Applied Data Analysis for the Military Decision Maker (Part II)

Professor Ron Fricker
Naval Postgraduate School
Monterey, California
Goals for this Lecture

• The hows and whys of data collection
  – Why you should care about the data first and foremost
  – Observational vs. experimental data and administrative datasets: strengths and weaknesses
  – Sources of error in data

• Principles of good data analysis
  – How should an analysis be approached and undertaken?
  – What makes an analysis "good"?
  – What should I ask during a briefing?
Good Analyses Start with Good Data

• Bad data in statistical analysis is just another form of *garbage in, garbage out*

• Determining whether you have good data is not always easy – it takes time and effort

• Real-world, large data sets are always messy
  – Missing data, data (entry) errors, duplicated records, etc.

• Good analysis is almost always:
  – 90 percent data collection, cleaning, and management
  – 10 percent analysis
Data Sources

- **Existing** (often called *administrative*) data
  - E.g., personnel records, logistics or maintenance records
  - Pro: Data has already been collected
  - Cons:
    - Data was collected *for some other purpose*
    - May have errors and/or idiosyncrasies making it problematic or unsuitable for your analysis

- **New data** – i.e., you have to go out and collect it yourself
  - E.g., direct measurement, surveys, experiments
  - Pros:
    - Data is collected explicitly for your analysis
    - Well-designed and executed data collection effort makes for better (and often easier) analysis
  - Con: Collecting data is usually quite expensive and time consuming
Bias and Variance

- Poorly constructed data collection methods suffer from bias
  - **Bias**: There is something systematically wrong with the study
  - If you did many studies and averaged the results would you get the right answer?
  - Solution: randomization

- Variance (variability) is a feature of all methods
  - If you did the study again you’d get a different answer
  - Relevant question: how precise are your estimates?
  - Solution: collect more data
Low bias
• Low variance

High bias
• High variance

Other sample statistics you could have observed

Observed sample statistic (sample mean)

True population value (pop. mean)
(Some) Sources of Bias

- **Sampling bias**: some people/units with a particular trait more likely to get into sample than others
- **Selection bias**: people/units with a particular trait (e.g., strong opinions) are favored
- **Sensitivity bias**: answers to questions of salary/sex/other social taboos might not be truthful
- **Non-response bias**: people that do not answer your survey are different from those that do
- **Reporting bias**:
  - Everything gives you cancer
  - Nobody reports the results from studies that show no link between “X” and cancer
Case #4: US Presidential Politics

• Literary Digest Poll, FDR vs Alfred Landon
  – Sampling frame taken from telephone numbers and club membership lists
    • Only one house in 4 had a phone
    • Rich tend to join clubs more than poor
    • Sampling (frame coverage) bias
  – Sent questionnaires to 10 million people
    • Only 2.4 million people responded
    • Nonresponse bias
Results of the Poll

• LD Prediction: FDR 43%, Landon 57%
• Election Result: FDR 62%, Landon 38%
• Largest error ever made by a major poll
  – Literary Digest had correctly forecast every election since 1916
  – Literary Digest went bankrupt soon after 1936
• The Gallup organization (a start-up in 1936)
  – Guessed the election more or less correctly
  – Guessed LD’s result using much less data
Even a survey chosen on purpose to be representative can suffer from bias

- 1948 Gallup Poll (others did similar things)
- Quota sampling: Survey crews interview people matching particular characteristics
- Interview 13 people:
  - 7 men, 6 women
  - Of the men, 3 under 40, 4 over 40
  - etc...
- What’s wrong with quota sampling?
Who Would *You* Interview?

Republican

Democrat

✓ Survey interviewers showed republican bias
Two Categories of Evidence

**Observational**

Simply observe A and B

A → B

or

B → A

or

A

C

B

Direction and source of cause not determined

**Experimental**

Modify one condition and compare results

A → B

Change A

See what happens to B

A → B

Direction of effect (cause) determined
Controlled Experiments

• What is the effect of a treatment?
  – E.g: Does the Salk vaccine reduce incidence of Polio?

• Assign a collection of experimental units (children) into two groups
  – Treatment group gets the Salk vaccine
  – Control group gets placebo

• Measure incidence of Polio each group

• Vaccine effect is (average reduction for treated)-(average reduction for control)
Bias in Experiments

• Confounding
  – Vaccine is given to all the girls and placebo to all the boys
  – Maybe girls have a lower incidence of Polio due to physiological differences?
  – Is the difference between treatment and control due to vaccine or gender?

