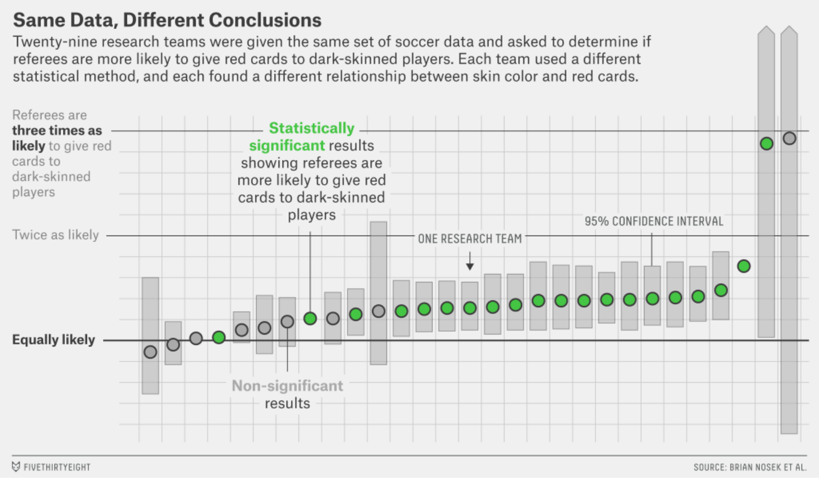
Researcher Choices and Bias

Data Analysis for Journalism and Political Communication (Fall 2024)

Prof. Bell


Same Data, Different Conclusions
Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.
Referees are three times as likely to give red cards to dark-skinned players
Statistically significant results showing referees are more likely to give red cards to dark-skinned players
Twice as likely
ONE RESEARCH TEAM
95% CONFIDENCE INTERVAL
Equally likely
Non-significant results
FIVETHIRTYEIGHT
SOURCE: BRIAN NOSEK ET AL.

Researcher Choices

1. What is my hypothesis?

Researcher Choices

1. What is my hypothesis?
2. What do I measure?

Researcher Choices

1. What is my hypothesis?
2. What do I measure?
3. How do I collect my data?

Researcher Choices

1. What is my hypothesis?
2. What do I measure?
3. How do I collect my data?
4. How much data do I collect?

RESEARCHER CHOICES

1. What is my hypothesis?
2. What do I measure?
3. How do I collect my data?
4. How much data do I collect?
5. What statistical analyses do I use?

Researcher Choices

1. What is my hypothesis?
2. What do I measure?
3. How do I collect my data?
4. How much data do I collect?
5. What statistical analyses do I use?
6. How do I handle outliers, missing data, and other peculiarities?

What is my hypothesis?

Choosing a hypothesis is all about avoiding error:

- **Type I error**: False positives
  - Sending the innocent to jail - this is bad!

What is my hypothesis?

Choosing a hypothesis is all about avoiding error:

- **Type I error**: False positives
  - Sending the innocent to jail - this is bad!
- **Type II error**: False negatives
  - Letting the guilty go free - we can accept this

What is my hypothesis?

Choosing a hypothesis is all about avoiding error:

- **Type I error**: False positives
  - Sending the innocent to jail - this is bad!
- **Type II error**: False negatives
  - Letting the guilty go free - we can accept this

Our goal is to reduce Type I error. Assume that the data is innocent (that the hypothesis is false) until it is proven guilty.

Measuring Type I Error

P-value

The chance that we would get a particular result from our test if the true answer is false

Measuring Type I Error

P-value

The chance that we would get a particular result from our test if the true answer is false

- The p-value is our chance of committing a Type I error - sending the innocent to jail

Measuring Type I Error

P-value

The chance that we would get a particular result from our test if the true answer is false

- The p-value is our chance of committing a Type I error - sending the innocent to jail
- Common p-value cut-offs in scientific research: .01, .05, and .1 indicate **statistical significance**

Measuring Type I Error

Hypothesis: A student cheated on an exam.

P-value (chance that we conclude that student cheated, but they did not): .50

Measuring Type I Error

Hypothesis: A student cheated on an exam.

- The student performed much better on this exam than on previous exams

P-value (chance that we conclude that student cheated, but they did not): .25

Measuring Type I Error

Hypothesis: A student cheated on an exam.

