Risk Analysis
and
Best Practices Benchmarking

Three Case Studies

"How Risky is My Risk Analysis?"

For

LONE STAR®

Topics

- Introduction to the Benchmarking Effort – MBP2

- Checklists from MBP2 applied to three risk analysis case studies
  - Policy/Legislation
  - Cyber Attack
  - Insurance Reserves

About the Benchmarking

- Benchmarking Project = MBP2 “Modeling Best Practices Benchmarking Project”
  - Began in 2015
- Data collection included surveys and multi-stage interviews
  - Cooperation and support from a number societies and non-profits
  - In-process papers and reports have been given with useful feedback
  - Initial reports have been issued
  - Book Draft in work
- Three Checklists Developed
  - Two used for the case studies today

The Best Practices

14 "yes" Answers, Please

1. **Intended use** – Is it clear why modeling is being done?
2. **Semantic Clarity** – Is there agreement on what words mean and which measures are preferred?
3. **Design Environment for low cost, high value** – Do tools enable rapid and cost efficient development of models?
4. **Process Discipline** – Are there clear model development processes
5. **Transparency (Glass Box Models)** – Do those with a "right to transparency" have easy insight into how the model works?
6. **People Driven; Subject Matter & Analysis Talent** - Can real humans put data in? Do they "get" the answers coming out?
7. **Open interfaces** – Is it easy to get data in and out?
8. **Accommodate Complexity** – Does the model adequately cope with real-world complexity and interconnections of systems represented? Is there what Box called "needless elaboration"?
9. **Accommodate Diversity** – Does the model accommodate disciplines who may not use the same measures or semantics?
10. **Accommodate Uncertainty (in cognition, representation, computation)** – Does the model incorporate the full span of mathematical uncertainty and is it preserved with correct computational methods? Is uncertainty provided to users in a way compatible with cognitive limits? Does the model do "the Arithmetic of Uncertainty" correctly?
11. **Accommodate Audit & Validation** – Does the process ensure error detection and correction is done?
12. **Provide Security** – Does the system provide security and privacy protection adequate to comply with applicable obligations, and to protect stakeholders?
13. **Processing and Network Compatibility** – Do processing loads and data flows fit within the time and cost constraints of the modeling purposes?
14. **Statutory and Regulatory Compliance** – are obligations clearly understood and is compliance documented?

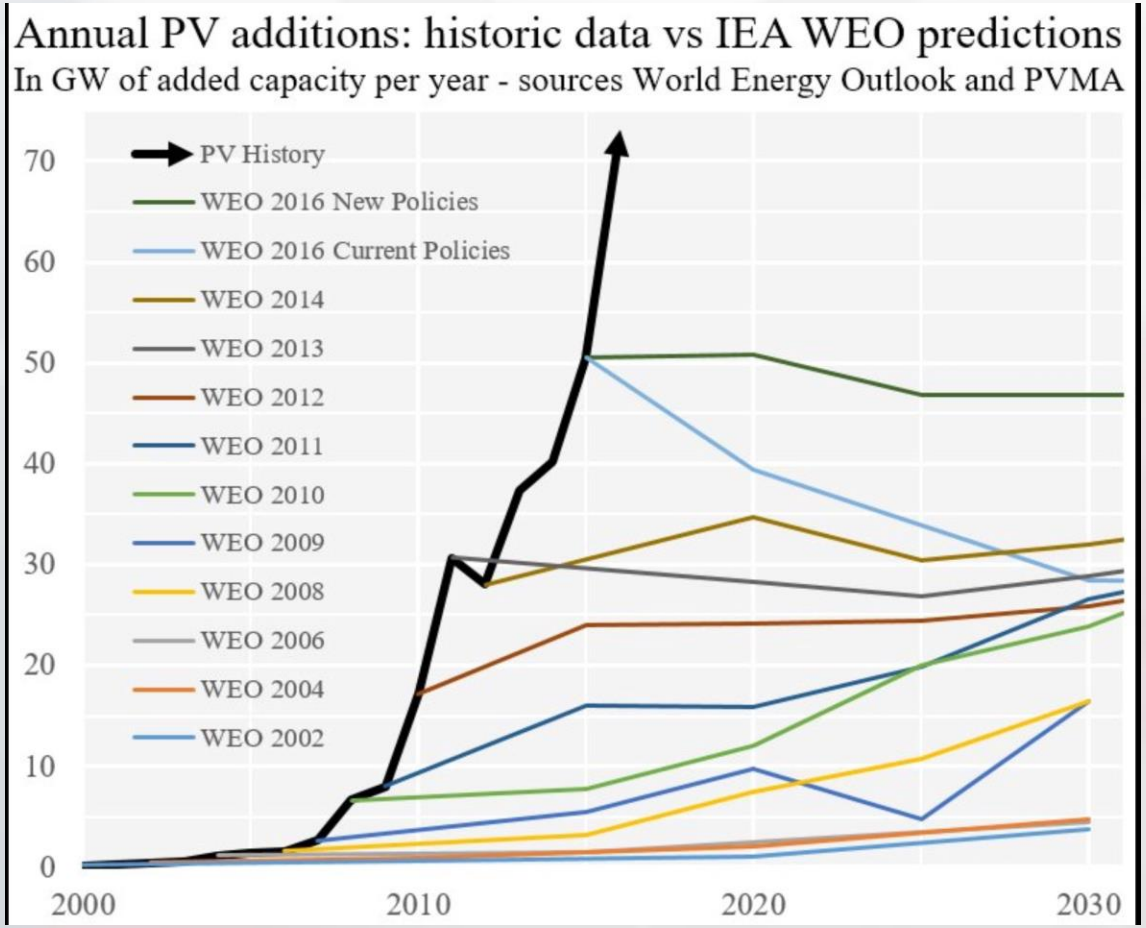
Best Practice Helps Avoid Risk, But Not Assess Risk

