung Cancer. A case control study by Doll & Hill (1950)

The occurrence of lung cancer increased rapidly in the first half of the twentieth century. Why? There were several ideas. During this period, in western nations, manufactured cigarettes were plentiful and easy to obtain. Could increased smoking account for the increase in lung cancer cases? A classic case control study carried out in 1947 by English epidemiologists Sir Richard Doll and Tony Bradford Hill examined the relationship between smoking and lung cancer. Doll and Hill compared the smoking habits of hospitalized London lung cancer patients to the smoking habits of patients hospitalized for other causes. All the patients were men and under the age of 75. Here are their results:

	Cases (lung cancer)	**Controls** (no lung cancer)
Cigarette Smokers	1350	1296
Non-smokers	7	61
Total	1357	1357

1. Fill out the Case Control Study Design below for Doll and Hill's study.

Case Control Study Design
Research Question:
Outcome:
Exposure:
Study Population:
Study Size:
Cases:
Controls:

2. Using the data provided for this study, calculate the percentage of cases who smoked and the percentage of controls who smoked.

3. What can you conclude from these percentages and from comparing them?

4. What are the odds that a case will be a smoker?

5. What are the odds that a control will be a smoker?

6. What is the odds ratio for this study? Show your work.

7. What does this odds ratio tell you about smoking and lung cancer? Give evidence from your calculations to support your answer.

Original Reference
Doll, R. & Hill, A. B. (1950) Smoking and carcinoma of the lung: preliminary report. British Medical Journal 221, 739-748.

Supplementary References
http://www.nature.com/milestones/milecancer/full/milecancer08.html
Nature Milestones Cancer. Milestone 8. 1950 Smoking and Cancer: Smoking gun. by Ezzie Hutchinson, 2006.

CDC Excite curriculum on Cigarette Smoking and Lung Cancer
http://www.cdc.gov/excite/classroom/smoking_q.pdf
In depth exploration of Doll and Hill's classic case control and cohort studies on smoking and lung cancer.

STUDENT SHEET 5.5. **Reading: How do we know if an association is significant or occurred by chance?**

You have learned that an odds ratio (OR) is used to show association between an exposure and an outcome.

> **Odds ratio**: *a measure of the strength of the association between an exposure and an outcome.*

Review the figure below to remind yourself how the odds ratio is calculated.

	Cases	**Controls**
Exposed	a	b
Not Exposed	c	d

Odds that a case was exposed:

$$\frac{\text{number of cases exposed}}{\text{number of cases not exposed}} = \frac{a}{c}$$

Odds that a control was exposed:

$$\frac{\text{number of controls exposed}}{\text{number of controls not exposed}} = \frac{b}{d}$$
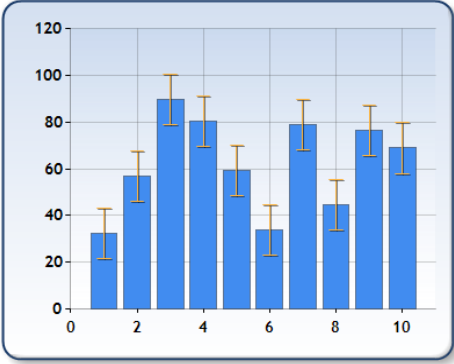
Odds ratio: $\dfrac{a/c}{b/d}$

"Cases are $\dfrac{a/c}{b/d}$ more likely to have been exposed than controls"

If we find an association (OR $\neq$ 1), then we need to determine whether that association is statistically significant or occurred simply by chance. In any case control study, we can only measure a sample of the total cases and controls in the entire population, so when we select subjects for the study, we try to get a good representation of the population. In general, associations that occur by chance are less likely when the sample size is large. Statistics, calculated from samples of the population, allows us to estimate the "true OR" found in the entire population.

STUDENT SHEET 5.5, continued

The 95% Confidence interval = error bar for odds ratios

Traditional error bars	95% Confidence interval
	A range of values that is believed to contain the true population OR with 95% confidence Expressed as: 95% CI = [lower bound, upper bound]
Often determined experimentally	Calculated: $\text{Lower Bound} = OR \times e^{-1.96 \times \left(\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right)}$ $\text{Upper Bound} = OR \times e^{1.96 \times \left(\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right)}$

In our analysis of data from the Smoking Behavior case control study, we will calculate the **95% confidence interval**, which is a range of values for the OR that we are 95% confident contains the true OR of the entire population. Another way to say this is if we repeated the study many times, drawing a different sample from the population each time, 95% of the time the confidence interval would contain the true OR of the population.

> ***Confidence interval:*** *a range of values that is believed to contain the true OR with 95% confidence. It is a measure of how confident we are that the OR in our study reflects the true OR of the population.*

The upper and lower bounds of the confidence interval are calculated using a complex formula involving the values of the OR and a, b, c, and d from the 2 x 2 table. In order for an association to be significant, the 95% confidence interval must not include the value 1. Remember, an OR = 1 means there is no association, so a statistically significant confidence interval should not include this value.

In general, ORs with small confidence intervals are more reliable than ones with large intervals. As the size of the study goes up, the confidence interval size decreases.

STUDENT SHEET 5.5, continued

The formula for the 95% confidence interval is given in the table above. This range of values, along with the OR, will be calculated automatically by the computer when you query the smoking behavior database.
.
Although you will not need to calculate the confidence interval yourself, you will need to know how the confidence interval is used to determine if the association between an exposure and outcome is significant. If you remember only one thing about the confidence interval, remember that it must not contain an OR=1 if the association is significant.

**How to interpret the 95% Confidence interval.**
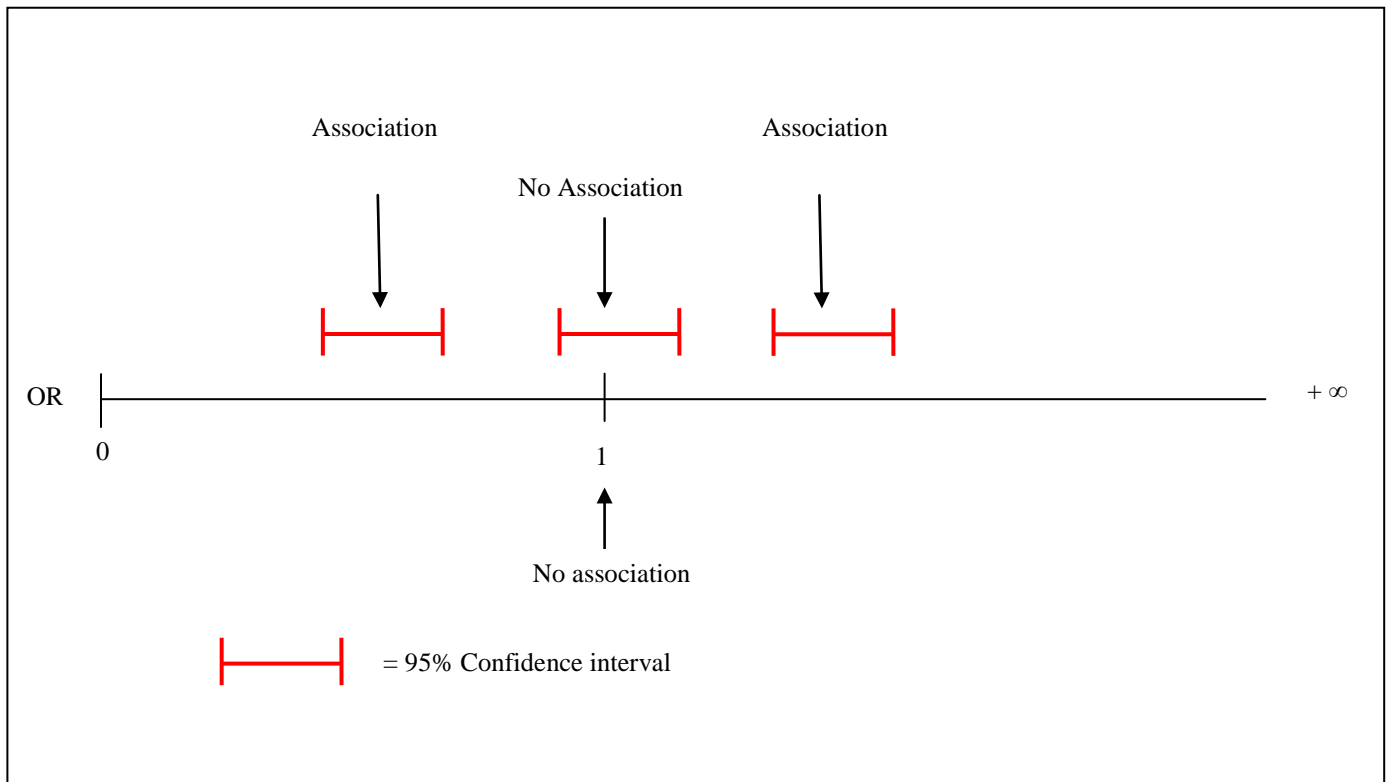
STUDENT SHEET 5.6 **Sources of Error in Case Control Studies**

