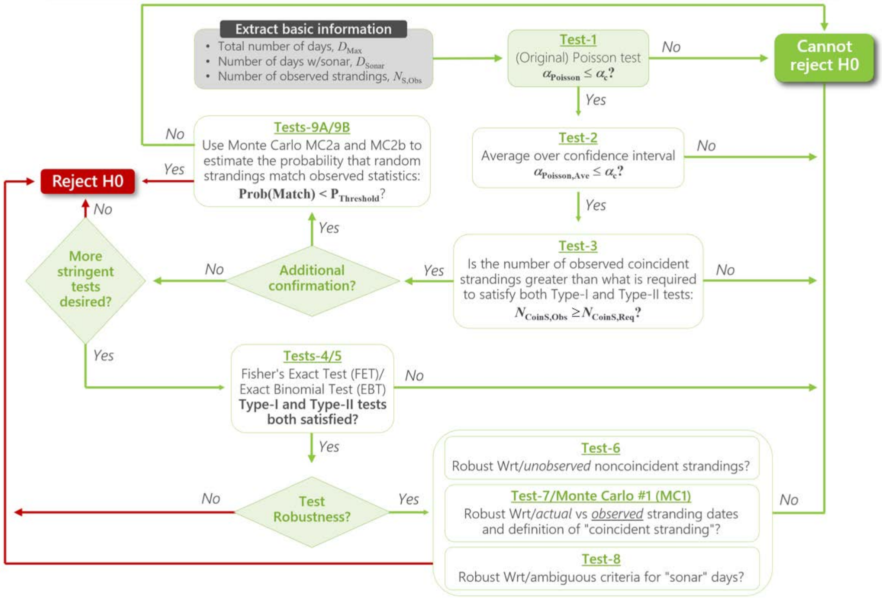
March 2025



Extract basic information
• Total number of days, D_{Max}
• Number of days w/sonar, D_{Sonar}
• Number of observed strandings, $N_{S,Obs}$
Test-1
(Original) Poisson test
$\alpha_{Poisson} \le \alpha_c$?
No
Cannot reject H0
Yes
No
Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
Prob(Match) < $P_{Threshold}$?
Yes
Reject H0
No
Test-2
Average over confidence interval
$\alpha_{Poisson,Ave} \le \alpha_c$?
No
Yes
Yes
More stringent tests desired?
No
Additional confirmation?
Yes
Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{CoinS,Obs} \ge N_{CoinS,Req}$?
No
Yes
Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?
No
Yes
Test-6
Robust Wrt/unobserved noncoincident strandings?
No
Test Robustness?
Yes
Test-7/Monte Carlo #1 (MC1)
Robust Wrt/actual vs observed stranding dates and definition of "coincident stranding"?
No
Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

Statistical Analysis of Marine Mammal Stranding Events

TOWARD A STRANDING CORRELATION ANALYSIS PLAYBOOK

Andrew Ilachinski and Ronald Filadelfo

DAB-2025-U-040942-2Rev

Abstract

Navy operations, training, and testing at sea (most notably active sonar) can potentially harm marine mammals and lead to stranding events under certain circumstances. However, strandings also occur because of natural reasons, so when strandings occur, it is difficult to determine whether the event was caused by Navy sonars (or any other human activity) or was a routine natural event. The Navy is thus often challenged by non-governmental organizations (NGOs) and federal agencies on the issue of active sonars harming marine mammals. The goal of the study was not to determine whether marine mammal stranding events and sonar are causally connected. No statistics-based method by itself could definitively answer such an open research challenge. Rather, the goal was to develop an approach for determining whether a given set of strandings (bounded in space and time) are correlated with sonar use. Toward this end, we refined existing methodology and developed a battery of new statistical tests that can be used to mutually confirm independent inferences. We also developed methods that take into account uncertainties in the underlying data. The main result of this study is the Stranding Correlation Analysis Playbook (SCAP), which is a visual flowchart for drawing inferences at varying levels of granularity and specificity. SCAP users may trace multiple pathways through the logic, depending on individual preferences and requirements.

CNA's annotated briefings are either condensed presentations of the results of formal CNA studies that have been further documented elsewhere or stand-alone presentations of research reviewed and endorsed by CNA. These briefings represent the best opinion of CNA at the time of issue. The views, opinions, and findings contained in this report should not be construed as representing the official position of the Department of the Navy.

This work was created in the performance of Federal Government Contract Number N00014-22-D-7001.

Cover image: CNA.

Approved by: **March 2025**

William Komiss

William Komiss, Research Program Director
Energy, Infrastructure, and Environment Program
Resources and Force Readiness Division

Statistical Analysis of Marine Mammal Stranding Events: Executive Summary

Navy operations, training, and testing at sea—most notably active sonar—can potentially harm marine mammals and lead to stranding events. However, it is difficult to determine whether a stranding was caused by Navy sonar (or any other human activity) or was a natural event. The Navy is thus often challenged by non-governmental organizations and federal agencies on the issue of active sonar harming marine mammals.