• Lurking variables
  – Unobserved variables with which your treatment and control assignments may be confounded
Does smoking cause cancer?

Smoking

Smoker genes or behavior

Cancer

Cancer
Randomization

• Big problem with lurking variables
  – We can’t see them because they’re “lurking”
• **Randomization**: Subjects (units) assigned to treatment or control groups at random
  – With high probability random assignment is not confounded with lurking variables
  – Minimizes chances of bias
Blinding

- **Blinding**: Subjects do not know if they are in the treatment or control group
- **Double blinding**: Doctors (for medical trials) do not know if their patients are in the treatment or control group
  - Controls for subtle biases that might occur
  - Subconscious bias can lead to lurking variables
  - Example: doctor prescribes the drug s/he thinks works best to the sickest patients
Trying to Determine Cause

- Controlled experiments try to determine cause and effect
- Randomization and blinding minimize chance of systematic differences between groups other than treatment
- Must think hard about right way to randomize and possible confounding
- Desired conclusion: Treatment causes the observed effect
Considerations in Experiments

• When experimenting on people (medical trials), must carefully consider ethics
  – Is treatment likely to be beneficial?
  – If data seems to show treatment benefit, when to stop giving placebo?

• Experiments apply to other than medical
  – Invented to improve agricultural practices
  – Often used for product/production improvement
Observational Studies

• In observational studies, not randomized into treatment and control by investigators
  – Study designer either
    • Can’t randomize units to treatment/control
    • Chooses not to randomize units
  – Investigators simply observe what happens

• Examples
  – Smoking studies (don’t assign people to smoke)
  – Consumer behavior studies
Observational Study Weaknesses

• Almost always subject to criticism due to possible lurking variables
  – Omitted variable bias:
    • A lurking variable can change your conclusion
    • Smoking: Is cancer caused by genetics or by smoking?
  – Simpson’s paradox:
    • Conditioning the analysis on a lurking variable can change the conclusion
Association vs. Causation

- If “a” causes “b” then “a” will be associated with “b”
  - Association: Existence or change in “a” related to the existence or change in “b”
- However, “a” associated with “b” does not mean that “a” causes “b”
  - Directionality: perhaps “b” causes “a”
  - Lurking variables: perhaps “c” causes both “a” and “b”
Calvin & Hobbes by Bill Watterson

GRAPHIC VIOLENCE IN THE MEDIA.

DOES IT GLAMORIZE VIOLENCE? SURE. DOES IT DESENSITIZE US TO VIOLENCE? OF COURSE. DOES IT HELP US TOLERATE VIOLENCE? YOU BET. DOES IT STUNT OUR EMPATHY FOR OUR FELLOW BEINGS? HECK YES.

DOES IT CAUSE VIOLENCE? WELL, THAT'S HARD TO PROVE.

THE TRICK IS TO ASK THE RIGHT QUESTION.
Without Randomization…

…how to tell if smoking causes cancer?
Hill’s criteria for evidence to support causation:

- **Strength**: strong association is firmer evidence of causality
- **Consistency**: similar studies using other methods give similar results
- **Specificity**: cause leads to only one outcome and outcome can only come from the one cause
- **Temporarily**: exposure to causal factor must precede outcome
- **(Biologic) Gradient**: an increase in level, intensity, duration or total level of exposure to the potential cause leads to progressive increase in (the likelihood of) the outcome
- **Plausibility**: association is plausible with known scientific facts about the potential cause and the result
- **Coherence**: all available evidence coheres to form a cohesive whole
- **Experimentation**: experimental studies provide results in support of the causal hypothesis (**provides strong evidence**)
- **Analogy**: similarity between things that are otherwise different (**weak evidence**)

From *Epidemiology Kept Simple: An Introduction to Classic and Modern Epidemiology* by B. Burt Gerstman
Principles of Good Research and Good Data Analysis

- The **problem** should be well formulated
- The **research approach** should be well designed and well executed
- The **data and assumptions** should be sound
- The **research and analysis** should:
  - Be objective, independent, and balanced
  - Demonstrate understanding of previous related studies
- The **research findings** should be useful and advance knowledge
- The **recommendations** should follow logically from the findings

Excerpted from www.rand.org/standards/
More Good Resources


Case #5: Survey of Pesticide Use in (First) Gulf War

• Almost 700,000 US troops participated in the Persian Gulf War
• After the war, veterans began to complain about a variety of symptoms
• In response, DoD and VA set up registries to catalogue symptoms and offer veterans a comprehensive medical exam
• Over 100,000 veterans have registered—not all have symptoms
Six Most Prevalent Symptoms Reported