What are some possible explanations for a strong association between an exposure and outcome but yet few criteria for causality are met?

Random error. A detected association could be due to chance. For example, the sample may not be representative of its source population. This type of false association may disappear if the study is repeated with a larger sample size or with a new sample drawn from the same population. When you use the Smoking Behavior database to look for associations, you will be able to check for the possibility of random error with a statistical measure called the **confidence interval**.

Systematic error. An association could be due to a systematic error in the study design. This type of error is called **bias** and does not diminish when the sample size is increased or when the study is repeated with a new sample. When you detect an association using the Smoking Behavior database, you will have to think carefully about whether it could be due to bias. There are several types of bias that are especially important in case control studies:

- **Selection bias** occurs when study participants are different than the population from which they are drawn, or when cases and controls come from different populations. Imagine a case control study that looks at whether exposure to car exhaust might cause asthma. The cases are asthma patients in an urban primary care clinic. The controls are patients in a rural dental clinic. Controls may have been exposed to car exhaust less than cases because of their location, leading to an association that may be false. Cases and controls should be drawn from the same area.

- **Information bias** occurs when there is an error in the measurement of the outcome or the exposure for some subjects. One type of information bias is called **recall bias**. This type of bias can occur if cases are better able to <u>recall</u> possible exposures than controls. For example, people with skin cancer may be more likely to recall times when they had extreme sunburns than people without skin cancer. This would result in a stronger association between extreme sunburns and skin cancer than the true strength of association (odds ratio).

- **Confounding** occurs when another factor is associated with both exposure and outcome, giving the appearance that the exposure and outcome are associated (see diagram that follows). Confounding is likely to occur when cases and controls not only differ in the exposure of interest but also in other exposures that may affect the same outcome. In case control studies, careful **matching or restriction** of cases and controls can help reduce possible confounding factors.

| Exposure | — ? → | Outcome |

Confounding factor
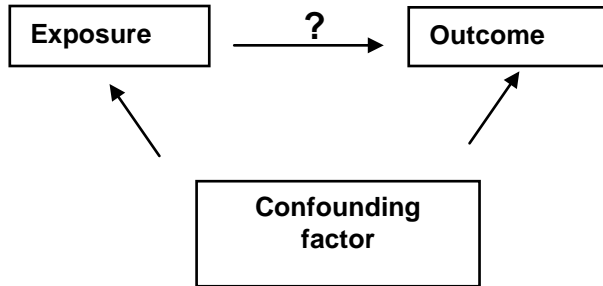
Could a confounding factor be associated with both the outcome and the exposure and account for some or all of the observed association between the two?

| Gambling | — ? → | Lung cancer |

Smoking?
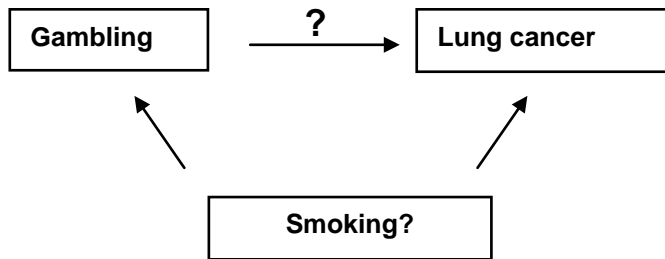
Gambling has been associated with cancer, but it seems unlikely that gambling would cause cancer (it does not make biological sense). What factors might be associated with both gambling and cancer?

How can we determine whether gambling causes cancer, or if another factor, like smoking, is associated with both gambling and cancer? During data analysis, we can **stratify** (separate) our data based on a suspected confounding factor (e.g. smoking). We would calculate the odds ratio for smokers and non-smokers separately, as shown in the two 2x2 tables below. If the resulting odds ratio is 1 for both groups then there we cannot conclude that gambling is directly related to cancer. .

Smokers

	Cases (have lung cancer)	Controls (no lung cancer)
Exposed (gambling)		
Not exposed (no gambling)		

Nonsmokers

	Cases (have lung cancer)	Controls (no lung cancer)
Exposed (gambling)		
Not exposed (no gambling)		

Confounding can also be minimized by designing the study carefully. We can try to **match** cases and controls for other known risk factors (exposures) for the outcome, or we can **restrict** the study to participants who do not have a suspected confounding factor, for example, by only including non-smokers in the gambling-lung cancer study.

STUDENT SHEET 5.6, Continued

Test Your Knowledge
Which of the following, made-up studies is an example of each type of error?

Type of error	Study number
a. Random error	
b. Selection bias	
c. Information bias	
d. Confounding	

Study 1: Sam hypothesized that smokers are more likely to have believed they could not become addicted to cigarettes during their experimental smoking phase. Therefore, Sam performed an odds ratio calculation from the smoking behavior database for Question 20: During your experimental smoking phase, did you believe that you could become addicted to cigarettes? Sam found that smokers were 1.23 times more likely to not believe they could become addicted compared to non-smokers. However, the 95% confidence interval (0.76 to 1.98) contains the number one, which means that there is not an association between believing you could become addicted to cigarettes and becoming a regular smoker.

Study 2: In a study of chronic back pain, cases were 2.5 times more likely than controls to recall having over-exerted themselves lifting heavy objects in the past 10 years.

Study 3: A recent study suggested that people who are overweight are less likely to attend college than people who are in the "normal" weight range. However, when the data were stratified for family income (high, average, and low), there was no observable association between being overweight and not attending college.

Study 4: Several participants mentioned that they learned about the study and the $30 gift card at the homeless shelter they frequented. This group was almost entirely smokers, and they tended to be older compared to control participants. On the other hand, several study subjects expressed delight that they could contribute to a study that involved high school students in doing research on why people smoke. In general, this population tended to be younger and was much more likely to be non-smokers. Based on these facts, what bias could occur in question 83: How well off is your current family/household?

Student Sheet 7.1. Using Hypothesis Generation to Explore Data

Directions
Work as a class to complete this activity, which models how to explore the data using Question 12.

1. Your teacher will project the database:
    https://gsoutreach.gs.washington.edu/database2/

2. Watch as your teacher selects Step 1.4 Hypothesis Generation, and then Question 12, "How old were you when you first tried a cigarette, even a single puff?"

3. Notice that in this view, you can see the number of cases and controls who gave each response to the question as both a table and a graph.

4. As a class, decide which responses you will use to define exposed and not-exposed. Select them, choose "Everyone" for the study population, and then "Get odds ratio."

5. In Query 1 below, write the responses you used for exposed and not-exposed, and fill in the 2x2 table, the odds ratio, the 95% confidence interval, and the sample size.

