Lies, Damned Lies, 
And What To Do With Them 

Ken Rice, CHRU Work in Progress 

March 2, 2005
A question for English majors

“There are three types of lies;

lies, damned lies, and statistics”

Which famous philistine said this?

A) Benjamin Disraeli
B) Mark Twain
C) George Bernard Shaw
D) Oscar Wilde
A question for English majors

“There are three types of lies;

lies, damned lies, and statistics”

Which famous philistine said this?

A) Benjamin Disraeli
B) Mark Twain
C) George Bernard Shaw
D) Oscar Wilde

E) Any of them, depending on who you ask (a Bayesian answer)
Structure of the talk

- Lies (from you)
- Damned lies (from data)
- Statistics (from me)
Raise your hand... if appropriate

Q.
Raise your hand... if appropriate

Q. Are you having an extra-marital affair?
Raise your hand... if appropriate

Q. Are you having an extra-marital affair?

Q. Have you taken money from drug companies?
Raise your hand... if appropriate

Q. Are you having an extra-marital affair?

Q. Have you taken money from drug companies?

Q. Have you had an STD/driven drunk/taken drugs...

A. We are unlikely to get 100% truthful answers

Can’t always collect ‘perfect’ data – but using a statistical model can help. (Get a coin ready)
A fun game about lying

1. Toss 2 coins
   – keep results secret

2. Rude Q:
   Tell the truth!  
   Lie!

   Yes  
   No

3. Math: \( p/4 \)  \( (1-p)/4 \)  \( 3(1-p)/4 \)  \( 3p/4 \)

   Proportion of hands up is about \( 3/4 - p/2 \)
What did we learn from that?

This is a ‘randomized response’ technique (Warner 1965). We get a **misclassified** version of the truth – and by design we can’t uncover the ‘truth’ for each person.

- The model for **misclassification** is important;
  - The ‘truth’ actually means something
  - Your coins are all ‘fair’ - 50:50 heads/tails
  - You followed the flowchart
  - You weren’t influenced by anyone else

- Have to make **assumptions** like this to get much done (e.g. estimate $p$)

- If you know **nothing** about ‘fairness’ you learn **nothing** about $p$

- Typical ‘solutions’ are more assumption-driven than we’d usually like
Some very nice genotype data
Some damned lies!

Ignoring the problem, or the whole dataset, is A Bad Idea
Assumptions about the error process

These are basically the same as in the coin-tossing game;

- Everyone has a ‘true’ genotype
- Only your true genotype affects your observed genotype (if anything)
- Process is same for cases/controls, men/women etc
- So-called non-differential misclassification
- Realistic for genotyping error, less so elsewhere (drinkers/smokers)

What practical effect will the error have?
Bias towards the null

For a binary exposure, non-differential misclassification biases towards the null

Event rate by genotype

True data (unseen) Observed data

A B
A B

Should expect estimates to get ‘more extreme’ - intervals less intuitive
Bias in lots of directions

For any other situation intuition is elusive;

Claiming ‘bias towards the null’ here is unhelpful and often inaccurate
Unmatched case-control setup

Switch to a retrospective analysis;

<table>
<thead>
<tr>
<th>Disease Status</th>
<th>Observed Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA</td>
</tr>
<tr>
<td>Control</td>
<td>$p_1$</td>
</tr>
<tr>
<td>Case</td>
<td>$q_1$</td>
</tr>
</tbody>
</table>

- Perhaps want to **test** equality, where $p = q$

- Or **estimate** odds ratios; $\frac{q_2 p_1}{q_1 p_2}$ and $\frac{q_3 p_1}{q_1 p_3}$

Normally the two approaches would be extremely similar – consider them both under misclassification;
Testing equality ($p = q$) under misclassification

- Assume non-differential misclassification, known error rates (sensitivity, specificity)
- **Likelihood Ratio** tests are **totally** unaffected (Bross, 1954)
Testing equality \((p = q)\) under misclassification

- Assume non-differential misclassification, known error rates (sensitivity, specificity)
- **Likelihood Ratio** tests are **totally** unaffected (Bross, 1954)
- Not as wacky as it seems (Rice & Holmans, 2003) - still just testing \(p = q\)
- Exact/mediocre/zero information on error rates... same result!
- Not a panacea: Type I error stays the same but **power** is reduced
- Doesn’t hold if we impose structure on \(p, q\)
### Estimation (regression) under misclassification

- Misclassification alters the (imposed) ‘structure’ of \( p, q \)

- ‘How much’ depends on the error rates \( \theta \) (sensitivity, specificity)

- We modify our ‘basic’ model appropriately (not trivial);

<table>
<thead>
<tr>
<th>Disease Status</th>
<th>Observed Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( AA )</td>
</tr>
<tr>
<td>Control</td>
<td>( \theta_{11}p_1 ) + ( \theta_{12}p_2 ) + ( \theta_{13}p_3 )</td>
</tr>
<tr>
<td>Case</td>
<td>( \theta_{11}q_1 ) + ( \theta_{12}q_2 ) + ( \theta_{13}q_3 )</td>
</tr>
</tbody>
</table>