Previous analysis of stranding events

Following a stranding event, researchers often examine time-space correlations between Navy sonar use and the stranding in the area of interest. However, there is no universally agreed-upon methodology for conducting such studies. The goal of this project is to develop a statistically rigorous approach for inferring correlations (or lack thereof) between strandings and sonar that respects the inherent limitations and uncertainties of the available data.

Previous analyses have been limited by two statistical shortfalls:

- ***Shortfall #1: Sole reliance on the null stranding rate.*** The *observed* null stranding rate (estimated by dividing the number of observed strandings on days without sonar by the total number of days without sonar) is typically used as the de facto average null stranding rate. However, if we assume that stranding events are distributed according to an underlying Poisson random process, the observed number of strandings represents only a single sample in a statistical distribution, the true average of which may be any number that lies within a range of numbers (called the confidence interval).
- ***Shortfall #2: Ignoring the probability of a false negative (Type II) error.*** Inferences are drawn on the basis of adjudicating only Type I (i.e., false positive) errors; however, doing so is insufficient because we must simultaneously minimize the probability of making Type II (i.e., false negative) errors. That is, we must also minimize the probability of erroneously accepting the null hypothesis (i.e., that strandings are uncorrelated with sonar) when it is actually false. Unfortunately, this test of power (as it is called) is seldom, if ever, performed.

In developing a more rigorous approach, including mitigating these statistical shortfalls, we have both refined existing methodology and developed a battery of new statistical tests.

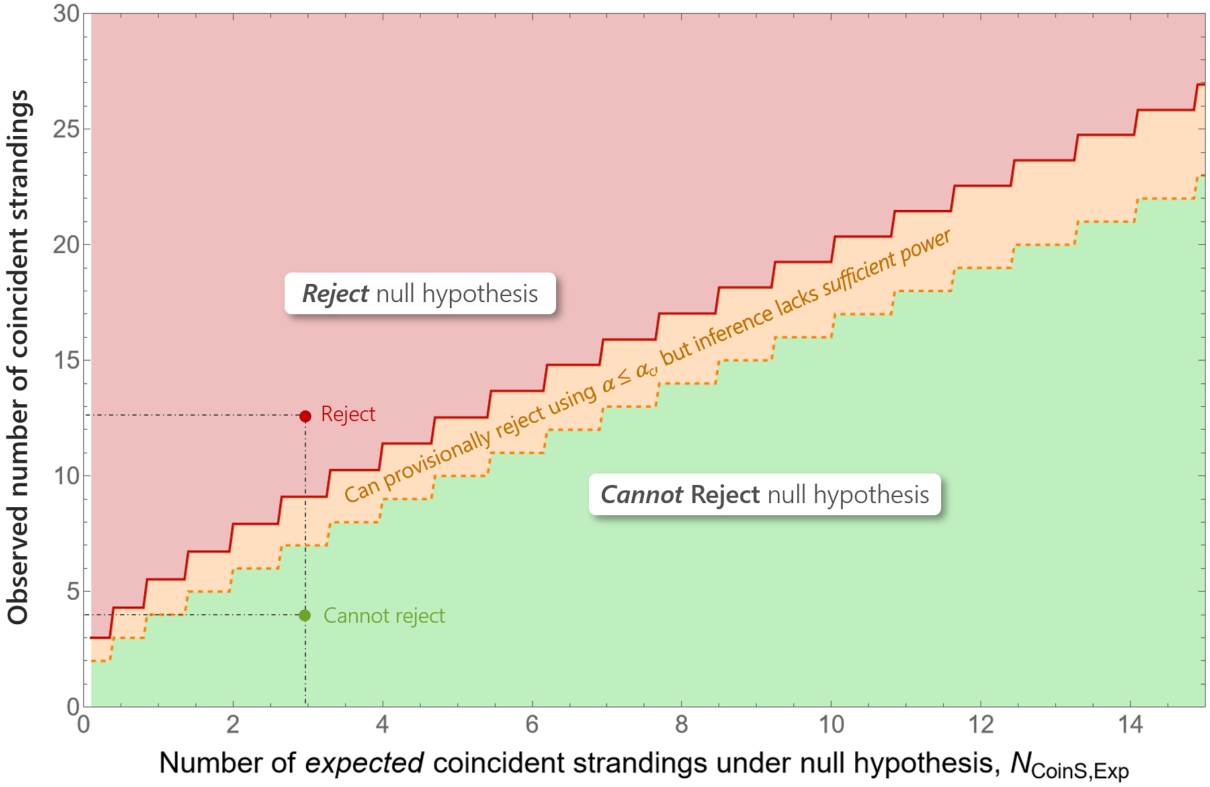
A more rigorous method

We have developed a method that effectively administers both Type I and Type II tests simultaneously. Rather than rejecting the null hypothesis based on a single means test of significance (or P-value), the decision to reject the null hypothesis follows only if the number of observed coincident strandings is greater than the minimum number required to *simultaneously satisfy both Type I and Type II tests.* This approach imposes a more stringent set of conditions that must be satisfied to reject the null hypothesis and is therefore a statistically stronger test to apply. Because it is stronger, we can generally expect fewer stranding events to be statistically correlated (i.e., coincident) with sonar than are found through existing methods that test for only Test I errors.