6. In Query 2, use different responses to define exposed and not-exposed, and again use "Everyone" for study population. Select "Get odds ratio." Hint: think about how to get a higher odds ratio.

7. In Query 3, use the same responses to define exposed and not-exposed as you used in Query 2, but select Male or Female for the study population.

Query 1

Exposed: Responses to question 12 _____ and description in words:

Not exposed: Responses to question 12 _____ and description:

Study population = Everyone

	Cases	**Controls**	
Exposed			Odds Ratio ____
	a=	**b=**	95% confidence interval _____
Not Exposed			Sample size _____
	c=	**d=**	

Query 2

Exposed: Responses to question 12 _____ and description in words:

Not exposed: Responses to question 12 _____ and description:

Study population = Everyone

	Cases	Controls
Exposed	a=	b=
Not Exposed	c=	d=

Odds Ratio _____

95% confidence interval _____

Sample size _____

Query 3

Exposed: Responses to question 12 _____ and description in words:

Not exposed: Responses to question 12 _____ and description:

Study population = Males or Females

	Cases	Controls
Exposed	a=	b=
Not Exposed	c=	d=

Odds Ratio _____

95% confidence interval _____

Sample size _____

Questions

1. Comparing Queries 1 and 2, what changes in odds ratio and 95% confidence interval do you notice as the way you defined exposed and not-exposed and thus the numbers in the four cells of the 2x2 table change?

2. Comparing Queries 2 and 3, how does changing the study population affect the odds ratio and 95% confidence interval?

Name:_____ Date: _____ Period: _____

Appendix B
Research Project Pages (RPP): Conducting your Research Investigation

The following pages provide documents to help you conduct your research investigation. Throughout this unit, you are addressing the question: "Why do some people become regular smokers and others do not?" by the carrying out these steps:

- Develop an overarching hypothesis from smoker profiles, your intuition/observations, and past research (Lessons 1-5).

- Create specific hypotheses that relate to your overarching hypothesis from the Smoking Behavior database questions (Lesson 6).

- Analyze many questions from the Smoking Behavior database to generate new hypotheses for future studies (Lesson 7).

The above bullet points are broken down into the following sections of RPP:

Name:_____ Date: _____ Period: _____

Lesson 1. Preliminary Ideas about Smoking Behavior

Write a paragraph describing three or four reasons why you think some people smoke and others do not. What evidence do you have for each reason you gave?

Name:_____ Date: _____ Period: _____

Lesson 2. Research Topic and Overarching Hypothesis

Directions
Work with your research team to provide the following information. Each person should complete their own sheet.

1. Research Team name:

2. Research Team members:

3. What is your research topic?

4. Why does this topic interest you?

5. How does your research topic relate to the driving question, "Why do some people become regular smokers and others do not?

6. Do a Google Scholar search to find out what other studies have shown related to your research topic. Summarize key findings below.

7. What is your overarching hypothesis?

8. What evidence from the smoker profiles, prior studies, and your own experience did you use to form this hypothesis?

9. List 4-5 question numbers from the research questionnaire that will help you to test your hypothesis.

Name:_____ Date: _____ Period: _____

Lesson 3. Genetic Component to Research Topic

1. Your group may be focused on a topic that is mainly environmental. Do you think that there is also a genetic component to the topic/factor you are investigating?

2. In a paragraph, explain your thinking. How do you think this might affect a person who is experimenting with smoking?

Name:_____ Date:_____ Period:_____

Lesson 4. The Smoking Behavior Case Control Study

What is the **research question** of the study?

What are the **outcome**, **cases**, and **controls** of the study?
Outcome:
Cases:
Controls:

What is the **study population** (check how the **study sample** was determined)?

What are the characteristics of the **study sample**? Answer this question by filling the actual numbers in the boxes with bold borders

Intended Characteristic	Category	Cases	Controls
Equivalent ratio of men and women	Men:		
	Women:		
Equivalent ratio across age	25-29:		
	30-34:		
	35-39:		
	40-44:		
	45-49:		
	50-54:		
Diverse racial backgrounds	American Indian / Alaskan Native:		
	Asian:		
	Black / African American:		
	Native Hawaiian / Other Pacific: Islander		
	White		

Answer the following questions

1. Do you find any bias in the **study sample**? How do you know there is a bias?

2. Why do you think there might have been problems properly matching cases and controls?

3. How might the problems matching cases and controls (as far as gender, age, and racial group) affect the results of your investigation? Refer to your specific exposure(s).

Name:_____ Date:_____ Period:_____

Lesson 6a. Hypothesis Testing

Please see the Directions for the Research Project Pages to complete the following RPP.

Go to https://gsoutreach.gs.washington.edu/database2/

Big Research Topic:

Overarching Hypothesis:

Complete this section before entering database

Your 1st database question (give question # and question):

Your specific hypothesis for this question:

Exposed: Responses _____ and description in words:

Not exposed: Responses _____ and description in words:

Why did you define "exposed" and "not-exposed" this way?

Study population (select one): Everyone__ Males__ Females__

Testing your Hypothesis: Go to "Hypothesis Testing" to complete this page.

Odds Ratio: _____   Sample size: _____

95% CI: _____   Does the 95% CI contain the value 1? _____

Interpret the result:
a) Use the odds ratio in a sentence that describes what it means.

b) Is there an association between the exposure and the outcome? How do you know?

c) Do you think this exposure <u>causes</u> people to become a regular smoker? Apply the criteria for causality to support your answer.

Name:_____ Date:_____ Period:_____

Lesson 6a. Hypothesis testing, continued

Your 2nd database question (give question # and question):

Your specific hypothesis for this question:

Exposed: Responses _____ and description in words:

Not exposed: Responses _____ and description in words:

Why did you define "exposed" and "not-exposed" this way?

Study population (select one): Everyone__ Males__ Females__

Odds Ratio: _____   Sample size: _____
95% CI: _____   Does the 95% CI contain the value 1? _____

Interpret the result:
a) Use the odds ratio in a sentence that describes what it means.

b) Is there an association between the exposure and the outcome? How do you know?

c) Do you think this exposure <u>causes</u> people to become a regular smoker? Apply the criteria for causality to support your answer.

Name:_____ Date:_____ Period:_____

Lesson 6a. Hypothesis testing, continued

Your 3rd database question (give question # and question):

Your specific hypothesis for this question:

Exposed: Responses _____ and description in words:

Not exposed: Responses _____ and description in words:

Why did you define "exposed" and "not-exposed" this way?

Study population (select one): Everyone__ Males__ Females__

Odds Ratio: _____   Sample size: _____
95% CI: _____   Does the 95% CI contain the value 1? _____

Interpret the result:
a) Use the odds ratio in a sentence that describes what it means.

b) Is there an association between the exposure and the outcome? How do you know?

c) Do you think this exposure <u>causes</u> people to become a regular smoker? Apply the criteria for causality to support your answer.

Name:_____ Date:_____ Period:_____

Lesson 6a. Hypothesis testing, continued

Your 4th database question (give question # and question):

Your specific hypothesis for this question:

Exposed: Responses _____ and description in words:

Not exposed: Responses _____ and description in words:

Why did you define "exposed" and "not-exposed" this way?

Study population (select one): Everyone__ Males__ Females__

Testing your Hypothesis: Go to "Hypothesis Testing" to complete this page.

Odds Ratio: _____   Sample size: _____

95% CI: _____   Does the 95% CI contain the value 1? _____

Interpret the result:
a) Use the odds ratio in a sentence that describes what it means.

b) Is there an association between the exposure and the outcome? How do you know?

c) Do you think this exposure <u>causes</u> people to become a regular smoker? Apply the criteria for causality to support your answer.

Name:_____ Date:_____ Period:_____

Lesson 6b. Mapping Activity

Directions: Fill in your overarching hypothesis in the oval below; then fill in each square with one of the four database question numbers and specific hypotheses. Using one of the three lines shown in the key, draw the relationship between each | Question # |

and | Becoming a Regular Smoker | at the bottom of the page.