- The student performed much better on this exam than on previous exams
- The student finished their exam more quickly than other students

P-value (chance that we conclude that student cheated, but they did not): .15

Measuring Type I Error

Hypothesis: A student cheated on an exam.

- The student performed much better on this exam than on previous exams
- The student finished their exam more quickly than other students
- The student's roommate saw them up all night studying before the exam

P-value (chance that we conclude that student cheated, but they did not): .40

Measuring Type I Error

Hypothesis: A student cheated on an exam.

- The student performed much better on this exam than on previous exams
- The student finished their exam more quickly than other students
- The student's roommate saw them up all night studying before the exam
- The student missed the same questions as other students

P-value (chance that we conclude that student cheated, but they did not): .75

The Scientific Method

- P-values are a product of the data we use and our choices about what we include and exclude from the analysis.

The Scientific Method

- P-values are a product of the data we use and our choices about what we include and exclude from the analysis.
- We follow the scientific method: Theory $\Rightarrow$ Hypothesis $\Rightarrow$ Test $\Rightarrow$ Analyze $\Rightarrow$ Report

The Scientific Method

- P-values are a product of the data we use and our choices about what we include and exclude from the analysis.
- We follow the scientific method: Theory $\Rightarrow$ Hypothesis $\Rightarrow$ Test $\Rightarrow$ Analyze $\Rightarrow$ Report
- But in practice, no analysis plan survives contact with the data

Are Democrats or Republicans Good for the Economy?

Use FiveThirtyEight's online modeling tool to test what you think is the best approach to answering the question. There are no right or wrong answers - just select the model you think is best, and report your results in the form:

https://bit.ly/smpa2152

(The link to the tool is on the form.)

Misuses of P-values

- **P-hacking**: Manipulating data or analysis in order to find a statistically significant result

Misuses of P-values

- **P-hacking**: Manipulating data or analysis in order to find a statistically significant result
- **HARKing**: **H**ypothesizing **a**fter **r**esults are **k**nown

Misuses of P-values

- **P-hacking**: Manipulating data or analysis in order to find a statistically significant result
- **HARKing**: **H**ypothesizing **a**fter **r**esults are **k**nown
- **File drawer bias/publication bias**: Only publishing statistically significant results

Misuses of P-values

- **P-hacking**: Manipulating data or analysis in order to find a statistically significant result
- **HARKing**: **H**ypothesizing **a**fter **r**esults are **k**nown
- **File drawer bias/publication bias**: Only publishing statistically significant results
- Confusing statistical significance with substantive significance

What do I measure?

Operationalization

The process of defining a measurable version of a concept

What do I measure?

Operationalization

The process of defining a measurable version of a concept

Suppose you were interested in measuring “study quality,” a variable indicating how well a student studies. What are some ways you would measure this concept?

What do I measure?

Operationalization

The process of defining a measurable version of a concept

- Principles of good operationalization:
    1. Unambiguous

What do I measure?

Operationalization

The process of defining a measurable version of a concept

- Principles of good operationalization:
  1. Unambiguous
  2. Concise

What do I measure?

Operationalization

The process of defining a measurable version of a concept

- Principles of good operationalization:
  1. Unambiguous
  2. Concise
  3. Familiar

What do I measure?

Operationalization

The process of defining a measurable version of a concept

- Principles of good operationalization:
  1. Unambiguous
  2. Concise
  3. Familiar
  4. Available

Exercise: Operationalization

1. You want to measure how happy people are
2. You want to measure people's driving ability
3. You want to measure the political ideology of a member of Congress

How do I collect my data?

Data Generating Process

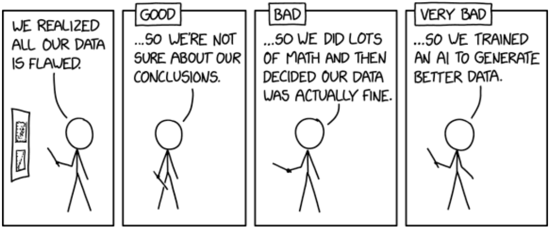
The rules and procedures that produce the data one is interested in

How do I collect my data?

Data Generating Process

The rules and procedures that produce the data one is interested in

- No amount of statistical wizardry can compensate for bad data



How do I collect my data?