- Fatigue
- Skin rash
- Headaches
- Muscle, joint pain
- Memory loss
- Sleeping problems

% of those registering

0 5 10 15 20 25
Possible GWI Causes

**Chemical Warfare**
- Sarin
- Cyclosarin
- Thiosarin
- Mustard
- Phosgene oxime
- Lewisite
- Soman

**Immunizations**
- Anthrax
- Botulinum toxin
- Adjuvants
- Contaminants
- Combinations

**Pesticides**
- Permethrin
- Deet
- Lindane
- Thallium
- Organophosphates
- Combinations

**Stress**
- Pyridostigmine bromide
- Depleted Uranium

**Environmental Exposures**
- Heat & cold
- Oil-well fires
- Water
- Sand

**Occupational Exposures**
- CARC paint
- Equipment retrograde
- Diet
- Fuels & lubricants
- Tent heaters

**Infectious diseases**
- Leishmaniasis
- Mycoplasma
- Strep
- Retroviruses
- Foodborne diseases

**Biological Warfare**
- Aflatoxin (mycotoxin)
- Anthrax
- Botulinum toxin

**Miscellaneous**
- Multiple chemical sensitivity
- Porphiria
- Fibromyalgia
- Chronic fatigue syndrome
- Baumsweiger syndrome
- Dental amalgams
Difficult to Distinguish the Role of the Gulf War from Genetics and Other Exposures

Genetics → Possible exposure before the war → Gulf War → Possible exposure after the war

Healthy → Sick
Pesticides can cause symptoms similar to those experienced by some veterans
  – Considered a potential contributing cause of illness in Gulf War veterans

Information about how average service members used pesticides not otherwise available

Survey not designed to link pesticide use to health outcomes
Key Research Questions

- Which pesticides were used?
- Who used them?
- How often and in what quantities?
Surveyed Veterans Representative of Entire Gulf War Population

- Telephone survey of Gulf War veterans
  - Fielded in late 1999
  - Statistically representative
    - U.S. personnel
    - on the ground
    - between August 1990 & July 1991
- Sample designed for +/- 2% margin of error
Issues in Conducting the Survey

• Choosing who to survey
• Maximizing recall of pesticide use
• Identifying pesticides and their active ingredients
Choosing Whom to Survey: Personnel on the Ground in KTO

- Sampling frame created from US Army Center for Unit Records Research (USACURR) data:
  - 535,942 personnel identified on the ground in KTO
    - Not afloat
    - Between August 1, 1990 and July 31, 1991
Stratified Random Sampling Scheme Used to Select Respondents

• Service
  – Army
  – Air Force
  – Marine Corps &
  – Navy Ashore/Coast Guard

• Urban Areas
• Senior Enlisted (E-6 To E-9)
• Food Service MOSs

Oversampled
## Sample Sizes

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Initial Sample Selected</th>
<th>Expected Number In Final Sample</th>
<th>Actual Number in Final Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ARMY</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban area</td>
<td>218</td>
<td>133</td>
<td>149</td>
</tr>
<tr>
<td>Food Service</td>
<td>109</td>
<td>67</td>
<td>54</td>
</tr>
<tr>
<td>Senior enlisted</td>
<td>272</td>
<td>167</td>
<td>168</td>
</tr>
<tr>
<td>All else</td>
<td>489</td>
<td>300</td>
<td>281</td>
</tr>
<tr>
<td><strong>Total Army</strong></td>
<td>1,088</td>
<td>667</td>
<td>652</td>
</tr>
<tr>
<td><strong>AIR FORCE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban area</td>
<td>218</td>
<td>133</td>
<td>160</td>
</tr>
<tr>
<td>Food Service</td>
<td>109</td>
<td>67</td>
<td>70</td>
</tr>
<tr>
<td>Senior enlisted</td>
<td>272</td>
<td>167</td>
<td>175</td>
</tr>
<tr>
<td>All else</td>
<td>489</td>
<td>300</td>
<td>317</td>
</tr>
<tr>
<td><strong>Total Air Force</strong></td>
<td>1,088</td>
<td>667</td>
<td>722</td>
</tr>
<tr>
<td><strong>MARINES/NAVY</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food Service</td>
<td>109</td>
<td>67</td>
<td>55</td>
</tr>
<tr>
<td>Senior enlisted</td>
<td>272</td>
<td>167</td>
<td>169</td>
</tr>
<tr>
<td>All else</td>
<td>707</td>
<td>432</td>
<td>405</td>
</tr>
<tr>
<td><strong>Total Marines/Navy</strong></td>
<td>1,083</td>
<td>662</td>
<td>629</td>
</tr>
<tr>
<td>Coast Guard</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>3,264</td>
<td>2,000</td>
<td>2,005</td>
</tr>
</tbody>
</table>
Types and Rates of Nonresponse