- Apply your favorite statistical method (MLE/GEE/Bayes)

- With good knowledge of error rates \( \theta \), and ‘well-behaved’ data, estimation is not too contentious
## Badly behaved data

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Misclassification?</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>10</td>
<td>Ignore it!</td>
</tr>
<tr>
<td>AB</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Observed data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possible ‘truth’ (nice)</td>
<td>5</td>
<td>Sens, Spec $\approx 95%$</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Possible ‘truth’ (not nice)</td>
<td>0</td>
<td>Sens, Spec $\approx 90%$</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

- If the data ‘point to’ a zero in $p, q$, odds ratios go to zero, infinity

- Report this! – awkward, but ‘fixes’ are worse
  – unless you’re really Bayesian (Rice, 2003)

- For finite samples, rare genotypes, not unrealistic behavior
Confidence intervals for estimates

- Probably of more interest than estimate!
- Again, well-behaved data allow use of standard methods (though not standard software)
- Most methods ‘don’t work’ when OR estimate is infinite
- ‘Naive’ interval coverage is (very roughly) $95\% \times P(\text{data was nice})$
- For infinite estimates, the data at best really just give an upper/lower bound. Report this! (not a fudge)
Really badly behaved data

- Odds ratios can go to zero, infinity
- They can also go to 0/0 (even for a 2x2 table)
- This is undefined!
- No estimate or interval makes any sense
- The data are telling you they can’t answer your question at all
Different models for regression

Plain regression

Under misclassification
Different models for regression

Plain regression

Fast
Generic, Reliable
Easy to use
Same power as everyone else

Under misclassification

Slow
Problem-specific, Can behave oddly
Requires expertise (perhaps extra data)
Extra power available

Misclassification can make us think harder about the output, but we still just fit models
Regression Methods which aren’t really models

Some ‘non-standard’ techniques don’t fit ‘full’ models (e.g. Conditional Logistic Regression, Cox Regression)

In vehicular form;

‘Non-standard’ regressions
Regression Methods which aren’t really models

Some ‘non-standard’ techniques don’t fit ‘full’ models (e.g. Conditional Logistic Regression, Cox Regression)

In vehicular form;

‘Non-standard’ regressions

- You can still ‘go places’ but in a rather different way.
- Limited comparisons with full models – perhaps fast/slow?
- Very hard to extend for misclassified data

- Car → Truck  Not too hard
- Sub → Truck  ???
We really need one of these!

Finding ‘full model’ ways of explaining non-standard methods is useful (but hard);

- Helps explain *why* non-standard methods work so well
- Much easier to allow for misclassification;
- Sub $\rightarrow$ Truck is just Car $\rightarrow$ Truck
- Just *really* cool
Examples of this re-interpretation

- Matched case-control studies, binary exposure (Rice, 2004)
- Pair-matched case-control studies, multi-category exposure (submitted)
- Genetic Trio studies (Clayton & Cordell method) (in progress)
- Cox regression (in progress)

Enables adjustment for misclassification (e.g. Rice, 2003) but also missing data – and maybe more?
Coronary Calcium Scoring

- White ‘blobs’ in heart
- Agatston scoring method
- This is far from perfect
- Most analysis currently ignores this;
  - ‘Agatston > 0’ ~ genotype
  - CHD ~ Agatston
- How much does it matter?
Uncertainty in Agatston score

Just how bad is it? Scan each person twice;

<table>
<thead>
<tr>
<th>Score 2</th>
<th>Agatson Score 1</th>
<th>&lt;0.5</th>
<th>0.5-10</th>
<th>10-100</th>
<th>100-400</th>
<th>400+</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.5</td>
<td>3388</td>
<td>103</td>
<td>42</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.5-10</td>
<td>97</td>
<td>190</td>
<td>85</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10-100</td>
<td>34</td>
<td>87</td>
<td>1056</td>
<td>88</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>100-400</td>
<td>0</td>
<td>0</td>
<td>86</td>
<td>773</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>400+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>41</td>
<td>632</td>
<td></td>
</tr>
</tbody>
</table>

- Clearly **some** unreliability
- Broad agreement with other studies
Agatston score

- Suppose we care about Agatston $=0$, $>0$ only
  - simple approach takes average score

- Perhaps both scores are misclassified version of unseen ‘truth’?

- Can estimating effects of e.g. age, gender, blood pressure...

- Get error rates ‘for free’ (typically 96-99%)

- But
  - ‘Non-differential’ probably not right (in progress)
  - ‘True’ Agatston score a slippery concept

- Imperfect, but a reasonable sensitivity analysis
Agatston score cutoff

We used Agatston =0, >0. What about Agatston <10, >10? 20? 50?

Misclass-adjusted estimate a bit higher, same ball-park change in estimates – never a huge effect
In summary

- There are lies (and damned lies!)
- Also statistics
- The latter can help interpret the former

“It is easy to lie with statistics.
It is hard to tell the truth without it.”

Andrejs Dunkels
References

- Warner SL (1965) A survey technique for eliminating evasive answer bias, JASA 60:63-69


- http://faculty.washington.edu/kenrice