New tool for analysis #1: the Accept/Reject Criteria Chart

We developed a chart (Figure 1) that identifies the statistical inference entailed by given numbers of *expected* coincident strandings (under the null hypothesis) and *observed* coincident strandings.

Figure 1. Accept/Reject Criteria Chart



Source: CNA.

The chart maps each pair of values (expected and observed) to one of three inferential results:

1. *Reject the null hypothesis*. The statistics pass both Type I and Type II error tests, so a statistical correlation exists. (The red area in Figure 1.)
2. *Cannot reject the null hypothesis*. Because the P-value exceeds the desired minimum, no statistical correlation exists. (The green area in Figure 1.)
3. *Provisionally reject the null hypothesis*. The statistics pass the Type I test but not the Type II test; the determination of significance therefore lacks sufficient statistical power. (The yellow area in Figure 1).

Because this chart can be precomputed to accommodate a large set of possible real-world scenarios, stakeholders can use it to identify—at a glance—scenarios in which the statistical evidence to reject or not reject the null hypothesis is strong enough to warrant drawing an immediate inference. Stakeholders can then distinguish these scenarios from others that require additional analysis (specifically, those in the "provisionally reject the null hypothesis" class).

New tool for analysis #2: the Stranding Correlation Analysis Playbook

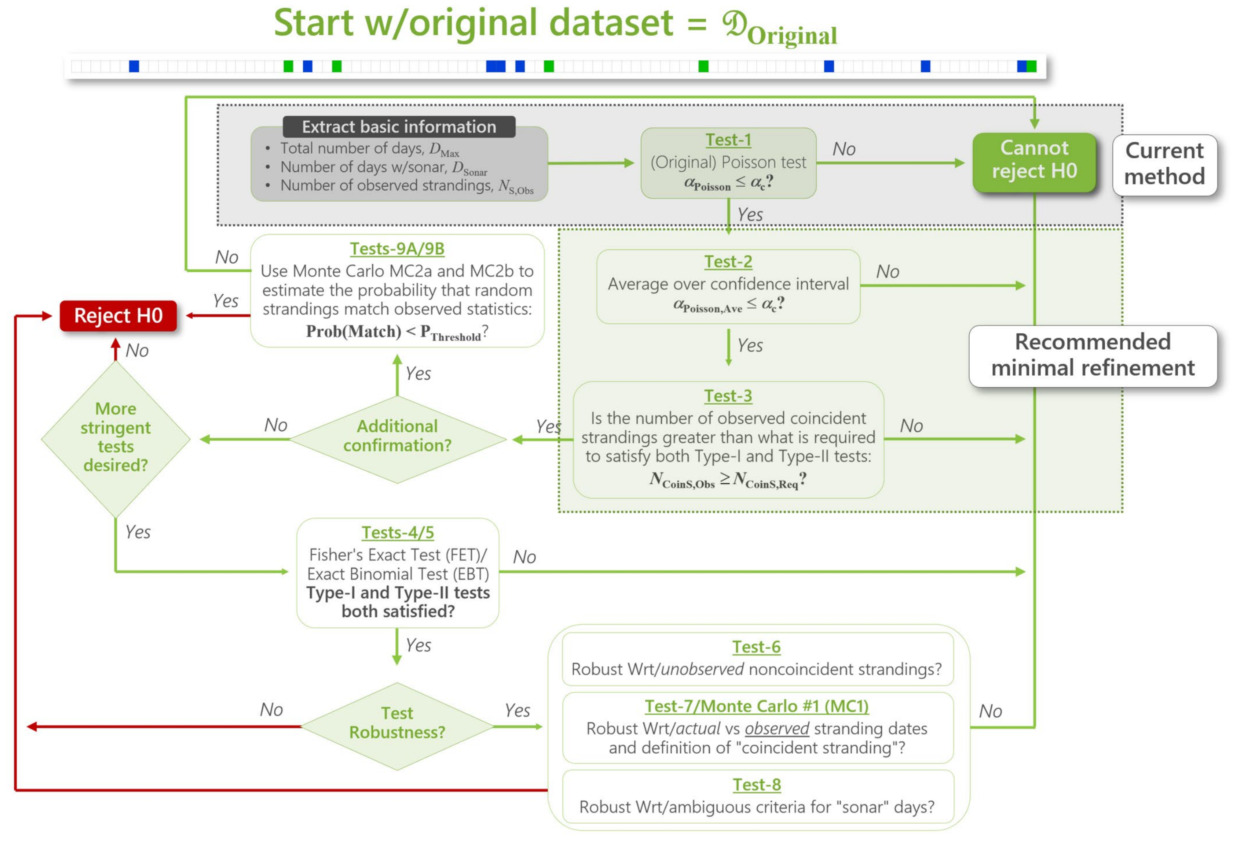
The underlying data contain several sources of uncertainty, including the following:

1. Whether a given stranding event is coincident with sonar.
2. The actual stranding date (which may be different from the observed stranding date).
3. The possibility that a given area of operations may include other unreported strandings.
4. The presence of non-Navy sonar.

To mitigate these sources of uncertainty, we developed a set of Monte Carlo simulations. Our report also introduces, describes, and provides illustrative use cases for 10 statistical tests and analysis tools, most of which we developed specifically for this study.

However, the proper application of these tests requires a threshold level of knowledge regarding mathematical and statistical modeling methods. Therefore, some stakeholders may not immediately recognize which tests are best suited for a given scenario or whether a given test is necessary (or even applicable). To mitigate this challenge, we developed the Stranding Correlation Analysis Playbook (SCAP) as the main result of this study (Figure 2).

Figure 2. Stranding Correlation Analysis Playbook



Start w/original dataset = $\mathcal{D}_{\text{Original}}$
Extract basic information
• Total number of days, D_{Max}
• Number of days w/sonar, D_{Sonar}
• Number of observed strandings, $N_{\text{S,Obs}}$
Test-1
(Original) Poisson test
$\alpha_{\text{Poisson}} \leq \alpha_c$?
No
Yes
Cannot reject H0
Current method
Test-2
Average over confidence interval
$\alpha_{\text{Poisson,Ave}} \leq \alpha_c$?
No
Yes
Recommended minimal refinement
Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\text{CoinS,Obs}} \geq N_{\text{CoinS,Req}}$?
No
Yes
Additional confirmation?
Yes
No
Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\text{Prob(Match)} < \text{P}_{\text{Threshold}}$?
No
Yes
Reject H0
No
More stringent tests desired?
Yes
Tests-4/5
Fisher's Exact Test (FET)/ Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?
No
Yes
Test Robustness?
No
Yes
Test-6
Robust Wrt/unobserved noncoincident strandings?
Test-7/Monte Carlo #1 (MC1)
Robust Wrt/actual vs observed stranding dates and definition of "coincident stranding"?
Test-8
Robust Wrt/ambiguous criteria for "sonar" days?
No

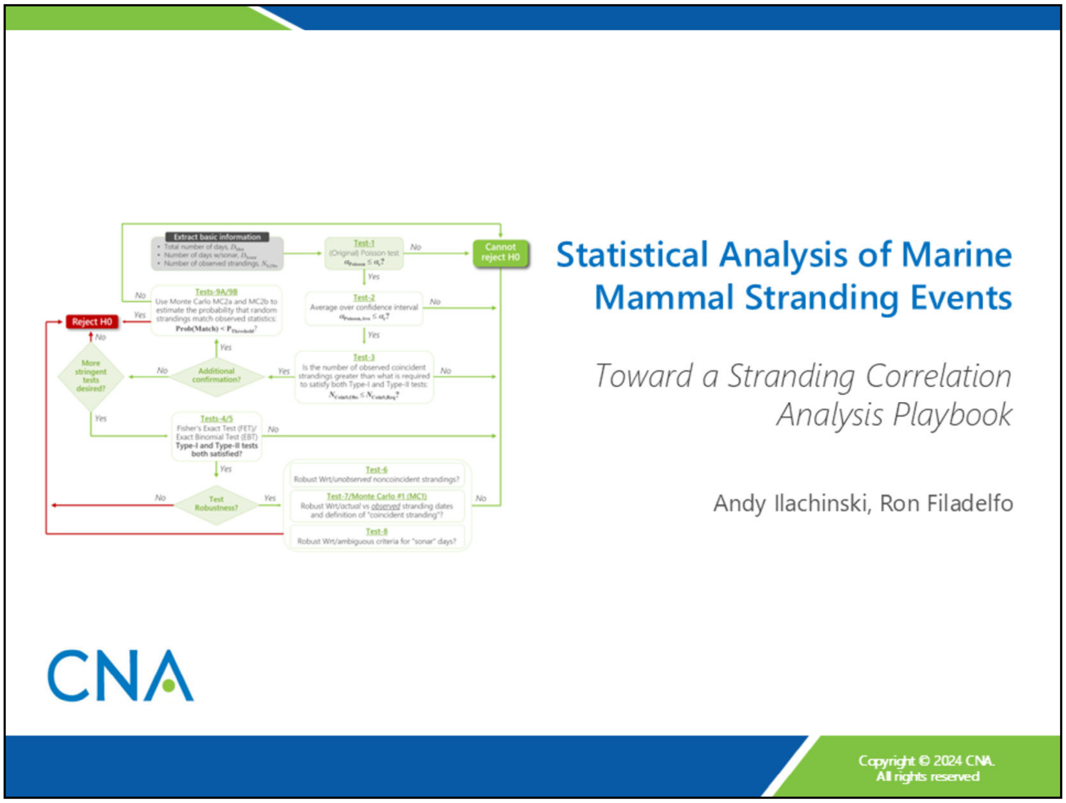
Source: CNA.

