Writing Your YES Research Paper

Tips for Success

1 Get Started

Choose Your Topic

Pick a topic that's important and meaningful to you. You might consider a health issue that affects people around you. After all, no one knows your community better than you do! Or you might consider one that has broader impact such as a health problem affecting populations globally.

- Make sure your topic is health-related and has a health outcome. Ask yourself, does it address something about human health, disease, or a risk factor for a health condition? (For example, academic performance is *not* a health topic.)

- If your topic has already been studied, make sure you look at it differently. Apply new methods, focus on a different population, or think of another possible risk factor. Or choose an entirely new topic.

Is it Epi?

Make sure your study falls into the field of epidemiology and not a related field. Organic chemistry, molecular biology and microbiology, for example, might be related to a public health issue, but a research project involving this needs to apply epidemiological design and methods.

- Think twice before considering a lab-based study. Even though some public health studies incorporate lab research methods, studies in which you work mostly in a lab and rely on results from bacteria, worms, mice or cell lines are out of the scope of this competition.

- Make sure your study has a population component. It must involve an issue that affects groups of people and draws upon that population for its data.

Quick Tips

✓ **Epi**
A search for genetic mutations that could explain a hereditary condition in a small population using blood samples from that population.

✗ **Not Epi**
A similar study on how specific mutations affect the gene function. Even though the gene function may be important to a large population, the study's focus is on the gene, not the population.

Imagine yourself with that award-winning YES paper. You can do it!

Writing Your YES Research Paper

Tips for Success

2 Develop a Research Question

The research question helps define what your study will answer. It asks whether there is some relationship between the focus of your study and the outcome you are interested in. Think of it as your study's objective or hypothesis stated as a question.

The rest of your study builds on the research question. Developing a good one is the hardest step of any study, so don't be discouraged if it takes several tries to get it just right. The pros do the same thing!

Quick Tips

- ✓ State your question clearly in your paper. It's perfectly acceptable to write, "My research question is…"
- ✓ Keep it simple. Try to avoid multi-part questions that could lead you to do multiple studies at once.
- ✓ Avoid questions that could become unmanageable. For example, it is hard to describe the many factors that contribute to obesity nationwide from a single study or dataset.
- ✓ Read previous winners' titles and abstracts for examples of good questions.

3 Find a Mentor

Though not required for the YES Competition, a mentor can be a great resource in evaluating your study design before you begin your project, reviewing your paper, and being your go-to person for questions and advice. Your mentor can be anyone, although someone with experience in your topic or research methods is best. Remember, however, that the actual work of the project, including the study design and analysis of the data, must be your own work.

Suggested places to find a mentor:

- Your school
- State and local health departments
- Your local college or university
- Journal articles

Many students seek out a teacher in their school, but you can also try a local college or university, even if they do not offer an epidemiology course:

- Search department faculty biographies for someone with experience researching your topic. (A statistics instructor who has worked on health surveys is a good choice and can also help you evaluate alternative ways to analyze your data. Nursing instructors make great mentors, too.)
- Try emailing them using the school's directory.

Try looking up the author of a paper on your topic. Remember, most people love to talk about their work. An author might suggest interesting directions for you to take the research, or even help you find data that has already been collected. Articles often include an email address, so don't be shy!

Another great resource is your local or state health department. Many have epidemiologists at the city or county level. Health departments also collect data, report it locally and send it for national reporting by federal agencies. They can help connect you to a variety of data sources for your project.

4 Choose a Design

Choosing a design for your study is a critical decision. Be sure to choose one that's appropriate for your research question and doable for the competition. Ask yourself:

- What type of design will I use?
- What type of data could best answer my research question?
- Can my data be obtained elsewhere or does it need to be collected especially for my study?

Quick Tip Visit the YES website for examples of study designs used by past competitors.

Two Common Study Types

Cross-sectional Studies

A practical choice for many YES projects, this design works well with survey studies. Think of it as a snapshot in time to look at different factors that might be associated with a health outcome. This type of design has a few drawbacks:

- It is hard to demonstrate the *cause* of an outcome or to tell what came first, since you are just looking at a snapshot of a condition in a population at one single point in time.
- The trends you see in the group may not explain outcomes seen in certain individuals. (For example, if you study the effect of industrial emissions on a neighboring community, those most bothered by air-borne contaminants or pollutants may have already moved away or avoided the area in the first place.)

Case-control Studies

Another common choice, this design is often used to compare the people who got sick (the cases) to those who did not (the controls). Be sure you have enough people enrolled in this type of study to be able to make conclusions about what you find. Keep in mind:

- Cases may be different from controls in a way that might also contribute to the outcome you are studying.
- Since cases and controls are matched based on certain characteristics such as age or residency, you can't look to see if those factors affected the outcome.

There are many other study types, so don't limit yourself to just these two possibilities!

5 Obtain Your Data

Once you've chosen a research question and study design, your next question should be:
What data do I need to gather to answer my research question?