Data Generating Process

The rules and procedures that produce the data one is interested in

- No amount of statistical wizardry can compensate for bad data
- The gold standard of data generating processes is the **random sample**

Sampling

- The group we are interested in studying is known as the **population**

Sampling

- The group we are interested in studying is known as the **population**
- Often, we are not able to count every unit in the population, so we take a **sample**

SAMPLING

- The group we are interested in studying is known as the **population**
- Often, we are not able to count every unit in the population, so we take a **sample**
- The key to any data analysis project is a quality sample, which is determined by two elements:

Sampling

- The group we are interested in studying is known as the **population**
- Often, we are not able to count every unit in the population, so we take a **sample**
- The key to any data analysis project is a quality sample, which is determined by two elements:
  1. A **random sample** of the population

Sampling

- The group we are interested in studying is known as the **population**
- Often, we are not able to count every unit in the population, so we take a **sample**
- The key to any data analysis project is a quality sample, which is determined by two elements:
    1. A **random sample** of the population
    2. The **sample size** is sufficiently large

Random Sample

Definition

The probability of any given unit being drawn from the population is uniform (the same)

Random Sample

Definition

The probability of any given unit being drawn from the population is uniform (the same)

- A failure of each unit to have a uniform probability of being drawn from the population is known as **selection bias**

Random Sample

Definition

The probability of any given unit being drawn from the population is uniform (the same)

- A failure of each unit to have a uniform probability of being drawn from the population is known as **selection bias**
- Units are "selecting" into our data because they are more observable than other units

Random Sample

Definition

The probability of any given unit being drawn from the population is uniform (the same)

- A failure of each unit to have a uniform probability of being drawn from the population is known as **selection bias**
- Units are "selecting" into our data because they are more observable than other units
- Selection bias reduces our **generalizability** to the population because the data is not representative of the population

Random Sample

Can I randomly sample 10 students from this class to generalize to:

- the population of GW students?

Random Sample

Can I randomly sample 10 students from this class to generalize to:

- the population of GW students?
- the population of SMPA students?

Random Sample

Can I randomly sample 10 students from this class to generalize to:

- the population of GW students?
- the population of SMPA students?
- the population of Data Analysis students?

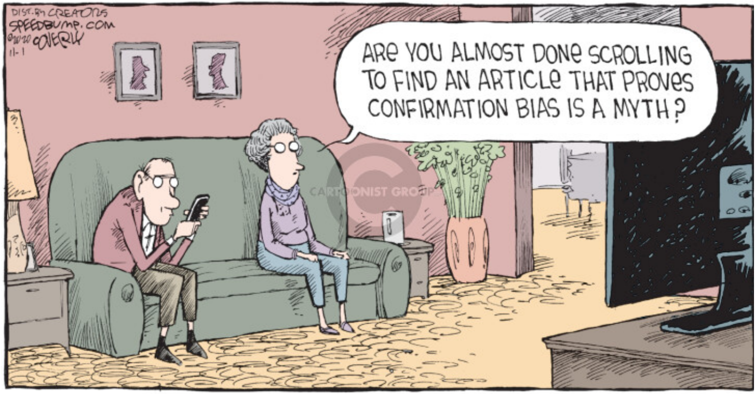
Exercise: Selection Bias

Biases in Research

1. Confirmation bias
2. Desirability bias
3. Authority bias
4. Availability bias
5. Certainty bias

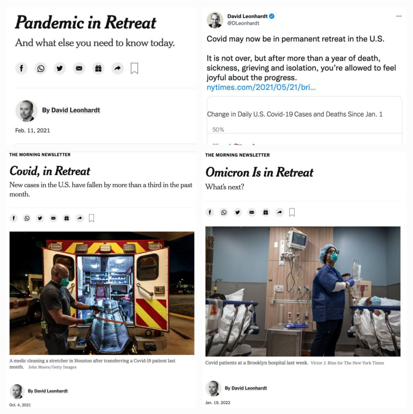
Confirmation Bias

We privilege evidence that supports our existing beliefs and discount evidence that challenges those beliefs.