Key

← → Association

→ Causation

− − − Neither

Overarching Hypothesis

Question # Specific Hypothesis	Question # Specific Hypothesis	Question # Specific Hypothesis	Question # Specific Hypothesis

Becoming a Regular Smoker
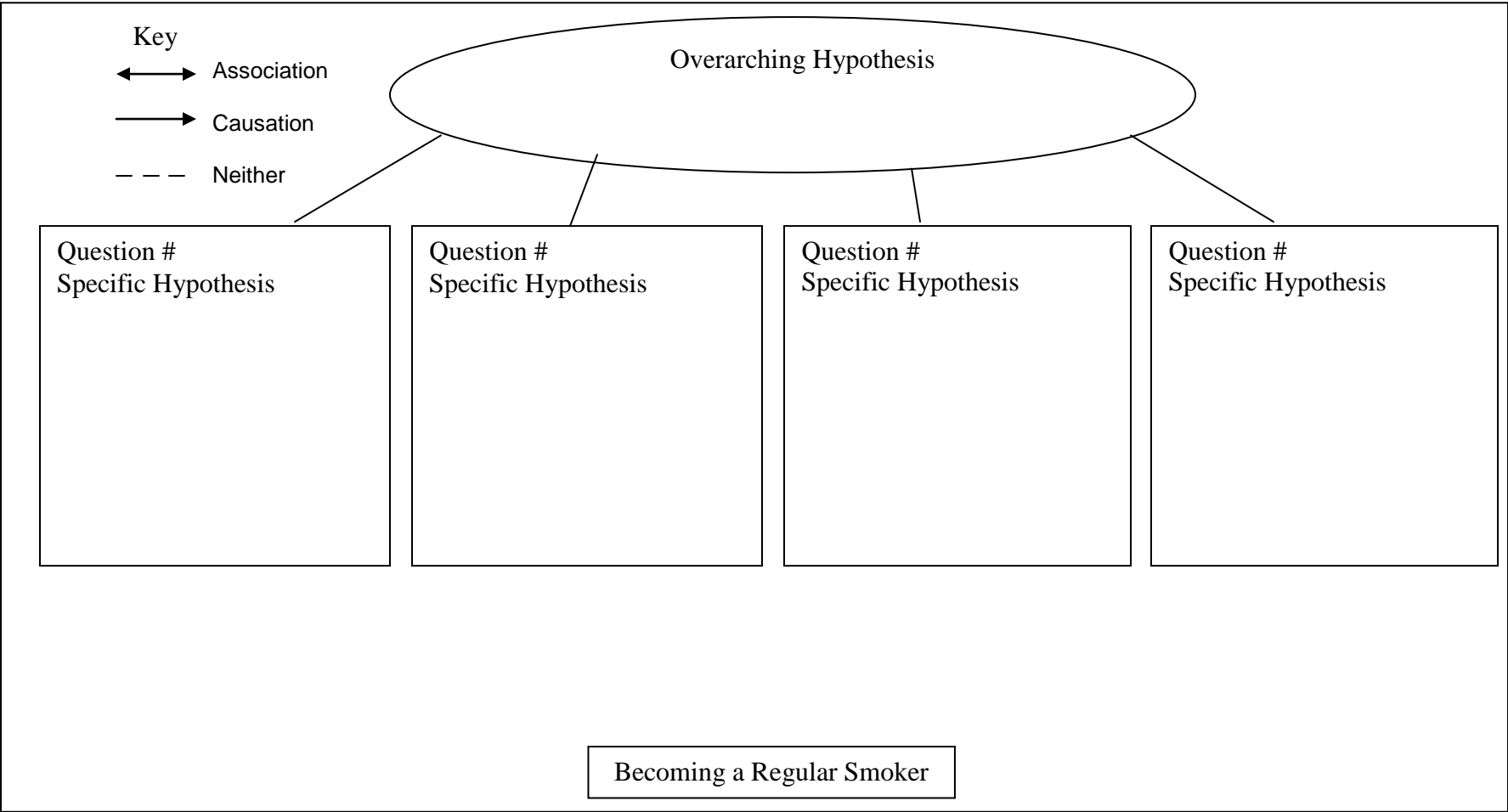
Name:_____ Date:_____ Period:_____

Lesson 6c. Drawing Conclusions

You will now look at the results for each of your four questions on RPP-7-10 and on the mapping activity to draw conclusions about your overarching hypothesis.

1. What do your results show about your overarching hypothesis?

2. Which pieces of evidence most strongly support your claim?

      a. Discuss one piece of evidence and why it supports your claim.

      b. Discuss a second piece of evidence and why it supports your claim.

3. How strong is the evidence that your overarching hypothesis is supported? Are there other ways to interpret your data?

4. What are the implications of your research for health? For your community?

Name:_____ Date: _____ Period: _____

Lesson 7a. Planning your Hypothesis Generation Research

During hypothesis testing, you were constrained in your queries to:

a. Use four questions from questionnaire
b. Plan and define your queries before doing them
c. Define exposed and not-exposed one way only for each question

In Hypothesis Generation, these constraints no longer apply because you are using your results to propose a future study, not to draw scientific claims. In the Hypothesis Generation view of the database, you see case and control responses to the questions before you do the query, so you can deliberately define exposed and not-exposed to explore the effects on odds ratio.

After you and your teammates have had the chance to explore several questions, you will compare notes and then work together to propose a new case control study.

Work with your research team to answer the following questions:

1. Describe two research topics that you would like to investigate in Hypothesis Generation, and explain why these topics interest you. One of your topics can be the one you used during Hypothesis Testing.

Research topic 1:

Why does this topic interest you?

Research topic 2:

Why does this topic interest you?

2. List at least four questions that interest your team and are related to one of the two research topics. You can include questions that you used in Hypothesis Testing as well as new questions. You may also include questions that only have graphical information.

Questions related to research topic 1: Q# _____, Q# _____, Q# _____

Questions related to research topic 2: Q# _____, Q# _____, Q# _____

Name:_____ Date: _____ Period: _____

Lesson 7b. Conducting Queries in Hypothesis Generation
Follow the instructions for **Lesson 7b** in the Directions for Student Research Pages. Go to "Step 4, Hypothesis Generation" in the database to fill in the table below. You may wish to do several queries on the same question. Record your results for each query in the table below. Use the Notes section to record your observations about trends in the data and interesting results.

Table 1: Keeping Track of Hypothesis Generation Searches

Question #	Exposed	Not Exposed	OR	CI	CI contains 1?	Odds Ratio Statistically Significant?	Sample Size
Notes:							
Notes:							
Notes:							
Notes:							

Name:_____ Date: _____ Period: _____

Lesson 7c. Hypothesis Generation and Proposing a New Case Control Study

Meet with your research team to complete this page after you have completed your
Hypothesis Generation queries.

1. Share the data you recorded in Table 1: *Keeping Track of Hypothesis Generation
   Searches* with your teammates. Record the most interesting queries in the table
   below. They may be on one or more research topics.

Research Topic	Question #	Why topic and question are interesting

2. Look at the data in the table above and decide which research topic your team will
   address in a proposed new case control study. Why did you choose this topic?

3. What trends do you notice in the queries you did on the questions related to this
   research topic?

4. What are any other questions you would like to ask related to this topic?

Lesson 7c, continued

Designing your proposed case control study:

5. Formulate a new hypothesis for your proposed case control study. What evidence did you use from your Hypothesis Generation queries to form this hypothesis?

6. What is the outcome of your proposed case control study?
   Note: It does not need to be the same as the outcome for this study. For example, you may be interested in designing a study focused on smoking cessation.

   Why did you choose this outcome?