The SCAP is a flowchart that analysts, stakeholders, and decision-makers can use to navigate the myriad options of the inferential process at various levels of granularity and specificity—a process that culminates in a decision to "Reject" or "Cannot Reject" the null hypothesis.

The SCAP flowchart weaves together five increasingly refined inferential pathways through the battery of statistical tests and analysis tools described in this report.

- The first and shortest pathway (highlighted in gray at the top of Figure 2) denotes the current method (Test 1), which typically consists of administering only a single test for significance.
- The second pathway (highlighted in the green box) adds two tests to strengthen the veracity of whatever final inference is drawn, and it simultaneously accounts for both Type I (false positive) and Type II (false negative) errors, as described earlier. **This is the pathway we recommend, at a minimum, as an immediate and rigorous refinement of the current method.**

For additional confirmation or to administer more stringent tests that better account for uncertainties in the data, other optional pathways are illustrated in Figure 2 and are described in the report.


Extract basic information
• Total number of days, D_{Max}
• Number of days w/sonar, D_{Sonar}
• Number of observed strandings, $N_{S,Obs}$
Test-1
(Original) Poisson test
$\alpha_{Poisson} \le \alpha_c$?
No
Cannot reject H0
Yes
Test-2
Average over confidence interval
$\alpha_{Poisson,Ave} \le \alpha_c$?
No
Yes
Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{CoinS,Obs} \le N_{CoinS,Req}$?
No
Yes
Additional confirmation?
Yes
Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
Prob(Match) < $P_{Threshold}$?
No
Yes
Reject H0
No
More stringent tests desired?
No
Yes
Tests-4/5
Fisher's Exact Test (FET)/ Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?
No
Yes
Test Robustness?
No
Yes
Test-6
Robust Wrt/unobserved noncoincident strandings?
Test-7/Monte Carlo #1 (MC1)
Robust Wrt/actual vs observed stranding dates and definition of "coincident stranding"?
Test-8
Robust Wrt/ambiguous criteria for "sonar" days?
No

Statistical Analysis of Marine Mammal Stranding Events

Toward a Stranding Correlation Analysis Playbook

Andy Ilachinski, Ron Filadelfo

CNA

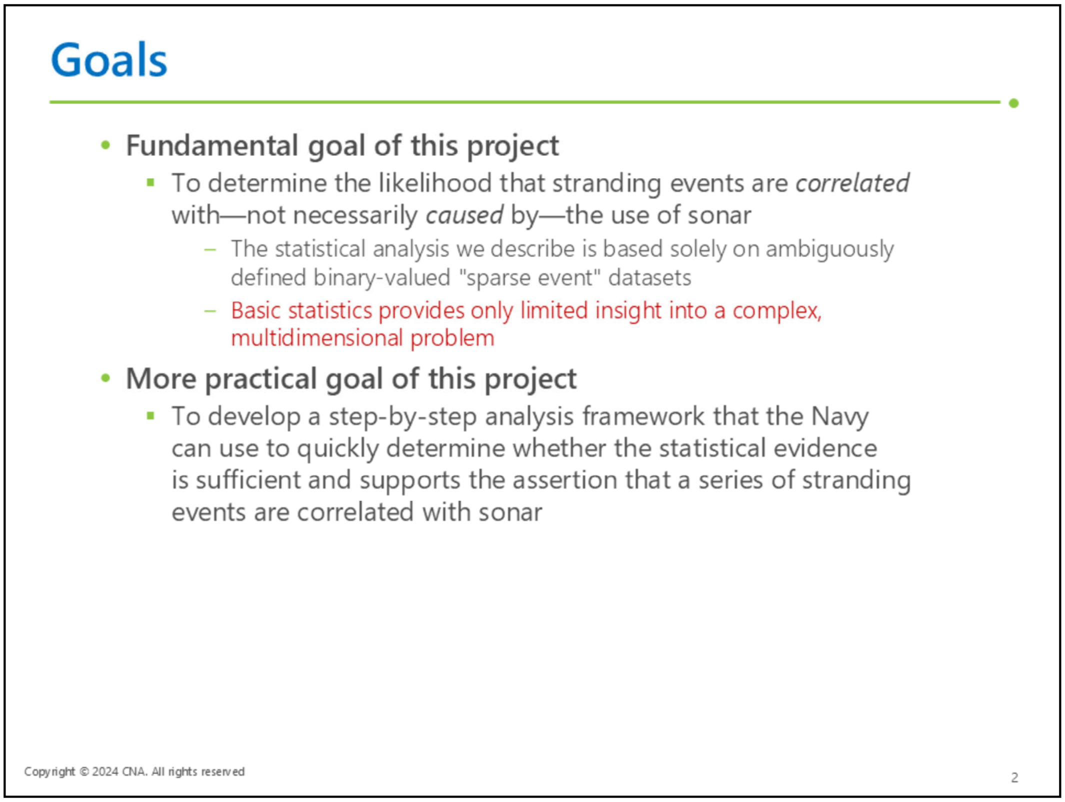
Navy operations, training, and testing at sea—most notably active sonar—can potentially harm marine mammals and lead to stranding events under certain circumstances. However, strandings also occur because of natural reasons, so when strandings occur, it is difficult to determine whether the event was caused by Navy sonars (or any other human activity) or was a routine natural event. The Navy is thus often challenged by non-governmental organizations (NGOs) and federal agencies on the issue of active sonars harming marine mammals.