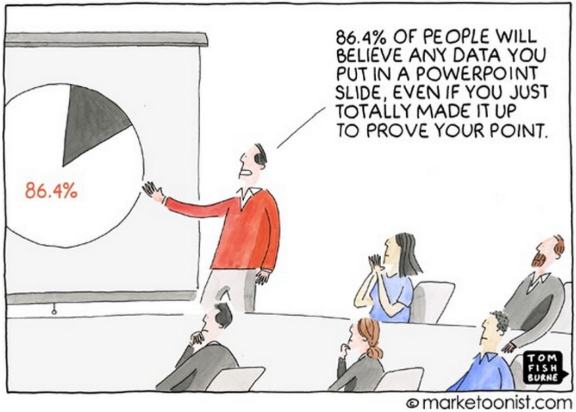
DESIRABILITY BIAS

We prefer evidence that supports a conclusion we want to be true and discount evidence that undermines that conclusion.



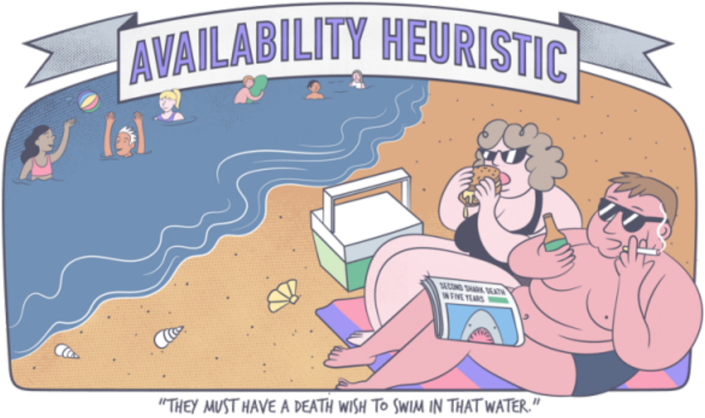
Authority Bias

We give greater weight to evidence offered by people in positions of authority.



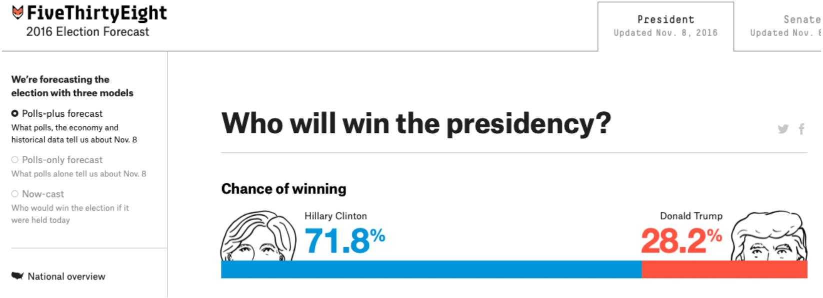
Availability Bias

We give greater weight to evidence that is most memorable.



Certainty Bias

We over- and under-state probabilistic evidence.



Retro Report (2021) - What's In a Number?

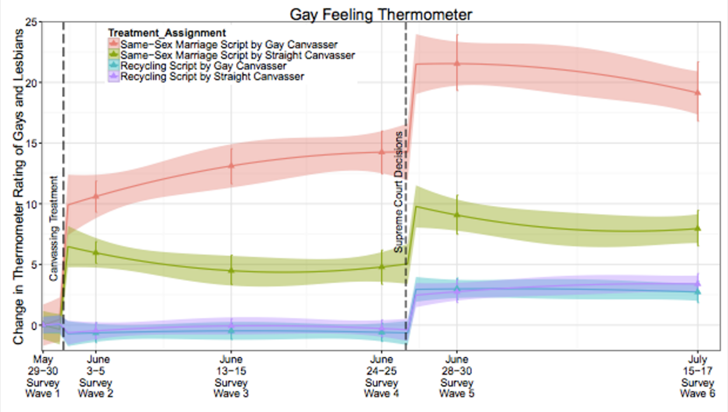
Fraudulent Research

- Unfortunately, fraud does happen - perhaps more than researchers would like to admit