7. Describe up to four exposures you would study related to your research topic and explain why you are interested in each exposure.
   a.

   b.

   c.

   d.

8. In your proposed case control study, what are the characteristics of your study population? *Note: your study population can be different from the population in the original Smoking Behavior study. For example, you may want to study males only, or people from a certain race/ethnicity or age bracket.*

   Why did you choose this study population?

Name:_____ Date: _____ Period: _____

Lesson 7c, continued

9. What are the characteristics of the cases and controls in your proposed study? *Note: They can be different from the cases and controls in the original Smoking Behavior study.*

    Why did you define cases and controls in this way?

10. Use your responses to Questions 2-9 above to complete the case control study design template.

Case Control Study Design

Research Topic:

Hypothesis:

Outcome for new study:

Exposures you will test:

Characteristics of Study Population:

Characteristics of Cases:

Characteristics of Controls:

Appendix C: Directions for Research Project Pages

Lesson 6: Hypothesis Testing

6a. Hypothesis Testing

1. Turn to Lesson 6a, Hypothesis Testing (RPP 7-10) in your Research Project Pages packet. Working with your research team, complete the top sections with your research topic, overarching hypothesis, and the questions you want to analyze. Write a specific hypothesis for each question.

2. Define "exposed" and "not-exposed" for each question you will analyze, using the responses to that question.
   Remember:
   a. Exposed: The responses to the question that match the factor you think might increase the risk of someone becoming a regular smoker. For example, believing smoking is not harmful for your health (response b to Q19) OR having a least 1 parent who smoked (responses b, c to Q26).

   b. Not-Exposed: The responses to the question that do not match having the factor. For example, believing smoking is harmful for your health (response a to Q19) OR not having a parent who smoked (response a to Q26).

   Also remember:

   - **"Don't know/not sure" should <u>not</u> be used as an exposure or non-exposure** because the study subjects were unsure if they were exposed or unexposed.

   - **Use all other responses to define either exposed or not-exposed.** This will ensure that you have as large a study population as possible, which will lead to a small 95% confidence interval.

   - **Each question from the database should relate to your overarching question/hypothesis.**

3. Select your study population (Everyone, Males, or Females).

4. Go to the database:
   https://gsoutreach.gs.washington.edu/database2/.

5. On the menu bar on the left side of your screen, click "Step 1.3 Hypothesis Testing," and select the question you wish to use in your OR calculation.

6. Use the information you already wrote in the Research Project Pages to define exposed and not-exposed. Beside each response, click either Exposed, Not-exposed, or Neither. Select the study population. Then click the "Get Odds Ratio" button. The analysis will come up below.

7. Keep track of the output by writing down the Odds Ratio, CI, and sample size in the Research Project Pages. Complete the three questions under "interpret the result" either by typing in your responses in the database or by writing them in the Hypothesis Testing pages. Here are some hints for each question:

   a. If your odds ratio is greater than one or less than one, there is an association. The association is significant if the 95% CI does not contain the value 1. The meaning of the OR is, "Regular smokers are (OR) x more likely than nonsmokers to have had the exposure (fill in the OR you calculated and the exposure you used for this question).

   b. Use the Criteria for Causality to decide whether this exposure is likely to cause someone to become a regular smoker.

   c. Relate the results of this query and your interpretation to your specific hypothesis for this question.

6b. Mapping Activity (Summarize Hypothesis Testing Results)

1. Fill in the oval with your overarching hypothesis, and fill in each square with one of the four database questions and the response(s) you used for the exposure.

2. Using one of the three lines shown in the key, draw the relationship between each question and becoming a regular smoker.

6c. Drawing Conclusions

You will now look at the results for each of your four questions on RPP-7-10 and on the mapping activity to draw conclusions about your overarching hypothesis.

Lesson 7: Hypothesis Generation

7a.   Planning your Hypothesis Generation Research
   1. Complete RPP-13 as a team before going to the computers.

7b.   Conducting Queries in Hypothesis Generation

   1. Work in pairs, and have each pair complete queries on half of the questions that your team wants to investigate. You can do additional questions if you have time.

   2. Access the database by going to:
       https://gsoutreach.gs.washington.edu/database2/.

   3. At the left side of the database, select Step 1.4 Hypothesis Generation.

   4. Select the first question you're interested in querying. Based on the data shown in the table and the graph, decide how you are going to define exposed and not-exposed. Choose the study population, and then select "Get odds ratio."

   5. Record your query and results in Table 1 on RPP-14. Fill in the question #, responses used for exposed and not-exposed, the OR, CI, etc. The Notes section is for any interesting observations and trends.

   6. Do another query with the same question, but change either the study population or how you define exposed and not-exposed. Record your data in Table 1.

   7. Conduct at least three more queries using the questions identified by your team (and other questions as well if you have time). For each query, use the data table and graph to help you define exposed and not-exposed so that you get the highest possible odds ratio. Record all of your searches in Table 1. After you fill up one table, get a clean copy from your teacher if you want to do additional searches.

7c.   Hypothesis Generation and Proposing a New Case Control Study

   1. Meet with the rest of your team and complete RPP-15-17.

Cheat Sheet: Odds Ratio Output Definitions and Hints

Cases: Regular Smokers

Controls: People who tried or experimented with smoking but never became regular smokers

Exposed: The factor you think might have an influence on someone becoming a regular smoker. For example, believing smoking is not harmful for your health OR having a least 1 parent who smoked.

Not Exposed: The factor you think might protect people from becoming regular smokers. For example, believing smoking is harmful for your health OR not having a parent who smoked.

2 x 2 Table: Is the way people organize data in a case control study. It tell you how many smokers and how many nonsmokers fall into the exposed or not exposed categories. How you define exposed and not exposed depends on how you drag and drop the answers to this question.

Sample Size: The total number of people who responded to this question, both cases and controls.

Odds Ratio: An odds ratio of 1 means there is no association of the exposure between regular smokers and nonsmokers.

95% Confidence Interval: The confidence interval is a tool to help you decide if your result is meaningful to the entire population- people in general. If the confidence interval has the number 1.00 in it, this means that even if your Odds Ratio is bigger or smaller than 1.00, there is not an association between the exposure you identified and regular smokers. However if 1.00 is not inside your Confidence Interval, this means there is an association between the exposure and becoming a regular smoker.

Risk Factor: An exposure (such as believing that smoking is not harmful to one's health) associated with an increased <u>risk</u> of outcome (such as becoming a smoker). Such an exposure is a risk factor when there is an association (OR >1, 95% CI does not contain one) and the causality criteria is fulfilled.

Protective Factor: An exposure (such as bicycle helmets) associated with a reduced risk of outcomes (head injuries). Such an exposure is protective when there is an association (OR <1, 95% CI does not contain one) and the causality criteria is fulfilled.

Criteria for Causality

Criteria for causality
Strength of association. A strong association between the exposure and outcome is demonstrated using statistical methods. (The larger the odds ratio, the stronger the association.)
Dose-response relationship. An increased dose of the exposure is associated with a greater risk for having the outcome. (For example, having more than 2 passengers is associated with a greater likelihood of getting into a car accident than having only 1 passenger.) <u>May not be applicable to the smoking behavior database study</u>.
Temporal sequence. The exposure must occur <u>before</u> the outcome. (Sometimes in case control studies, this can be difficult to verify.)
Consistent with other studies. The result should be mostly consistent with what is already known in the field. (If it is not, there is always the possibility that you have discovered something new and unexpected, but there is also the possibility that your study design or assumptions were in some way flawed.)
Biological plausibility. The result should make biological sense. (For example, it makes sense that having passengers in the car would cause one to get into a car accident.)
Lack of confounder or significant bias. *Can the association be explained by another factor?* Is there a factor or bias that explains the association?