This project seeks to build a framework for analyzing future stranding events to determine whether evidence suggests they are statistically correlated with Navy sonar.

Following stranding events, studies are often performed that seek to examine time-space correlations between Navy sonar use and whale strandings in the particular area of interest. These studies are subject to many mathematical pitfalls, including limited observations, possible observational bias, and a great deal of uncertainty in the data supporting them.

To address this issue, the Navy, CNA analysts, and National Oceanic and Atmospheric Administration (NOAA) scientists agreed on the need to develop a rigorous, standard methodology for these types of studies, given how important they are in regulatory decision-making and their implications for the Navy being able to train at sea.

This study represents the first step toward developing this methodology.



Goals

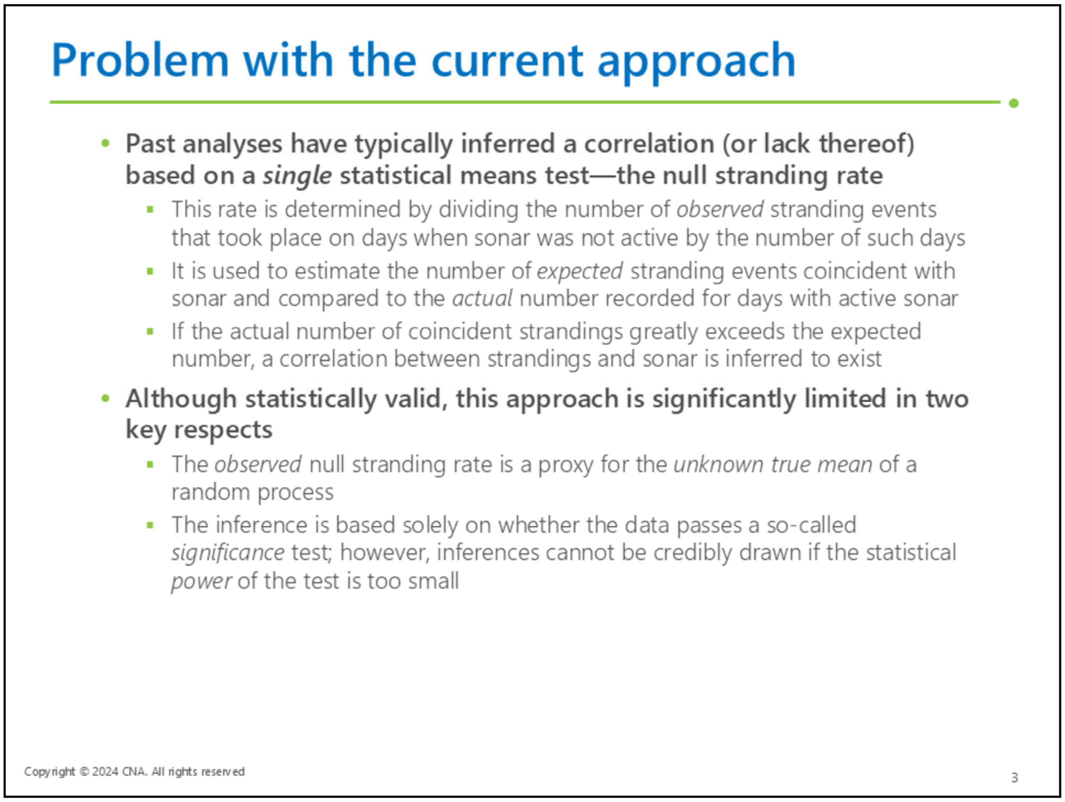
- **Fundamental goal of this project**
  - To determine the likelihood that stranding events are *correlated* with—not necessarily *caused* by—the use of sonar
    - The statistical analysis we describe is based solely on ambiguously defined binary-valued "sparse event" datasets
    - Basic statistics provides only limited insight into a complex, multidimensional problem
- **More practical goal of this project**
  - To develop a step-by-step analysis framework that the Navy can use to quickly determine whether the statistical evidence is sufficient and supports the assertion that a series of stranding events are correlated with sonar