Fraudulent Research

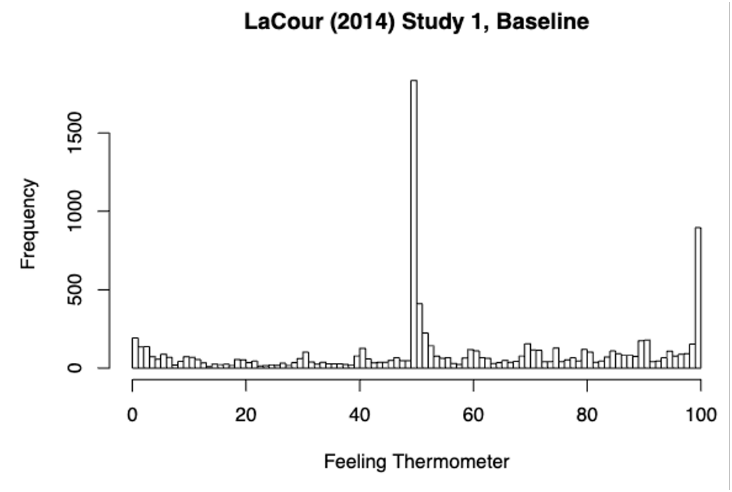
- Unfortunately, fraud does happen - perhaps more than researchers would like to admit
- One of the most infamous cases of fraud was a study linking the childhood MMR (mumps, measles, and rubella) vaccine to autism