Why Some People Smoke and Others Don't

Many people try smoking, but not everyone continues to smoke. In this study we are trying to find clues about environmental and genetic factors that may influence whether a person becomes a regular (daily) smoker. Please do your best to answer each question thoroughly. Each question was chosen carefully, and is designed to increase our understanding of why some people become smokers and others don't.

CIRCLE THE BEST ANSWER FOR EACH QUESTION.

Today's Date:

_____ / _____ / _____
MM   DD   YYYY

Definitions
The terms below are used throughout this survey. This list of definitions is provided for your easy reference.

Regular smoker. A person who smokes every day or almost every day for 30 days straight.

Experimental smoking phase. The period when a person is first starting to smoke, beginning with the time they smoke their first complete cigarette and extending until they either become a regular smoker or quit.

Parents/guardians. Adults who had regular responsibility for raising you. Can include parents, step-parents, or other caregivers.

Your smoking and non-smoking practices and attitudes

Your smoking history
The following questions ask about how much you currently smoke, if at all, as well as how much you have smoked in your lifetime.

1. **About how many cigarettes have you smoked in your entire life?**
   a. 10 cigarettes or less
   b. 11 – 100 cigarettes (up to 5 packs)
   c. More than 100 cigarettes (more than 5 packs)

2. **At any time in your life, have you smoked 1 or more cigarettes per day for 30 days straight?**
   a. Yes
   b. No

3. **Do you currently use any of the following tobacco or nicotine products: chewing tobacco or snuff, cigars, tobacco pipes, clove cigarettes or bidis, nicotine replacement products such as gum or patch, or any other tobacco products besides cigarettes?**
   a. Yes
   b. No

4. **When was the last time you smoked?**
   a. Over 12 months (1 year) ago → GO TO QUESTION #12
   b. Between 1 and 12 months (1 year) ago → GO TO QUESTION #12
   c. Within the last 30 days (1 month) → GO TO QUESTION #5

5. **During the past 30 days (1 month) on how many days did you smoke cigarettes?**
   a. Every day or almost every day
   b. Some days
   c. No days

6. **How soon after you wake up do you smoke your first cigarette?**
   a. After 60 minutes
   b. 31-60 minutes
   c. 6-30 minutes
   d. Within 5 minutes

7. **Do you find it difficult to refrain from smoking in places where it is forbidden?**
   a. No
   b. Yes

8. **Which cigarette would you hate most to give up?**
   a. The first cigarette in the morning
   b. Any other

9. **How many cigarettes per day do you smoke?**
   a. 10 or less
   b. 11-20
   c. 21-30
   d. 31 or more

10. **Do you smoke more frequently during the first hours after awakening than during the rest of the day?**
    a. No
    b. Yes

11. **Do you smoke even if you are so ill that you are in bed most of the day?**
    a. No
    b. Yes

Experimental Smoking

The next questions refer to your experimental smoking phase, the time when you were first starting to smoke. This phase begins when you first smoke a complete cigarette and extends until you either become a regular (daily or almost daily) smoker or quit.

12. **How old were you when you first tried a cigarette, even a single puff?**

    a. I have never smoked cigarettes
    b. 7 years old or younger
    c. 8 or 9 years old
    d. 10 or 11 years old
    e. 12 or 13 years old
    f. 14 or 15 years old
    g. 16 or 17 years old
    h. 18 – 21 years old
    i. 22 years old or older

13. **How old were you when you first smoked an entire cigarette?**

    a. I have never smoked cigarettes
    b. 7 years old or younger
    c. 8 or 9 years old
    d. 10 or 11 years old
    e. 12 or 13 years old
    f. 14 or 15 years old
    g. 16 or 17 years old
    h. 18 – 21 years old
    i. 22 years old or older

14. **About how long did your experimental smoking phase last?**

    * If you are a current non-smoker, please indicate the length of time you experimented with smoking before you stopped completely.
    * If you are a current smoker, please indicate the amount of time you experimented with smoking before becoming a regular (daily or almost daily) smoker.

    a. Less than 1 month
    b. At least 1 month, but less than 6 months
    c. At least 6 months, but less than 12 months (1 year)
    d. At least 1 year, but less than 2 years
    e. At least 2 years, but less than 4 years
    f. 4 years or more
    g. Don't know/not sure

15. **During your experimental smoking phase, you may have smoked more frequently at some times and less frequently at others. When you were smoking most frequently during this phase, about how often did you smoke?**

    a. Daily or almost daily
    b. At least weekly, but less than daily
    c. At least monthly, but less than weekly
    d. Less than monthly
    e. Don't know/not sure

16. **During your experimental smoking phase, you may have smoked more cigarettes at some times than others. When you were smoking the most during this phase, about how many cigarettes did you smoke on days when you did smoke?**

    a. less than 1 cigarette per day
    b. 1 cigarette per day
    c. 2-5 cigarettes per day
    d. 6-10 cigarettes per day
    e. 11-20 cigarettes per day
    f. 21 or more cigarettes per day
    g. Other

17. **During your experimental smoking phase, where did you *usually* smoke? (Select only one response.)**

    a. At home
    b. At school
    c. At work
    d. At friends' homes
    e. At social events
    f. In public places (e.g. parks, shopping centers, street corners)

18. **During your experimental smoking phase, did you experience any of the following regularly?**

	Yes	No	Don't know/ not sure
A pleasurable feeling	☐	☐	☐
A "buzz"	☐	☐	☐
A good taste in your mouth	☐	☐	☐
A bad taste in your mouth	☐	☐	☐
A relaxed feeling	☐	☐	☐
An increase in focus or concentration	☐	☐	☐
A decrease in stress	☐	☐	☐
A craving for cigarettes	☐	☐	☐
A decrease in appetite	☐	☐	☐

19. **During your experimental smoking phase, did you believe that smoking cigarettes could be harmful to your health?**
   a. Yes
   b. No
   c. Don't know, not sure

20. **During your experimental smoking phase, did you believe that you could become addicted to smoking cigarettes?**
   a. Yes
   b. No
   c. Don't know, not sure

21. **During your experimental smoking phase, did you think that smoking was "cool"?**
   a. Yes
   b. No
   c. Don't know/not sure

22. **During your experimental smoking phase, did you use any of the following tobacco or nicotine products? Chewing tobacco or snuff, cigars, tobacco pipes, clove cigarettes or bidis, nicotine replacement products such as gum or patch, or any other tobacco products besides cigarettes.**
   a. Yes
   b. No

Regular smoking

After going through your experimental smoking phase, you then either stopped smoking or went on to become a **regular smoker**. Regular smokers have smoked daily or almost daily for 30 days straight.

23. **When did you first become a regular smoker?**
   a. I have never been a regular smoker
   b. Before age 10
   c. 10-13 years old
   d. 14-17 years old
   e. 18-21 years old
   f. 22 years or older

24. **Why did you become a regular smoker? (Check up to 3 reasons)**
   a. I have never been a regular smoker
   b. I craved cigarettes if I didn't smoke regularly
   c. I was around smokers a lot of the time
   d. I found smoking pleasurable
   e. I got a "buzz" when I smoked
   f. I enjoyed the taste
   g. Smoking relaxed me
   h. Smoking helped me focus and concentrate better
   i. Smoking made me feel less stressed
   j. I smoked to fit in with other people
   k. I liked the image of a smoker
   l. Smoking helped me control my weight
   m. Since others in my family smoked, it was easy to see myself as a smoker too
   n. Other: _____

25. **Why didn't you become a regular smoker? (Check up to 3 reasons)**
   a. I did become a regular smoker
   b. I didn't enjoy smoking
   c. Smoking cigarettes made me feel sick
   d. I didn't like the taste and/or the smell of cigarettes
   e. I didn't want to become addicted to nicotine
   f. I was concerned about the effects of smoking on my health
   g. I was concerned about the effects of my smoking on the health of the people around me
   h. I stopped hanging out with people who were smokers
   i. I was encouraged to stop smoking by my friends
   j. I was encouraged to stop smoking by an adult such as my parents/guardians, a teacher, or a doctor
   k. I didn't like the image of being a smoker
   l. It was hard for me to obtain cigarettes
   m. I couldn't afford to buy cigarettes
   n. I had trouble finding places I could smoke
   o. Since I grew up in a non-smoking family, I just couldn't see myself as a smoker
   p. Other: _____

Influences on smoking

Your parents and family

These questions focus on your family's smoking behavior and attitudes while you were growing up (from birth through age 18). Your parents/guardians are adults who had regular responsibility for raising you. Parents/guardians can include parents, step-parents, or other caregivers.