2

The fundamental goal of the study was *not* to determine whether marine mammal stranding events and sonar are causally connected. No statistics-based method by itself could definitively answer such an open research challenge. Rather, the goal was to develop an approach for determining whether a given set of strandings (bounded in space and time) are *correlated* with sonar use. Correlation represents a far weaker relationship that does not immediately imply causation, although it is often erroneously assumed to do so (if only for expediency) and equally rarely acknowledged.

Saliently, statistics-based correlative methods are inherently limited because of the nature and paucity of data related to stranding events. When available, the data often consist of only three or four pieces of basic information, such as the following: (1) the time stranding events were observed (not necessarily when they occurred), (2) their location, (3) the state of decay of an animal when it was found (which is reported only rarely), and (4) the presence or absence of sonar (within some set of space-time coordinates that overlap with strandings).

At a basic level, the data consist of two strings of time-stamped binary-valued elements. In the first string, each date is assigned a one if a stranding was observed and a zero if not; in the second string, each date is assigned a one if sonar was active and a zero if not. The analytical goal is to determine whether these two binary-valued strings are correlated beyond mere chance.



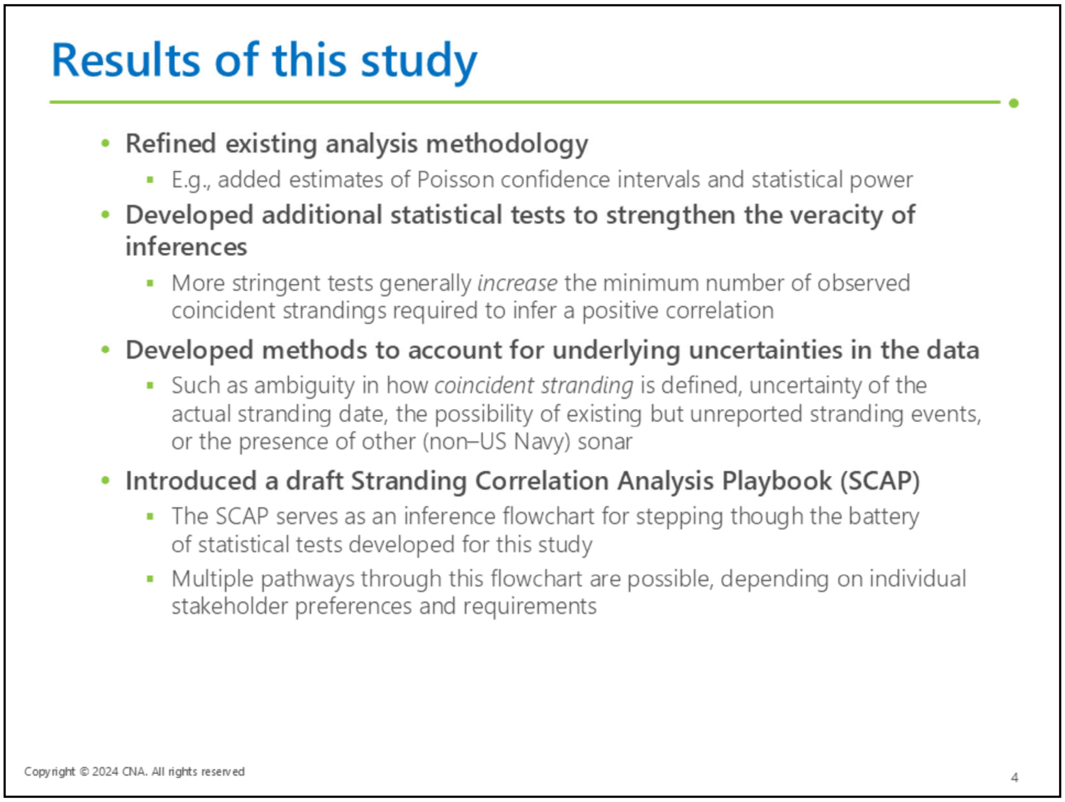
Problem with the current approach

- **Past analyses have typically inferred a correlation (or lack thereof) based on a *single* statistical means test—the null stranding rate**
  - This rate is determined by dividing the number of *observed* stranding events that took place on days when sonar was not active by the number of such days
  - It is used to estimate the number of *expected* stranding events coincident with sonar and compared to the *actual* number recorded for days with active sonar
  - If the actual number of coincident strandings greatly exceeds the expected number, a correlation between strandings and sonar is inferred to exist
- **Although statistically valid, this approach is significantly limited in two key respects**
  - The *observed* null stranding rate is a proxy for the *unknown true mean* of a random process
  - The inference is based solely on whether the data passes a so-called *significance* test; however, inferences cannot be credibly drawn if the statistical *power* of the test is too small