The Best Practitioners

- We found zero organizations who could consistently say “yes” to all 14
- We found four organizations who fully understood all 14, and who took steps to cope with shortcomings
- Two of the best practitioners agreed to be publicly acknowledged

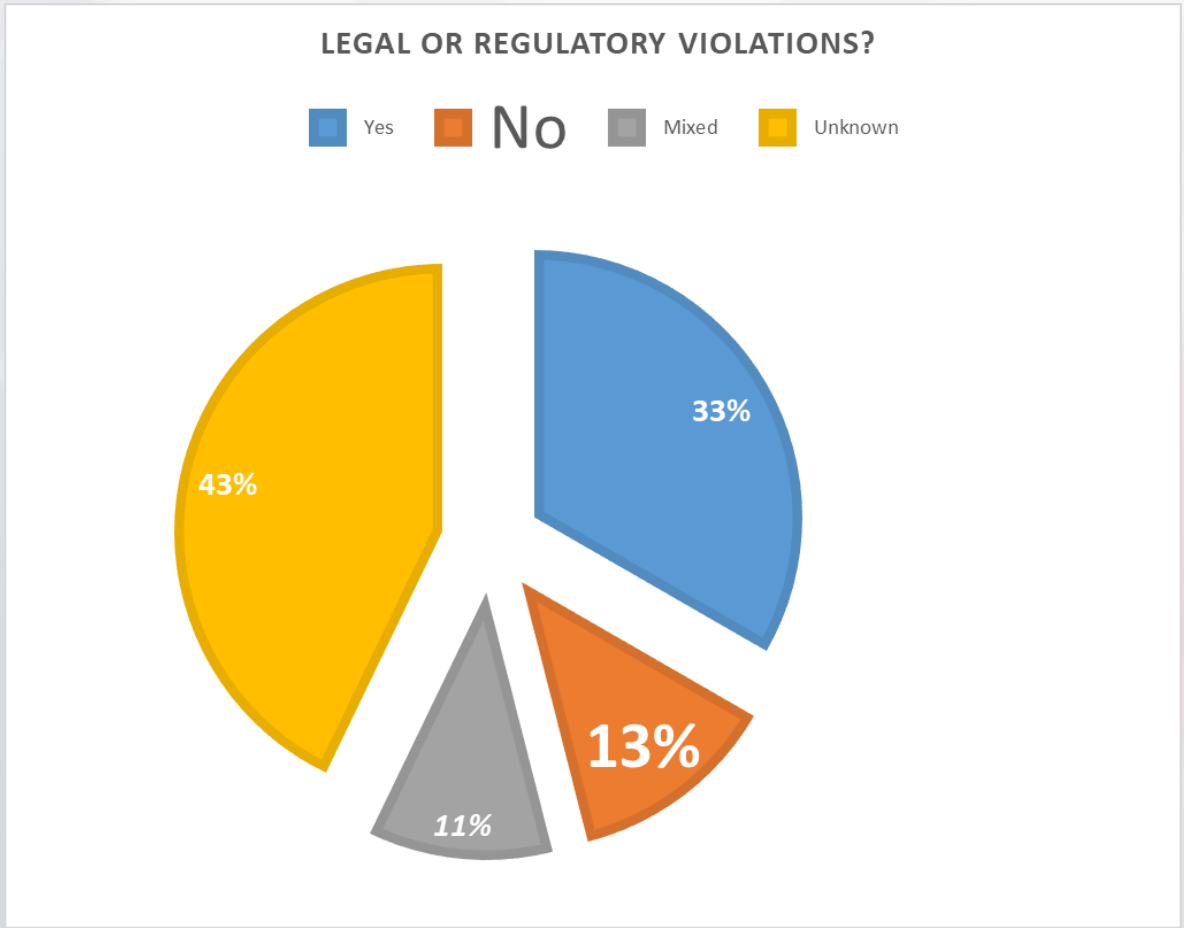
Lots of Bad Practice to Critique

- The supply of bad practice in modeling and simulation seems inexhaustible
- One example – more than a decade of IEA's modeling to predict solar energy capacity additions
  - The colored lines are the forecasts
  - Black line is what happened
  - ***Why this persistent badness?***



Less than "Best" – Legal/Regulatory

- Shown here – Interviews
  - Only 13% clearly compliant
- Survey probably worse?
- Remember – these are supposed to be best practice candidates, not the average practitioners
- Logical question – ***Do most analytics break some rule, law, or guideline?***



Learning From Contrast

- The U.S. EIA and the E.U. IEA were a striking contrast
  - EIA was one of the best practitioners
  - IEA did not even enter the early elimination round – strikingly bad practice and repeatedly bad results
- We noticed some of the best and worst practitioners were *government agencies*
- There was no particular or obvious pattern at first
  - Not associated with mission – EIA and IEA do the same thing
  - Not associated with national laws, rules or departments – one anonymous *best practitioner* was in the same agency with *a group so bad* we coined the term “mathematical malpractice”
- These contrasts seemed worthy of serious consideration

Regression Derived Risk Checklist

1. Relying on repetition?
2. Focused on budgets, revenues, heft?
3. Using analytics professionals?
4. Frozen MS&A tools and processes are dangerous.
5. Trusting in formal approvals or audits?
6. Looking at single numbers, not spans of uncertainty? Use of averages is a danger signal.
7. Is the analytics provider held accountable? Accountability significantly reduces risk. Lack of accountability raises risks.