26. While you were growing up, how many of your parents/guardians smoked at all?

    a. None
    b. One
    c. More than one
    d. Don't know/not sure

27. While you were growing up, what was your oldest age when at least one of your parents/guardians smoked?

    a. None of my parents/guardians smoked when I was growing up
    b. 6 years old or less
    c. 7-10 years old
    d. 11-13 years old
    e. 14-18 years old
    f. Don't know/not sure

28. While you were growing up, did any of your brothers or sisters who lived with you smoke?

    a. Yes
    b. No
    c. I don't have any brothers or sisters who lived with me while I was growing up
    d. Don't know/not sure

29. While you were growing up, did any of your parents/guardians smoke inside your home?

    a. Yes
    b. No
    c. Don't know/not sure

30. While you were growing up, did your parents/guardians allow visitors and guests to smoke inside your home?

    a. Yes
    b. No
    c. Don't know/not sure

31. While you were growing up, did any of your parents/guardians allow you to smoke at home?

    a. Yes
    b. No
    c. Don't know/not sure

32. While you were growing up, did any of your parents/guardians ever talk to you about the dangers of smoking?

    a. Yes
    b. No
    c. Don't know/not sure

33. While you were growing up, did any of your parents/guardians tell you that you were not allowed to smoke?

    a. Yes
    b. No
    c. Don't know/not sure

34. While you were growing up, did you have a close relationship with any of your parents/guardians?

    a. Yes
    b. No
    c. Don't know/not sure

35. When you were growing up and your family went out to eat dinner in a restaurant, where did your parents/guardians usually ask to be seated?

    a. My family never went out to restaurants
    b. Non-smoking section
    c. Smoking section
    d. Non-smoking or smoking section (no preference)
    e. My family chose restaurants where smoking was not allowed
    f. Non-smoking sections in restaurants did not exist back then
    g. Don't know/not sure

36. When you were growing up, did your parents/guardians ever ask nearby smokers to stop smoking?

    a. Yes
    b. No
    c. Don't know/not sure

Your peers

These questions focus on your peers' smoking behavior and attitudes during your experimental smoking phase.

37. During your experimental smoking phase, was it common for people your age to smoke cigarettes?

    a. Not at all common
    b. Less than half the people my age smoked
    c. About half the people my age smoked
    d. Almost everyone my age smoked
    e. Don't know/not sure

38. **During your experimental smoking phase, did any of your closest friends smoke cigarettes?**
    a. None of them
    b. Some of them
    c. Most or all of them
    d. Don't know/not sure

39. **During your experimental smoking phase, did people your age think that smoking was "cool"?**
    a. None of them
    b. Some of them
    c. Most or all of them
    d. Don't know/not sure

40. **During your experimental smoking phase, did any of your friends think that smoking was "cool"?**
    a. None of them
    b. Some of them
    c. Most or all of them
    d. Don't know/not sure

School

These questions are about your school environment and what you learned about smoking in your classes (from kindergarten through 12th grade).

41. **When you were in school, were you taught about the dangers of smoking (for example, lung cancer, heart disease, fire hazards)?**
    a. Yes
    b. No
    c. I did not attend school
    d. Don't know/not sure

42. **When you were in school, did you discuss the reasons why people smoke in any of your classes?**
    a. Yes
    b. No
    c. I did not attend school
    d. Don't know/not sure

43. **When you were in school, were you taught about the side effects of smoking, such as making teeth yellow, causing wrinkles, or making smokers smell bad?**
    a. Yes
    b. No
    c. I did not attend school
    d. Don't know/not sure

9

44. **In what grades did you discuss smoking and health in your classes at school? (Check all that apply.)**
    a. Kindergarten to 6th grade
    b. 7th to 9th grade
    c. 10th to 12th grade
    d. I did not discuss this in school
    e. Don't know/not sure

45. **When you were in high school, were there designated smoking areas for students ?**
    a. I did not attend high school
    b. Yes
    c. No
    d. Don't know/not sure

46. **When you were in school, were adults such as teachers and administrators allowed to smoke at the school?**
    a. Yes
    b. No
    c. Don't know/not sure

Work

These questions are about smoking behaviors and attitudes at your place of work during your experimental smoking phase.

47. **During your experimental smoking phase, did you work at a job for at least 10 hours per week?**
    a. Yes
    b. No
    c. Don't know/not sure

48. **During your experimental smoking phase, were you serving on active duty in the military?**
    a. Yes
    b. No
    c. Don't know/not sure

49. **During your experimental smoking phase, was smoking allowed in any of the following areas in your workplace? (Check all that apply.)**
    a. In my work area
    b. In common areas (lobbies, rest rooms, lunch rooms)
    c. In designated smoking areas
    d. Don't know/not sure
    e. Smoking was not allowed anywhere in my workplace
    f. I was not working during my experimental smoking phase

10

50. During your experimental smoking phase, how many of your co-workers smoked?

   a. I was not working during my experimental smoking phase
   b. All or most of my co-workers smoked
   c. Some of my co-workers smoked
   d. None of my co-workers were smokers
   e. Don't know/not sure

Religion and ethnicity

These questions are about the smoking behaviors and attitudes of the religious and ethnic communities you belonged to when you were growing up.

51. When you were growing up, did the religious community you belonged to have a general attitude toward smoking tobacco?

   a. I did not belong to a religious community
   b. My religious community did not approve of smoking
   c. My religious community did not have a dominant attitude toward smoking
   d. Smoking was accepted by my religious community
   e. My religious community used tobacco in its ceremonies
   f. Don't know/not sure

52. When you were growing up, did the ethnic community you belonged to have a general attitude toward smoking tobacco?

   a. I did not belong to an ethnic community
   b. My ethnic community did not approve of smoking
   c. My ethnic community did not have a dominant attitude toward smoking
   d. Smoking was accepted by my ethnic community
   e. Don't know/not sure

Advertising and the media

These questions are about the influences of the media on your smoking behavior when you were growing up.

53. While you were growing up, how often did you see or hear anti-smoking messages on television, radio, billboards, posters, newspapers, magazines, or movies?

   a. Never
   b. Sometimes
   c. A lot
   d. Don't know/not sure

54. While you were growing up, how often did you see smoking advertisements on television, radio, billboards, posters, newspapers, and magazines?

   a. Never
   b. Sometimes
   c. A lot
   d. Don't know/not sure

11

55. While you were growing up, how often did you see actors smoking when you watched TV, movies, or videos?

   a. Never
   b. Sometimes
   c. A lot
   d. I didn't watch TV, movies, or videos
   e. Don't know/not sure

Public policy

These questions focus on how public policies can affect smoking behavior.

