Disease Detectives Exam Tips

A well-written Disease Detectives event is very similar to a **case study**. Finding good scenarios (i.e., articles or reports that serve as the basis for the case study) is one of the most challenging parts of developing a good Disease Detectives event. A good scenario should tell a story, include sufficient detail and use relatively simple methods to answer its research questions. Research and publishing trends are making it increasingly difficult to find articles to use as a basis for events — especially those where one is hoping to base fifteen or twenty good questions on a single scenario. Reports of outbreak investigations are a great place to start but run-of-the-mill outbreaks seldom appear in the newer literature. If one has access to a university library, the American Journal of Hygiene (now American Journal of Epidemiology) is a great place to look.

CDC's Morbidity and Mortality Weekly Report (MMWR) is a good place to start (https://www.cdc.gov/mmwr/publications/index.html#). The archives (listed under the MMWR resources) go back to the very first issue. The earlier reports tend to provide more detail than recent ones and are generally good for five or six questions. Stringing several of these together can result in a really great regional or state event for either Division. MMWR reports on major health events are often published before the investigation has been completed. They tend to be brief and incomplete. However, they are often followed detailed reports published in refereed journals. These can be used as the basis for larger events with more questions. MMWR articles can be used as a source for search terms that can be used in MEDLINE or PubMed to find more complete articles. The fungal infection outbreak that served as the basis for part of the 201^ National Tournament was picked up this way.

Official reports can also serve as the basis for event scenarios. The Ebola outbreak that served as the basis for the one of the scenarios in the 2014 National Tournament was based on a WHO report that was referenced on a CDC webpage listing Ebola outbreaks (https://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html).   The waterpark outbreak that was used in the 2017 national tournament was downloaded from a collection of NIOSH Health Hazard Evaluations (https://www.cdc.gov/niosh/hhe/). In looking for this report, I found a similar list of NCEH investigations at https://www.cdc.gov/nceh/ehs/activities/env-investigations.html .

While most of the outbreaks used in national competitions really happened, there have been a few that were fictitious. The child care crypto outbreak used as a scenario in the 2015 National Tournament is one such example. For this scenario, we created a line list with both demographic and illness information. The number of children and care-givers in each room was based on national licensing standards. Attack rates, incubation periods and clinical symptoms were based on published reports. This line list was used to produce figures and tables and allowed us to be sure that the numbers were consistent as the outbreak progressed.

Examples of types of performance indicators that could be used in Division B and C events at various levels of competition.

Competition Level	Division	
	Division B (Middle School)	Division C (High School)
Regional/State	List and recognize examples of different modes of transmission Calculate health-related rates (attack, incidence, prevalence, case fatality) Calculate a simple relative risk and describe what it means Interpret epi curves, temporal patterns and other simple graphic presentations of health data. List, discuss and recognize examples of disease causing agents (physical and biological) Demonstrate an understanding and ability to use terms such as endemic, epidemic and pandemic; population versus sample, association versus cause. Describe various types of prevention and control strategies (e.g. immunization, behavior change, etc) and situations where they might be used	Recognize differences between study designs and **types of error**. Calculate measures of risk (e.g. relative risk or odds ratio) when given a description of the study design Calculate measures based on data that is not given but that can be readily extracted. Recognize how gaps in information influence the ability to extend conclusions to the general population. Do **Statistical Analysis** on Data for both Descriptive and Analytical Epidemiology
National	Understand how units affect the relative magnitude of a set of rates with different units. Calculate appropriate measures of risk when given the study design Complete tables when given all data needed to complete calculations. Propose a reasonable intervention to a public health problem. Recognize gaps in information	Recognize unmentioned factors that may influence results and **types of error**. Convert between rates with different basic units (e.g. incidence per 10000 persons/year to incidence per 100 persons/week). Propose a means to evaluate the effectiveness of an intervention or control program. Do **Statistical Analysis** on Data for both Descriptive and Analytical Epidemiology

Potential Types of Error in Data Collection
- False Relationships
  - **Random Error** - the divergence due to chance alone, of an observation on sample from the true population value, leading to lack of precision in measurement of association
  - **Bias** or **systematic error** in an epidemiologic study that results in an incorrect estimation of the association between exposure and health-related event

- Non-Causal Relationships – "guilt by association"
  - **Confounding** – occurs when the effects of two risk factors are mixed in the occurrence of the health-related event under study - when an extraneous factor is related to both disease and exposure

Random Error:
- Is the result of fluctuations around a true value due to sampling variability
- It can occur with data collection, coding, transfer, or analysis of data
- Affects measurement in a inconsistent manner
- Ways to reduce random error may include – increasing the sample size and reduce the variability in measurements

Systematic Error:
- Occurs when there is a difference between the true value (the population) and the observed value (the sample)
- The error is in the system used for measurement so it occurs in each occasion hence systematic error
- Conclusions drawn on this data will be inaccurate – too great or too little
- Validity of a study depends upon the degree of systematic error – less error equals more validity
  - Internal validity – the amount of error in measurements including those for exposure, disease, and the association between these variables.
  - External validity –relates to the process of generalizing the finding of a study to the population from which the study is taken

Types of Bias: Describe problems in how the study is organized
- **Selection bias –** occurs when study subjects are selected for the study as a result of a third unmeasured variable which is association with both the exposure and the outcome. There may be association between diseases or between characteristics and a disease related to the admission to a hospital for those with a disease, without the disease but with symptoms, and those with only the characteristics of the disease.
- **Information bias –** occurs from systematic error in the assessment of a variable. Examples **are information bias, response bias, interviewer bias, recall bias**

Confounding: the co-occurrence or mixing of effects of extraneous factors
- May lead to overestimating or underestimating the true association between exposure and outcome.
- A **confounding variable** would be a variable (pollution) that can cause the disease under study (cancer) and is also associated with the exposure (smoking).

Statistical Concepts for Disease Detectives- Division C

Some **statics analysis** may be used at Division C – but it must be less than 10% of the competition.
Descriptive and analytic statistical methods are important in the life and health sciences and it is reasonable to expect competitors at the high school level to demonstrate a basic competency in these methods in events such as Disease Detectives. However, statistics is not a major component of the event and should not define the event. **Many competitors will not have had statistics and event supervisors should use it like a habanero, sparingly if at all but with purpose. One or two short, well-placed questions are adequate.**

Key principles

- Competitors should be able to use statistics – not do statistics. Competitors should not be expected to do lengthy calculations such as paired T-tests, life tables or analysis of variance using hand calculators. In real life, these are done with computers and statistical software packages such as SAS, SPSS or EpiInfo.
- A more realistic approach would be to explain the study design and data collection methods and have competitors interpret printouts from one of the above programs, to use intermediate statistics such as standard deviations and population means to calculate test statistics or tell how changes in the value of intermediate statistics would affect final results.
- Students should be provided formulas for all calculations and tables. It is reasonable to expect them to know which test to use, assumptions and limitations of each but not to memorize the formulas.
- Interpretation and translation of statistics into plain language is a valid competency that fits easily into a multiple choice format.   For example, "Which of the below statements best describes these results?"
- Asking students to select a test or what test they would use to test a particular hypothesis is also appropriate. This question may be best suited for multiple choice unless graders have sufficient expertise to know which tests are clearly inappropriate and which are marginally appropriate and recognize the different names for identical tests.
- Questions concerning statistics should be part of a larger scenario so that competitors can see how they fit into the larger analysis.
- Free-standing questions on statistics just for the sake of including statistics into the competition are contrary to the intent of the event.   Event supervisors should not feel compelled to include questions about statistics in the events if they do not fit into the larger picture.