1. Being repeatedly bad is sadly common
2. Bigger is not better, budgets don't assure excellence
3. Pros can lower risk… sometimes
4. Unable or unwilling to change = danger (the world keeps changing)
5. The smiley face sticker mattered in kindergarten
6. There is no reliable means to conduct analysis without correct representation of uncertainty
7. Absolute power corrupts absolutely

Applying the Checklists
3 Case Studies

- A "Giant Company's" Long Term Care Insurance Reserves

- General Accountability Office response to Congressional policy concerns over contracting protests

- UK National Health Service Response to the WannaCry Cyber Attack

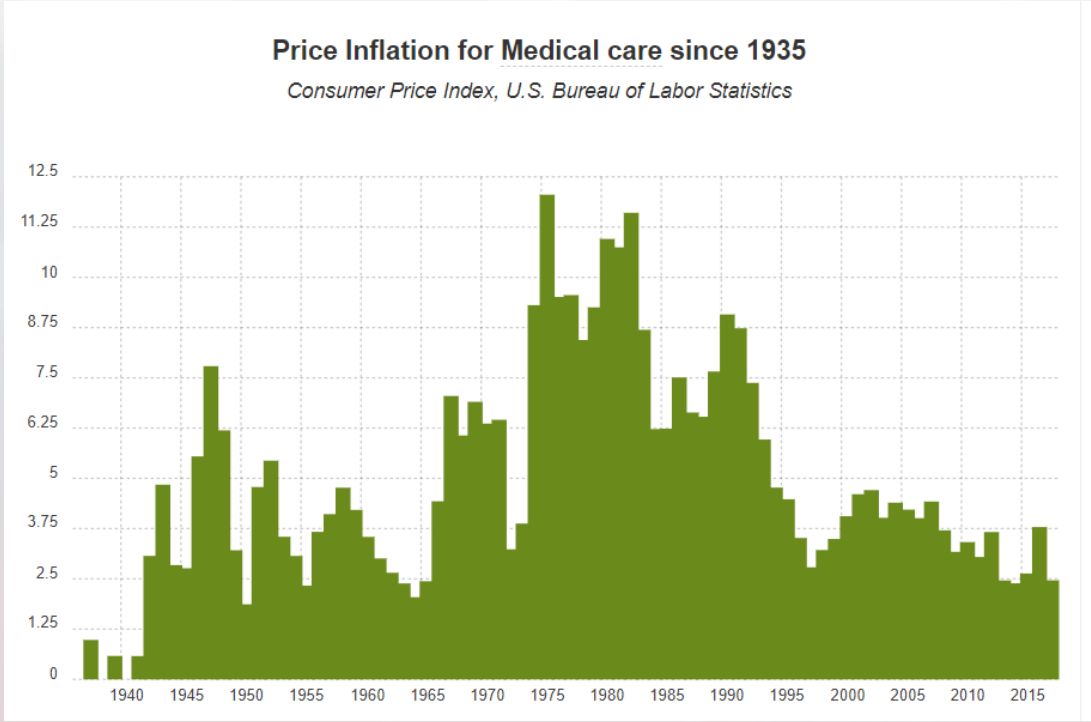
A Giant Company's Logo Here

"GC"

Case Study 1
Long Term Care Underwriting Risks

GC's Long Term Care Write-off

- A very complex story
- Now and SEC investigation
- Write off to date is probably about $15B
- Most recent write-off was $6 – 8 Billion (depending on tax treatment and the analyst)
- Seeds were sown in the 1980's
- How could these "smart guys" be so wrong?

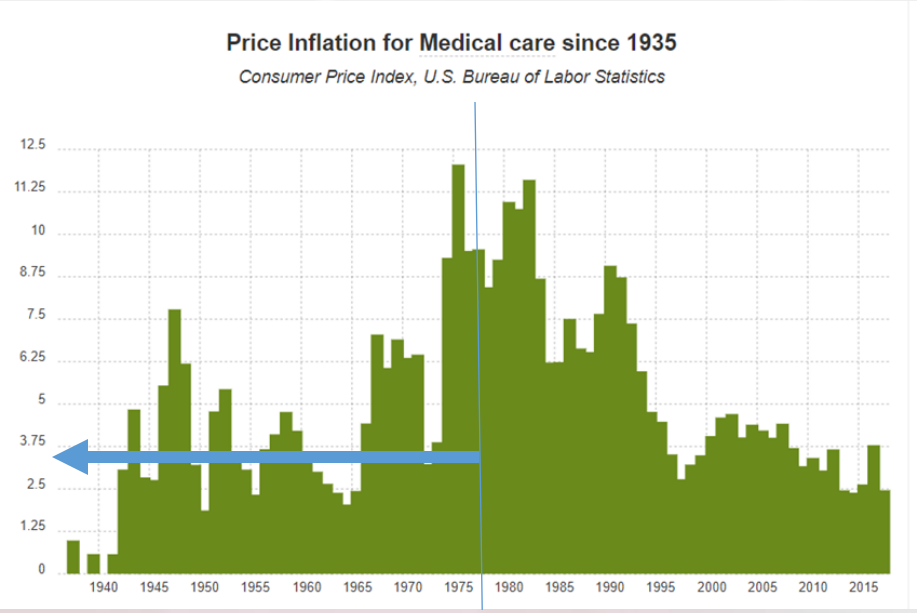


Biggest Error – Deterministic Assumptions?