Consider how large your dataset needs to be and who it should include. Your study's "sample" describes all of the individuals who are part of your dataset. When anyone talks about a sample, they often discuss how representative the sample is.

How large should your sample be? The answer is often how large the sample could realistically be. Ideally the sample includes everyone affected by the problem or risk factor you are studying.

Larger Datasets

- Broad topics that affect a lot of people, like obesity and the flu, are often best described from larger, national datasets, unless a component of the issue is important locally.
- Sometimes a dataset from a government source already exists that fits the bill.
- Broad topics can still be done with smaller datasets, provided you describe them as preliminary data.

Smaller Datasets

- Local area topics, like illness due to a shared local issue, can be described with smaller datasets.
- For practical reasons you may choose to look at a smaller group that represents the larger group of everyone affected.
- Ensure your sample has a sufficient number of exposures and outcomes you wish to study in order to have meaningful data. If you are studying something rare, it is often necessary to draw from a larger sample.
- A limited sample size will limit the conclusions you can make about the larger population. This happens a lot in epidemiology. Articles will often say "more research should be done…" to describe what could be done with a larger sample.

Collecting Your Own Data

Next you should decide whether to collect your own data or use an existing dataset. Some YES competitors choose to collect their own data, for example from their classmates or community. This approach works best for topics that are specific to your school, age group or community. It also works when you want to study something that is not already in a dataset elsewhere.

However, realize that writing survey questions can be tricky:

- Be careful not to lead your respondents to answer a certain way while still making sure they undestand your question.
- Your response categories and the order in which they are presented should not lead respondents to answer a certain way.
- A good place to start is previously published surveys. You can borrow questions or adapt them to what you are doing. Be sure you properly reference these sources.

Randomization has two different roles in epidemiology:

1) The first role is for selecting your study group. If the total number of people in the population you are interested in is large, you probably cannot collect or handle data from all of them. The usual strategy is to work with a sample of the population that is representative of the total population. A representative sample is one that reflects the proportions of various characteristics in the total that you decide are important – such as gender, age group, race or other characteristics. Your sample must mirror the overall population if you want to be able to generalize your findings to the total population. It is not difficult to choose a random sample. You can find helpful articles and tools on the internet and your school's mathematics or statistics teachers should be excellent resources in this area. One online example is found at **http://www.randomizer.org/tutorial.htm.**

Be aware that randomized samples are not the same as "systematic" samples. For example, if you selected every student in your school who has a birthday in February, June or October, this would be a systematic sample and not a randomized sample. In planning your study design, you should consider how your methods of selecting your study population will affect your results due to bias and other factors.

2) The second role is a randomized experimental design. This means you are going to perform an intervention, comparing the group of subjects exposed to the intervention to the group who was not exposed. In a randomized experimental design, the intervention each subject gets is based on a random selection process. The purpose is to obtain roughly equivalent distributions of various characterics in the two groups, the intervention group and the control group, and thus avoid bias. Randomization does not assure that this will happen every time, but it makes it more likely! Thus, you still need to check to see if there are any important differences in the two groups that might have affected your results.

Quick Tips: Random number generators

- ✓ You can find random number generators on the internet for selecting samples from a larger population. First you create a numbered list of all the people in the population – say 1 to 500. You then input the total population size and the sample size (let's say it is 50) that you want to research into the random number generator. It will give you back a list of 50 numbers selected randomly from the 500 possible subjects. The 50 represent your random sample.

- ✓ You can also find random number generators for research that involves assigning different interventions to different groups.

Using an Existing Dataset

An existing dataset may be the right choice when you want to study a larger population. It can save a lot of leg work for you, though it may not include everything you need. Gaining access to existing datasets can also take time. And you need to be sure that this dataset has not already been used to address the same question you are researching especially if this has been reported, presented and/or published.

Quick Tip

There are many national- and state-wide datasets available through the resources listed on the YES website and your state or local health department.

Institutional Review Board (IRB) Approval

Medical records and insurance claim databases are good potential secondary data sources and are well suited to cross-sectional studies. However, both require you to gain approval from a human rights committee or Institutional Review Board (IRB) to look at confidential medical information on individuals. This type of data should also be de-identified before you are given access, meaning that you cannot connect the data in the dataset with any particular individual.

It is important to protect the identity and personal data of those in your study. Any time you are collecting data about other people, you should obtain IRB approval to be sure your study conforms to the highest ethical standards. Even though it is not required for the YES Competition, most epidemiological studies are subject to human subjects' protection through an IRB. Keep in mind:

- If you gather data from an existing source, IRB approval of your study may be required prior to gaining access to the data.
- Allow plenty of time to complete your IRB application and receive approval prior to beginning your work. It may take a month or more to receive a response to your application.

In many cases, if you are collecting anonymous survey data in your school, getting your school's administrative approval and asking for student and/or parental consent to participate may be considered adequate. Again, always be sure to protect the confidentiality of the information you receive from participants.