Gay Canvassers Scandal



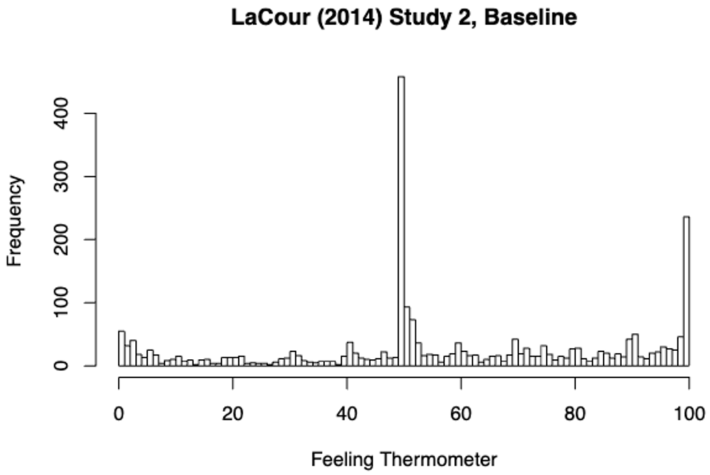
Fraudulent Research

Gay Canvassers Scandal



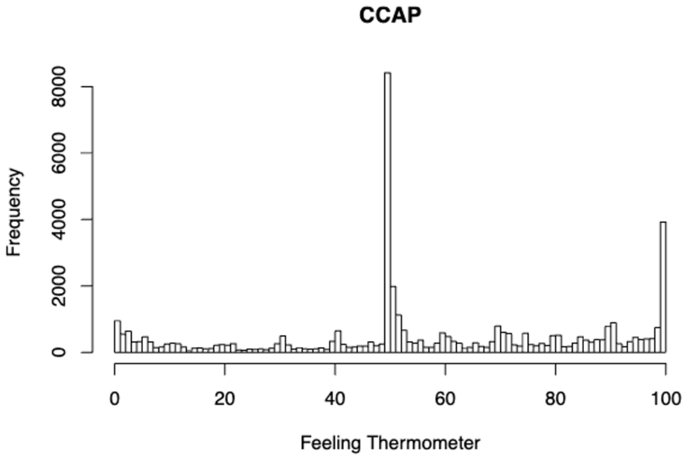
Fraudulent Research

Gay Canvassers Scandal



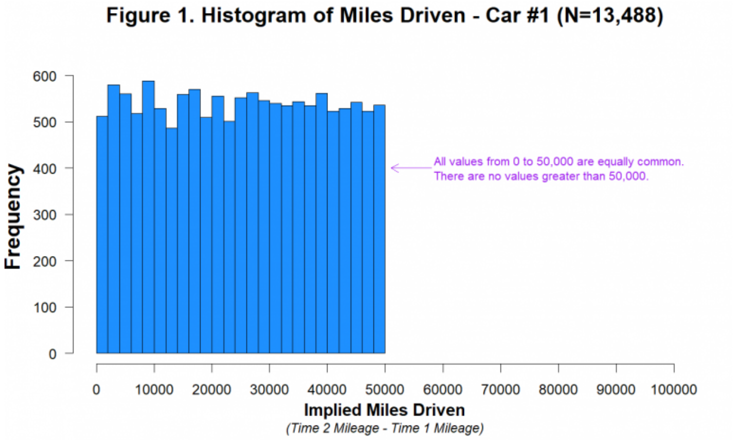
Fraudulent Research

Gay Canvassers Scandal



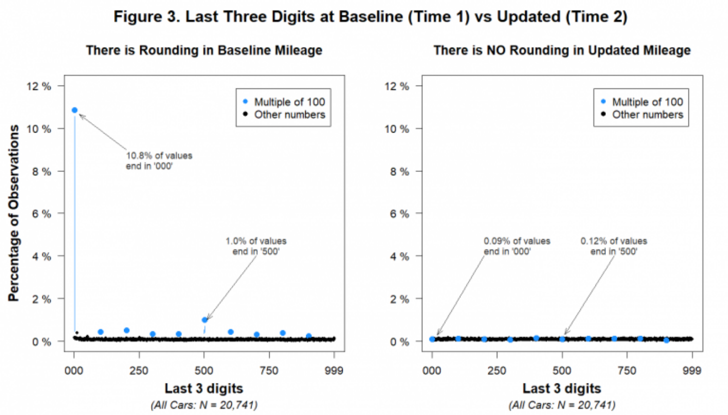
Fraudulent Research

Dishonesty in Dishonesty Research (uncovered by Data Colada)



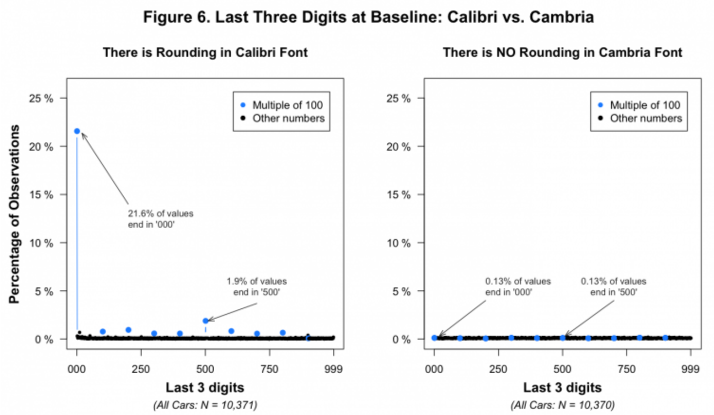
Fraudulent Research

Dishonesty in Dishonesty Research (uncovered by Data Colada)



Fraudulent Research

Dishonesty in Dishonesty Research (uncovered by Data Colada)



Fraudulent Research

- Unfortunately, fraud does happen - perhaps more than researchers would like to admit
- One of the most infamous cases of fraud was a study linking the childhood MMR (mumps, measles, and rubella) vaccine to autism
- Red flags:
  - Not sharing data or methods

Fraudulent Research

- Unfortunately, fraud does happen - perhaps more than researchers would like to admit
- One of the most infamous cases of fraud was a study linking the childhood MMR (mumps, measles, and rubella) vaccine to autism
- Red flags:
  - Not sharing data or methods
  - Unusual patterns in the data

Fraudulent Research

- Unfortunately, fraud does happen - perhaps more than researchers would like to admit
- One of the most infamous cases of fraud was a study linking the childhood MMR (mumps, measles, and rubella) vaccine to autism
- Red flags:
  - Not sharing data or methods
  - Unusual patterns in the data
  - Non-reproducible outcomes or outcomes not supported by theory

Fraudulent Research

- Unfortunately, fraud does happen - perhaps more than researchers would like to admit
- One of the most infamous cases of fraud was a study linking the childhood MMR (mumps, measles, and rubella) vaccine to autism
- Red flags:
  - Not sharing data or methods
  - Unusual patterns in the data
  - Non-reproducible outcomes or outcomes not supported by theory
  - Conflicts of interest

Fraudulent Research

- Unfortunately, fraud does happen - perhaps more than researchers would like to admit
- One of the most infamous cases of fraud was a study linking the childhood MMR (mumps, measles, and rubella) vaccine to autism
- Red flags:
  - Not sharing data or methods
  - Unusual patterns in the data
  - Non-reproducible outcomes or outcomes not supported by theory
  - Conflicts of interest
- These "red flags" do not necessarily mean the data is fraudulent – but be vigilant