- SEC investigations and other disputes make it impossible to know objectively what happened back in the 1980's
- By most accounts GC used deterministic numbers
  - Other long term risks had benefited from this; e.g., underwriting life insurance as life spans increase.
- Early 1980's were a bad time to pick point estimates
  - Investment returns were at historic highs
  - Health care inflation (in retrospect) had a moderate median expectation
- Seems to be compounded with other "unlucky" swings
  - Cost Per Stay
  - Duration of Stays
  - Longevity of Customers

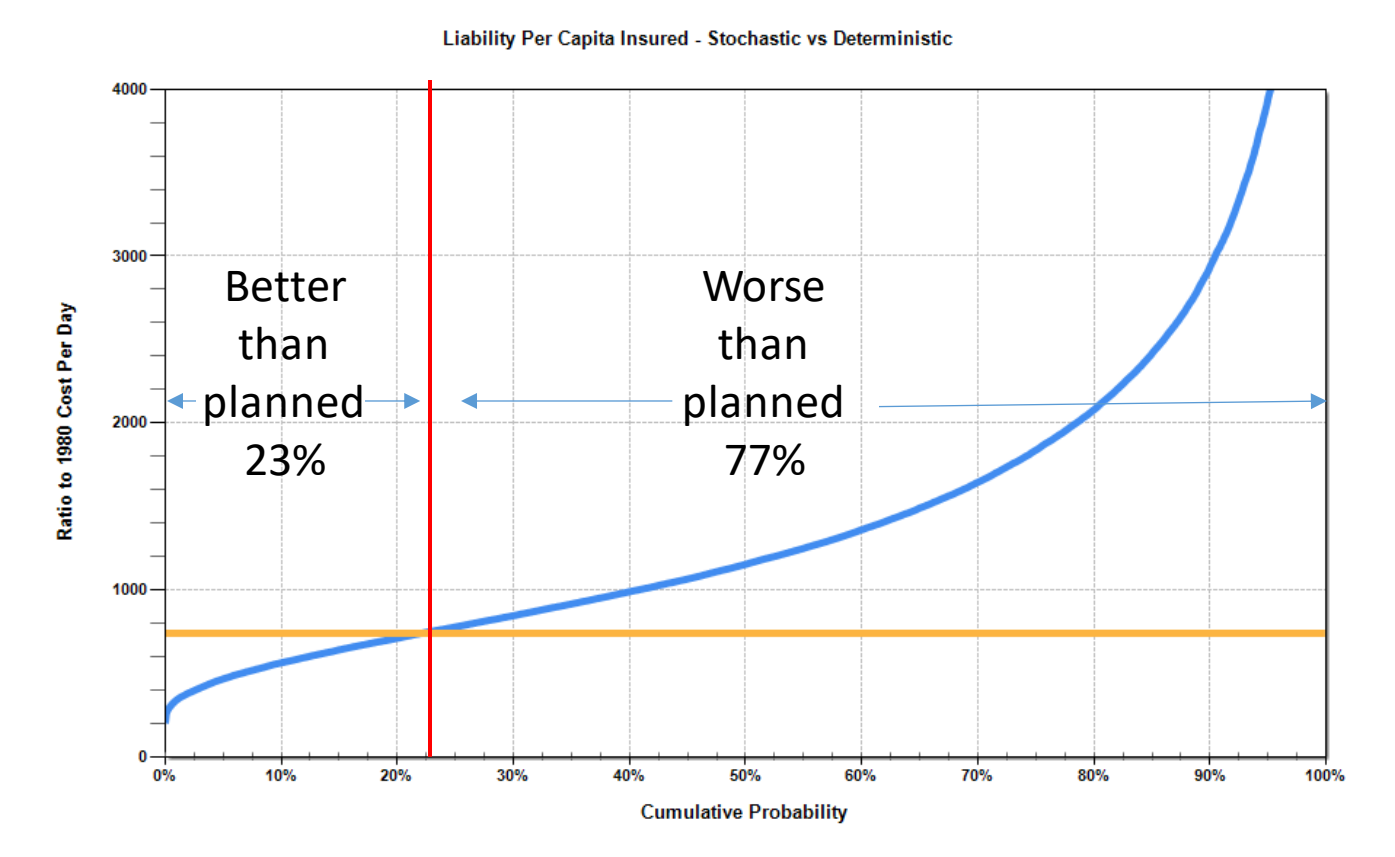
Example; Inflation - Returns

- Historical data in the 1980's suggested a median healthcare inflation rate around 3% (50 years prior)
- GC business model reportedly assumed a 7.5% rate of return on invested funds (remember 10% bank CDs?)
- **Net spread of 4.5% Assumed?**
- But looking at a spread of returns and inflators analysis *at the time* should have show there was a less than a 40% chance this would be true
- Reality was worse
  - The net returns proved to be more like 4% and the inflator was more like 5%
  - Actual spread was 1% the wrong way
  - 5.5% from the apparent assumption – compounded over many years



Brutal Math

Multiplying Distributions is Unforgiving



- Deterministic Number methods It probably looked like the per capita cost insured would be about 800 days x 1980 cost per day
- There are seven key assumptions
- All of them have asymmetric distributions
- All of them are skewed the “wrong” way – even with what was knowable in the 1980s
- GC seems to have had about a 20-25% chance meeting or exceeding their targets in the 1980’s but the next 30 years were even worse

Risk Checklist – GC Long Term Health Care

1. Relying on repetition?
2. Focused on budgets, revenues, heft?
3. Using analytics professionals?
4. Frozen MS&A tools and processes are dangerous.
5. Trusting in formal approvals or audits?
6. Looking at single numbers, not spans of uncertainty? Use of averages is a danger signal.
7. Is the analytics provider held accountable? Accountability significantly reduces risk. Lack of accountability raises risks.