6 Analyze Your Data

This part of the project can be nerve-wracking. It's a great time to ask for guidance from your mentor or the statistics teacher at your school. The best type of data analysis will vary depending on the type of study you did. This guide will focus on survey data, whether collected by you or from another source. You can get through this part of your study one step at a time.

First, check the response frequency for each variable in your dataset.

- Keep your eyes out for missing responses, unexpected responses, and so on.
- This step will help you see if you did not get enough information to analyze a particular aspect of your project.
- It may also help you identify some unexpected results.

Next, consider your hypothesis and the data you obtained.

- What is the best way to tell the story you need to tell to support or refute your hypothesis? This varies depending on your research question. In a few cases, frequencies of responses might be enough.
- Often though, you need to go a step further and begin using statistical tests.

Statistical Tests

Consider these three statistical tests commonly used in epidemiology:

1. **Correlation tests** measure just what they sound like. They test to see how much of the variance between the two variables you are testing can be explained by the relationship.
2. **Chi-squares** are often called two by two comparisons, since you can think of them as comparing two groups against two conditions.
3. **T-tests** compare the average values in two groups.

There are many other common tests that could be more appropriate for your study. In short, there are a lot of tests you *could* use. Figuring out which ones you *should* use is the tricky part. Luckily there are plenty of resources available. Some might already be in your school.

7 Present Your Data

Epidemiological papers follow a standard format: title page, abstract, acknowledgements, introduction, methods, results, discussion, references. Here are a few quick tips for the first parts of your paper:

- Write your abstract after completing everything else. It will come together more quickly, and you won't have to worry about it changing.
- The introduction should clearly state what you plan to do and why. If you plan to collect information from a small or convenient population, justify it.
- Where available, include findings from similar journal articles on the topic and cite them properly. If you have trouble finding similar published work, offer some reasons why.
- Acknowledgements: For the YES Competition, acknowledgements will be included as part of your registration rather than in the actual research paper.

Results

In your results section, summarize the data you collected. Don't just present a table listing all the variables for every respondent. Instead, summarize the key variables using a mixture of charts, graphs and text.

- Charts or tables are great for presenting numbers or proportions.
- Graphs or figures are well-suited to showing a trend or comparing groups. Include text that describes the findings and highlights the key points of your tables and figures.

Description of the sample

Classically, epidemiological papers have a table that describes the sample population and includes things like age, race, gender, etc., though this is only necessary for studies that involve data from participants. A description of your sample is also a good way to start your results section.

You should mention how your study population compares to the larger population of those affected by your topic. For example, if you were studying patients in care for a certain condition at a local clinic, you should compare your respondents to what you know about people with that condition in your community, state or the country. You might look to see if the respondents were more likely to be female or any other characteristic that might be important to your target population.

Quick Tip

Share your interpretation of the results in the discussion section, not here.

Discussion

You're in the home stretch now. In the discussion section, describe what story your data tell based on the results you just presented. Be careful not to overstate them.

If you used data from a small number of respondents, you cannot make broad statements about the larger population. If you are lucky enough that your results support your hypothesis, remember that your findings are specific to your study and are not enough to "prove" your hypothesis.

What should you do if your findings don't support or confirm your hypothesis?

If your findings do not confirm your hypothesis or do not reach statistical significance, that's okay. You still learned something that is important to communicate. You should review your study design and methods and see if there have been places where bias crept in and skewed your results. Or you could redesign the study to make it better and possibly reach a different conclusion. The judges will be impressed if you can describe what you have learned from the experience of having a negative finding (that is, you didn't get the result you expected) and that you have also learned ways to improve your design. The caveat is that you should not just say, "I would redo the study with a larger sample size."

Use caution with these words

Some terms in epidemiology have a special meaning. Be careful when using:

- ***Proof:*** Your study's findings are not enough to prove your hypothesis.
- ***Significant:*** Refers to the outcomes of a statistical significance test, which measures the p-value of different calculations. Don't use significant unless you have done a statistical test that resulted in a significant p-value.
- ***Associations and correlations:*** These are shown through specific statistical correlation tests.
- ***Cause:*** We often want to show that something was the cause of the outcome we are seeing, but this is not possible when we ask about the outcome and possible associations at the same time, as in a cross-sectional study.

Epidemiological paper format checklist:

- ☑ Title Page
- ☑ Abstract
- ☑ Acknowledgements
- ☑ Introduction
- ☑ Methods
- ☑ Results
- ☑ Discussion
- ☑ References

Find out more about the format on the YES website under "Content & Organization."

Academic Integrity

Your Research Report must be your own work; coauthored reports are not eligible. Ideas, techniques, facts, words, images, someone else's validated instrument, or information from other sources (including the internet) must be properly cited. The YES Competition uses specific procedures as part of the judging process to detect plagiarized materials. See the competition guidelines for more details.