- Nonresponse
  - Refused 3%
  - Unable to respond 2%
- Database errors (wasn’t in GW) 7%
- Unable to locate respondent 23%
- Located, didn’t interview 3%
- Other 1%
Maximizing Pesticide Use Recall

- Pesticide use probably hard to remember
  - Minor detail compared to other events
- Issues:
  - How to aid recall?
  - How to link responses to locations and time periods?
  - How to most efficiently collect information?
We Employed Numerous Aids to Enhance Respondent Recall

• Each participant asked to review package of materials before interview:
  – Background brochure describing the survey
  – Map of the Persian Gulf Theater of Operations
  – Gulf War calendar and list of key events
  – Respondent asked to fill out fact sheet on their Gulf War service
Survey Structured to Enhance Veterans’ Recall

• Shaped to reflect way veterans recalled their experiences
• Structure proceeds from easy to remember events to harder to remember
• Interviewers provided key events to respondents to spark memories of particular time periods
Basic Survey Instrument Structure

Basic Resp. Identifying Information

Pick random month

Information on living, working, sleeping conditions

List of pests seen and pesticide forms used

personal use pesticides

Form 1

Name?

Y

How used? Frequency?

N

Use? Color? Smell?

How used? Frequency?

(repeat for field use pesticides)

Exit questions
Estimated Use of Some Active Ingredients

• Could map some form, color, smell combinations to unique active ingredients

• Estimated probability of use for some active ingredients in personal pesticides

• For all others, results limited to form of pesticide
  – Some forms unique:
    • No-pest strips
    • “Fly bait”
Results for Personal Pesticides
Used on Body and Uniform

• In general, Army used pesticides most, followed by Marine Corps, Navy, and Air Force

• Survey data show that:
  – Insect repellents used most frequently
    • 50% used DEET (median use: 30 times/month)
  – Other personal pesticides used less extensively
    • E.g., 6% used Permethrin (median use: 20 times per month)
  – 1/3 did not use any pesticides

• Usage varied less by other factors
“Misuse” defined as violating EPA-approved label instructions in existence at that time

Pet flea/tick collars most visible example of misuse
  - Approximately 13,000 US service members used pet flea/tick collars at any particular time during the Gulf War
  - Most wore collars over clothes or shoes

Other intentional or unintentional misuse rare
  - Caveat: Can not determine if lack of misuse is because events were truly rare or were hard to remember
  - No evidence found suggesting widespread misuse
Multiple Exposure

• Some respondents reported multiple pesticide use
  – 31% used more than one type of pesticide at a time, but only 9% used three or more types
  – More use of pesticides on the person positively correlated with higher use of pesticides in other ways

• Also found modest positive association between pesticides and pyridostigmine bromide (―PB‖) pills
  – PB pills used to guard against attack with nerve agent soman
  – Correlation may be the result of season, location, or predisposition to use pesticides

• Result: some service members experienced more pesticides or greater concentrations than others
Take-Aways

• Good surveys are more than a bunch of questions posted on the web or sent out to everyone in the population
• Data collection can be both expensive and time consuming
• Association is not the same as causation
Some Briefing Questions

• How did you get/collect your data?
  – If administrative data:
    • What is the original purpose of the data (collection system)?
    • In what ways does the data not fit/address the analytical objective(s)?
    • What sort of unusual things did you find in the data?
  – If new data:
    • What sort of compromises/trade-offs did you have to make when collecting your data?
    • Did you have to exclude any sub-populations? Why or why not?
    • Did you use an experimental design and/or randomization? Why or why not?
Some Briefing Questions (cont’d)

• What sort of biases may be present in your results as a result of the
  – data you used?
  – way you collected the data?

• Could your results be confounded in any way? Why or why not?

• If the conclusions are causal: What’s your evidence that “X” actually causes “Y”?
  – Particularly if it’s an observational study: Are there other explanations for the observed association?


Case #3: Fricker, R.D., Jr., *Some Thoughts on INSURV Data Analysis and Display (FOUO)*, briefing, July 2011.