1. **GC was thought to be risk & finance savvy; life insurance**
2. **GC wanted to be #1 or #2 and seems to have chased volume**
3. GC used actuaries – not clear if they were listened to benchmarking project included insurance company actuaries who generally felt ignored
4. **GC seems to have used methods for risks which were more static**
5. **Unclear that regulators understood the risks either**
6. **GC apparently used single number proxies for uncertainty**
7. **Long time between pricing and payout means no one in the '80s would ever be accountable**

Six Warning Flags – GC Risk Pricing Assessments Were Clearly Risky

Case Study 2
Public Policy Risks

GAO and Protest Reform

- For nearly a decade, the House Armed Services Committee (HASC) wrote the same letter to the GAO asking two questions

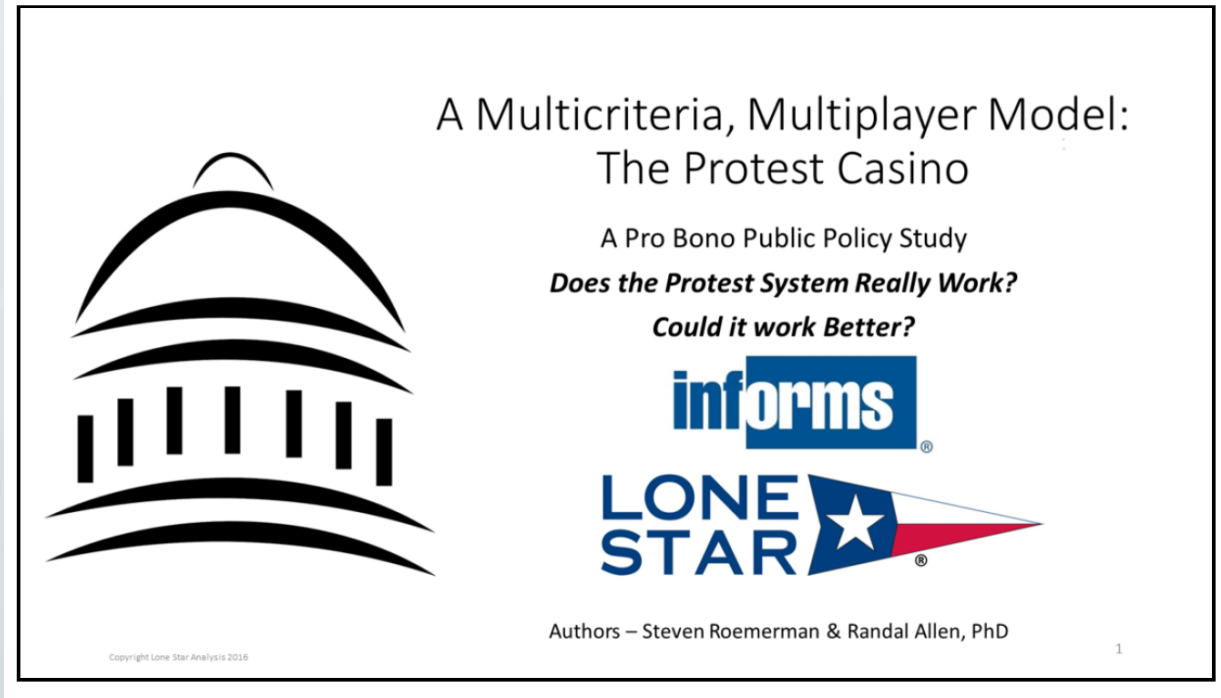
1. *How many protests are frivolous?*
2. *Why is the problem getting worse?*

- Each year the GAO wrote back with roughly the same two answers:

1. *Silly congress, there are none*
2. *Silly congress, the data shows no problem*

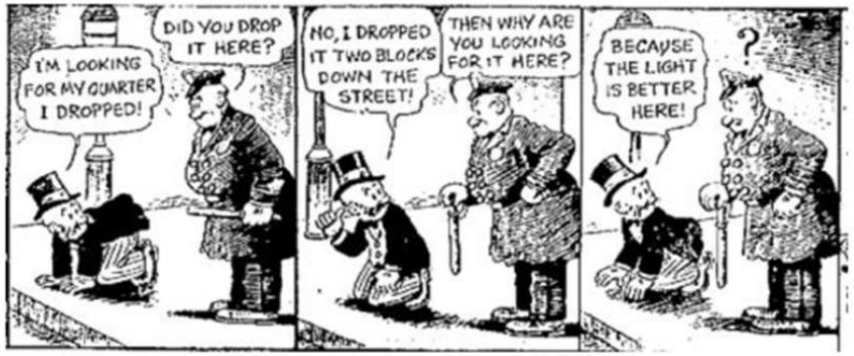
The “Protest Casino”

- For the HASC, we conducted research, interviews and built game theory models
  - Congress made the rules but was the game rigged so the house would lose?
- Our work was reviewed by the Congressional Research Service with no resulting corrections
- The HASC has proceeded along the lines our study suggested – written into law
- **Dramatically different findings than the GAO…why?**



Excerpts from the Protest Casino Study

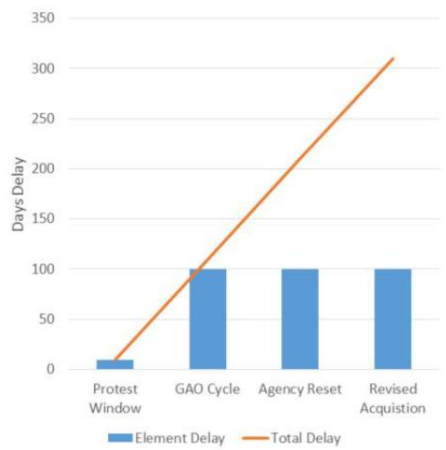
Survey Vs. GAO Metrics



- GAO collects and reports a number of metrics; Some are quoted because they are available – not because they have real meaning
- "Streetlight Effect" form of observational bias
- We had to carefully consider which GAO metrics had meaning for our simulation purposes