56. During your experimental smoking phase, were you influenced by the warning labels on cigarette packages?

   a. They influenced me to quit smoking
   b. They made me worry about my smoking
   c. They had no effect
   d. They amused me
   e. They angered me
   f. There were no warning labels on cigarette packages
   g. Don't know/not sure

57. During your experimental smoking phase, how did you get your cigarettes? (Check all that apply.)

   a. I bought them in a store
   b. I bought them from a vending machine
   c. I gave someone else money to buy them for me
   d. Someone else gave them to me
   e. I stole them
   f. I was given a free sample to try
   g. I bought them from a catalog
   h. I bought them over the Internet
   i. Other: _____

58. During your experimental smoking phase, was there a minimum legal age for buying cigarettes?

   a. Yes
   b. No
   c. Don't know/not sure

59. During your experimental smoking phase, was it easy for you to get cigarettes?

   a. Yes
   b. Somewhat
   c. No
   d. Don't know/not sure

12

60. During your experimental smoking phase, did anyone refuse to sell you cigarettes because of your age?
    a. No
    b. Sometimes
    c. Often
    d. I did not try to buy cigarettes
    e. Don't know/not sure

61. During your experimental smoking phase, could you usually afford cigarettes?
    a. Yes
    b. No
    c. I did not try to buy cigarettes
    d. Don't know/not sure

You and your life during your teen years

62. As a teen, do you think you had a healthy lifestyle?
    a. Yes
    b. Somewhat
    c. No
    d. Don't know/not sure

63. As a teen, how often did you eat fruits and vegetables?
    a. Less than 1 serving a day
    b. Usually 1 serving a day
    c. Usually 2-3 servings a day
    d. Usually 4-5 servings or more a day
    e. Don't know/not sure

64. As a teen, did you have a healthy diet? A healthy diet contains a wide variety of foods and is high in fruits and vegetables, lean protein, low-fat dairy products, and whole grains. It is low in sugar, fat, and salt.
    a. Not at all
    b. Some of the time
    c. Most or all of the time
    d. Don't know/not sure

65. As a teen, how often did you get moderate to vigorous exercise (for example, fast walking, running, swimming, skating, or playing a sport)?
    a. Less than once a week
    b. 1-2 times a week
    c. 3-5 times a week
    d. 6-7 times a week
    e. Don't know/not sure

13

66. As a teen, about how often did you participate in supervised activities after school such as organized sports, scouting, volunteer activities, etc.?
    a. 0 times a week
    b. 1-2 times a week
    c. 3 or more times a week
    d. Don't know/not sure

67. As a teen, how often did you drink alcohol?
    a. Often
    b. Sometimes
    c. Rarely
    d. Never
    e. Don't know/not sure

68. As a teen, how satisfied were you with your life? Consider all aspects of your life at this time.
    a. Not at all or slightly satisfied
    b. Somewhat satisfied
    c. Very satisfied
    d. Don't know/not sure

69. As a teen, how much stress did you experience in your life?
    a. None or very little
    b. Some
    c. A lot
    d. Don't know/not sure

70. As a teen, did you have high self-esteem?
    a. Not at all
    b. Somewhat
    c. Very
    d. Don't know/not sure

71. When you were a teen, did a doctor or dentist ever talk to you about the dangers of smoking?
    a. Yes
    b. No
    c. I didn't see a doctor or dentist
    d. Don't know/not sure

72. When you were a teen, do you think you received adequate information about smoking and its risks?
    a. Yes
    b. No
    c. Don't know/not sure

14

About you...

73. In what year were you born?

_____

74. What is your age in years?

_____

75. What is your sex?

a. Male
b. Female

76. Are you Hispanic or Latino?

a. Yes
b. No
c. Don't know/not sure

77. What is your race or ethnic group? (Check all that apply.)

a. American Indian or Alaska Native
b. Asian
c. Black or African American
d. Native Hawaiian or Other Pacific Islander
e. White
f. Other (specify) _____
g. Don't know/not sure

78. Which one of these groups would you say best represents your race? (Check only one.)

a. American Indian or Alaska Native
b. Asian
c. Black or African American
d. Native Hawaiian or Other Pacific Islander
e. White
f. Other (specify) _____
g. Don't know/not sure

79. When you were growing up, what language did you speak most often at home?

_____

80. What is the highest grade or year of school you have completed?

a. Did not graduate from high school
b. High school graduate or equivalent (GED)
c. Some college or technical/vocational school, but no degree or certificate
d. Junior college or technical/vocational school degree or certificate
e. Bachelor's degree or higher

15

---

81. What is the highest grade or year of school your parents/guardians completed? (Answer on a separate line for each parent/guardian.)

	Did not graduate from high school	High school graduate or GED	Some college, but no degree or certificate	Junior college or vocational school degree	Bachelor's degree or higher
Parent/guardian 1					
Parent/guardian 2					
Parent/guardian 3					
Parent/guardian 4					

82. When you were a teen, how well off was the family you lived with?

a. We struggled to make ends meet
b. We lived comfortably
c. We were very well off
d. Don't know/not sure

83. How well off is your current family/household?

a. We struggle to make ends meet
b. We live comfortably
c. We are very well off
d. Don't know/not sure

84. In what county do you currently live?

_____

85. When you were a teen, in what *one* state did you mostly live?

_____ (write in state)

a. I didn't spend those years mostly in one state
b. I spent those years mostly outside of the U.S.
c. Don't know/not sure

86. How would you describe the area where you currently live?

a. Urban
b. Suburban
c. Rural

You have completed all the questions. Is there anything else that you would like to tell us?

16

Your answers to these questions may help us learn more about the environmental and genetic factors that influence why some people become smokers and some people don't.

If you are filling this questionnaire out at home, please bring it with you to your appointment at the University of Washington.

Thank you!

Extracted Questionnaire and Genotyping Data

101. Is the subject a case, a control, or neither?
a. Case
b. Control
c. Neither

102. What is the FTND for this subject?
a. 0
b. 1
c. 2
d. 3
e. 4
f. 5
g. 6
h. 7
i. 8
j. 9
k. 10

103. Subject's genotype for *DRD2* C957T is:
a. C/C
b. T/T
c. C/T
d. not known

104. Subject's genotype for *DRD2* -141 Ins/Del is:
a. I/I
b. D/D
c. I/D
d. not known

105. Subject's genotype for *DDC* is:
a. G/G
b. A/A
c. G/A
d. not known

106. Subject's combined genotype is:

Introduction
In Exploring Databases curriculum, students test their specific hypotheses with an online database (Lesson 6) in a case control study format (Lesson 5) after they develop their overarching hypothesis from smoker profiles (Lesson 2), their intuition/observations, and past research. Thereafter, they explore the database further and develop new hypotheses to write their research proposal (Lesson 7). This appendix provides you (teachers) with a tutorial for how to start and open the database for your class.

1.1. Where is the database?
Go to https://gsoutreach.gs.washington.edu/database2/

1.2. System Requirement
The database application is optimized for *Firefox* (http://www.mozilla.com/en-US/firefox/) or *Chrome* (http://google.com/chrome) and we strongly recommend you to use one of them if available on your classroom computers. It will also work with *Internet Explorer* (default on Windows OS) and *Safari* (default on Mac OS), but with some limitations (see Table 1 below).

We have also learned that some teachers experience INTERNET or NETWORK issues due to student access restrictions, and we strongly recommend that you CHECK this in advance (see 1.3 for details).

Table 1. Web Browser Support

Browser, version, OS	Known Issues	Recommendation Level
Firefox on OSX and Windows	None	Strongly Recommended
Chrome on OSX and Windows	None	Recommended
Safari 4.0+ on OSX	None	
IE 9 on Windows	None	

Other browsers are not yet recommended

1.3. Network Issues
If your school and district require students to log into classroom computers with their accounts, they may not be able to access the online database or it may not work properly because of their access restrictions. NOTE that this could happen even if it works properly with your TEACHER account/computer. To identify/solve the issue in advance, please check the following with a STUDENT account.

1. On the database page, click one of the items in the questionnaire table and check whether you can see a table and bar graph under the questionnaire table. If so, the database is working properly.
2. Check another two or three computers in the same way for confirmation.

   If you identify any issues using the above list, your school/district may restrict your student access to the database page. Please contact your network administrator and ask them to open your student access to https://gsoutreach.gs.washington.edu/.

2.   Contact

If you have any question with the online database, please contact Maureen Munn mmunn@u.washington.edu.