15

Confusing Metrics Example - Protest Delay



- Supporters of the current system point to GAO promptness; nearly always meets the 100 day time limit to "deal with protests"
- But – what does this metric mean?
- Chart on the left is an example
  - 100 Day GAO cycle is about 32% of the total time impact for this one example
  - Some crude modeling suggests that GAO's cycle time is **less than half the total impact in vast majority of cases**

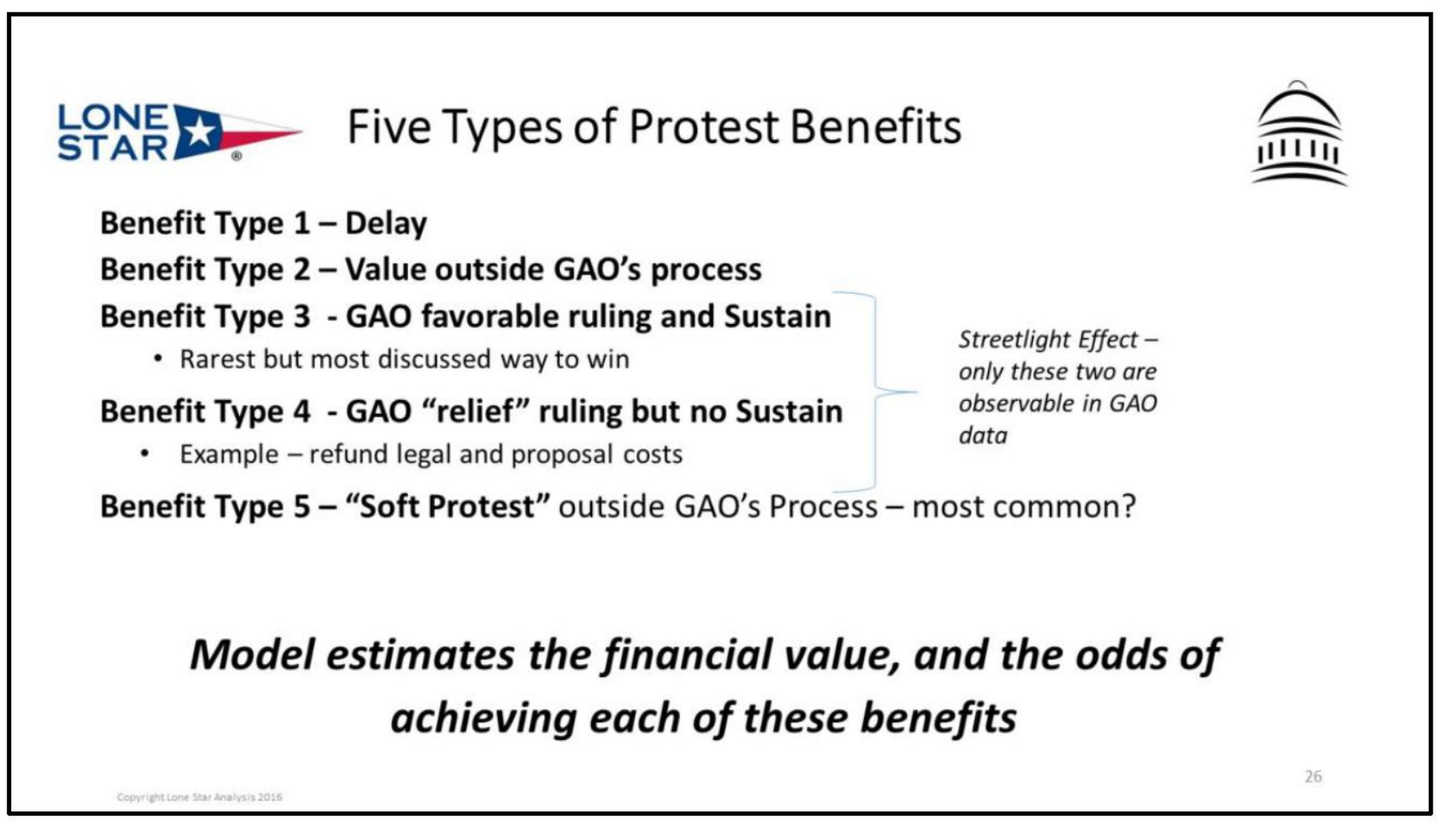
GAO's 100 Days – Is that the right measure?

16

Key Finding

How Companies Filing Protests Measure "Benefits"

Hint … Differently than GAO



GAO Protest Analysis vs Best Practices

3 "yes" Answers, 11 "no" Answers

1. **Intended use** – Is it clear why modeling is being done?
2. **Semantic Clarity** – Is there agreement on what words mean and which measures are preferred?
3. **Design Environment for low cost, high value** – Do tools enable rapid and cost efficient development of models?
4. **Process Discipline** – Are there clear model development processes
5. **Transparency (Glass Box Models)** – Do those with a "right to transparency" have easy insight into how the model works?
6. **People Driven; Subject Matter & Analysis Talent** - Can real humans put data in? Do they "get" the answers coming out?
7. **Open interfaces** – Is it easy to get data in and out?
8. **Accommodate Complexity** – Does the model adequately cope with real-world complexity and interconnections of systems represented? Is there what Box called "needless elaboration"?
9. **Accommodate Diversity** – Does the model accommodate disciplines who may not use the same measures or semantics?
10. **Accommodate Uncertainty (in cognition, representation, computation)** – Does the model incorporate the full span of mathematical uncertainty and is it preserved with correct computational methods? Is uncertainty provided to users in a way compatible with cognitive limits? Does the model do "the Arithmetic of Uncertainty" correctly?
11. **Accommodate Audit & Validation** – Does the process ensure error detection and correction is done?
12. **Provide Security** – Does the system provide security and privacy protection adequate to comply with applicable obligations, and to protect stakeholders?
13. **Processing and Network Compatibility – Do processing loads and data flows fit within the time and cost constraints of the modeling purposes?**
14. **Statutory and Regulatory Compliance – are obligations clearly understood and is compliance documented?**

So – not good practice but is that the same thing as RISK?

Risk Checklist – GAO Protest Analysis

1. Relying on repetition?
2. Focused on budgets, revenues, heft?
3. Using analytics professionals?
4. Frozen MS&A tools and processes are dangerous.
5. Trusting in formal approvals or audits?
6. Looking at single numbers, not spans of uncertainty? Use of averages is a danger signal.
7. Is the analytics provider held accountable? Accountability significantly reduces risk. Lack of accountability raises risks.

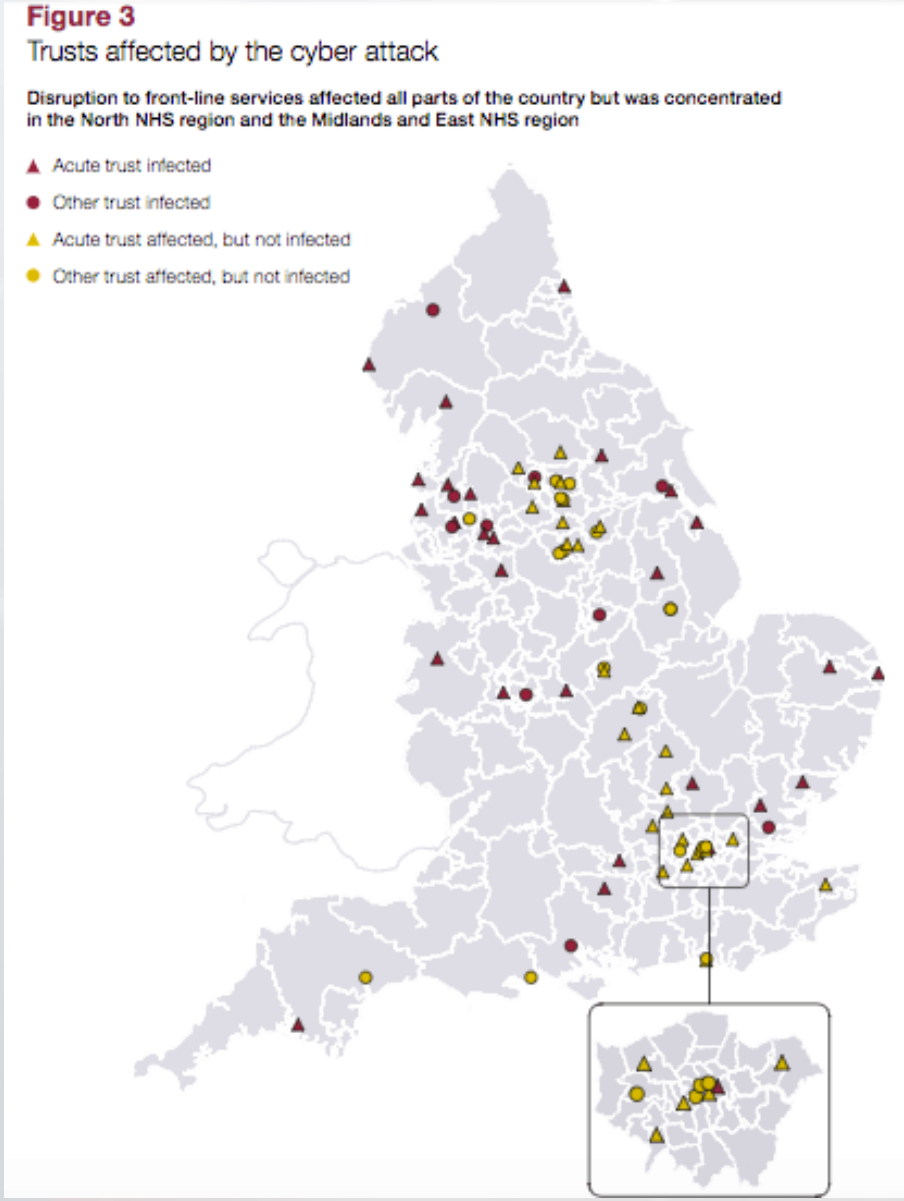
1. **GAO's processes & measures went back decades repetition was key to trend assessments**
2. Not clear "heft" was a problem
3. **Key GAO staff were attorneys**
4. **GAO resisted HASC suggested changes**
5. **An echo chamber from the protest bar suggested this was really good**
6. **GAO consistently used single number proxies for uncertainty**
7. **GAO's charter is to hold others accountable, not be actually *be* accountable**

Six Warning Flags – GAO Protest Assessments Were Clearly Risky

Case Study 3
Cyber Risks

The Attack

- May 12, 2017 – The Wannacry Ransomware attack strikes roughly 200,000 organizations in more than 100 nations
- None were hit harder than the National Health Service (NHS)
- Two UK Government Reports used for this analysis
  - October 2017 National Audit Office (NAO)
  - February 2018 NHS Report



What NHS Said

On Friday 12 May 2017, a global ransomware attack, known as WannaCry, affected a wide range of countries and sectors. Although WannaCry impacted the provision of services to patients, the NHS was not a specific target.

The NHS responded well to what was an unprecedented incident, with no reports of harm to patients or of patient data being compromised or stolen. In total, 1% of NHS activity was directly affected by the WannaCry attack. 80[2] [3] out of 236 hospital trusts across England were affected[4], which means that services were impacted even if the organisation was not infected by the virus (for instance they took their email offline to reduce the risk of infection). 595 out of 7,454[5] GP practices (8%) and eight other NHS and related organisations were infected. This disruption to patient care has made it even clearer how dependent the NHS is on information technology and, as a result, the need for security improvements to be made across the service.

Not Aimed at us

We did good

Zero Harm?

1% of Activity?

- The NHS document is supposed to be a lessons learned document
- It comes to conclusions much different than the earlier NAO Report

Are NHS improvements based on solid risk assessments?

One Percent?

- Was it 100% (all IT systems shut down)
- At least 34% (“at least 81 out of 236 trusts…” according to NAO)
- 8% (600 or more local offices out of about 7500)
- 3% (number of patients turned away)

A further 603 primary care and other NHS organisations were infected by WannaCry, including 595 GP practices. However, the Department does not know how many NHS organisations could not access records or receive information, because they shared data or systems with an infected trust. NHS Digital told us that it believes no patient data were compromised or stolen (paragraphs 1.2 to 1.5 and 1.9, and Figure 1).

NAO Report

No Harm?

- 19,000 to 30,000 Patient Appointments and Treatments Canceled?
- 139 Urgent Cancer Treatments Postponed
- 5 hospitals turned away emergency patients
- Unknown number of ambulances were unavailable due to shuttling patients

on the normal rate of follow-up appointments to first appointments. NHS England told us it does not plan to identify the actual number because it is focusing its efforts on responding appropriately to the lessons learned from WannaCry. As data were not collected during the incident, neither the Department nor NHS England know how many GP appointments were cancelled, or how many ambulances and patients were diverted from the five accident and emergency departments that were unable to treat some patients (paragraphs 1.7, 1.8 and 1.10, and Figure 1).

NAO Report

“No Harm” Claim

- If people have emergences and can’t be treated what are the odds harm was done?
- If about 25,000 appointments were canceled or delayed, what are the odds harm was done?
- If 139 (or more) cancer patents had “urgent” treatments delayed for a week or more what are the odds harm was done?

Conclusions Psychological distress is associated with increased risk of mortality from several major causes in a dose-response pattern. Risk of mortality was raised even at lower levels of distress.

Conclusion from a 2012 BJM study funded by the NHS

A note to The NHS

- On February 8 we sent a request for a response to the following points about the NHS Report:

1. The 1% claim lacks semantic clarity and objective meaning
2. Use of deterministic measures when faced with significant uncertainty is flawed
3. “No reports of harm” violates both of the above, conflates absence of evidence with evidence of absence, and contradicts NHS science showing disruption and stress is harmful to the ill, injured and those at risk
4. Transparency is a best practice, but this analysis is a opaque

William Smart , Chief Information Officer for Health and Social Care
Skipton House,
80 London Road,
London,
SE1 6LH
Email: England.CIOReview@nhs.net

Multiplying Distributions - Again

- What the NHS did was to multiple distributions
- They committed several forms of mathematical malpractice when they did it
- They used single digit proxies for what was really spans of uncertainty
- In EVERY case they picked the lowest possible number for the impact

Multiplying the Distributions

Chances the number of infected systems is right – about 1%

Chances impact to care access per system is right – about 1%

Chances impacted care to patient per impacted system is right – about 1%

So… Chance the NHS is right 1 x 10^-6 ??

Risk Checklist – NHS Cyber Analysis

1. Relying on repetition?
2. Focused on budgets, revenues, heft?
3. Using analytics professionals?
4. Frozen MS&A tools and processes are dangerous.
5. Trusting in formal approvals or audits?
6. Looking at single numbers, not spans of uncertainty? Use of averages is a danger signal.
7. Is the analytics provider held accountable? Accountability significantly reduces risk. Lack of accountability raises risks.

1. **NHS repeated the same general assessments which had been used to assure Parliament before the attack**
2. **NHS CIO's office does seem to want more money**
3. **NHS analysts are unnamed, but the "analysis" is consistently bad – IT guys?**
4. **NHS resisted using assessment methods NAO wanted?**
5. **NHS seems to rely on cyber checklists and advisories**
6. **NHS consistently used single number proxies for uncertainty**
7. **NHS seems largely unaccountable, and can lean on the independence of the Trusts and GPs, as well as patient privacy in order to avoid meaningful accountability**

Perfect 7 for 7 – NHS Cyber Assessments Are Clearly Risky

Summary

Risk Checklist Comparison

	EIA	Met Office	GAO Protests	GC LTC	NHS Cyber
Relying on repetition	No	No	Yes	Yes	Yes
Focused on budgets, revenues, heft?	No?	Maybe	No	Yes	Yes
Using Analysis Pros	Yes	Yes	No	Sort of	No?
Frozen Methods	Sort of	No	Yes	Yes	Yes
Trusting Formality	No	No	Yes	Yes	Yes
Deterministic?	Sort of	No	Yes	Yes	Yes
Accountable?	Yes	Yes	No	No	No

Risk Analytics are… Risky

Three flawed risk assessments

1. Protest policy has been flawed, and at best wasted a taxpayer money
2. Long Term Care insurance will cost stockholders well over $25 Billion dollars and some seniors will not have insurance they paid for
3. NHS probably killed someone with bad cyber policy

The Analytics Benchmarking Risk Checklist quickly warns how risky your analysis is… including the risk of bad risk analysis