Moreover, virtually all past sonar-stranding correlation analyses have used a single statistical test—the null stranding rate—to determine whether the number of *observed* strandings coincident with sonar exceeds the *expected* number of coincident strandings (as estimated by the null stranding rate). Although this approach is prima facie valid, it is also deficient in two ways:

- The null stranding rate is based on a *single number*—the number of observed strandings that took place on days when sonar was not active; as a result, the observed number of null strandings is the de facto proxy for the unknown *true mean* of a random process.
- The inference itself is based solely on comparing the observed number of coincident strandings to what is expected using the proxy-based null stranding rate. A correlation between strandings and sonar is inferred if these two numbers are significantly different, as determined by the magnitude of the resulting P-value (defined on a later slide). But, statistical inferences cannot credibly be drawn on the basis of significance alone; they also require a test of *power*, which determines whether the data actually warrant making any inferences at all. If the power is too low, a correlation between two random processes cannot be inferred to exist, even if the P-value suggests that it does.

Later slides show how to mitigate both of these limitations and how these mitigations can be used to develop more powerful methods to test for correlations.



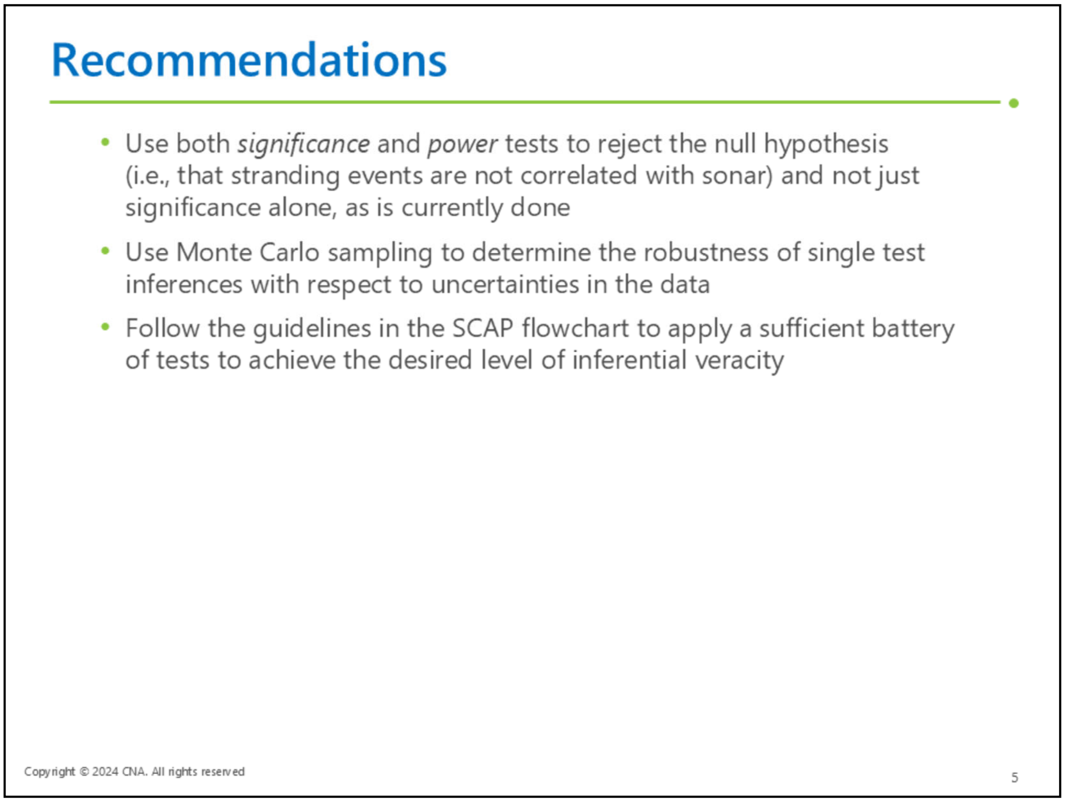
Results of this study

- **Refined existing analysis methodology**
  - E.g., added estimates of Poisson confidence intervals and statistical power
- **Developed additional statistical tests to strengthen the veracity of inferences**
  - More stringent tests generally *increase* the minimum number of observed coincident strandings required to infer a positive correlation
- **Developed methods to account for underlying uncertainties in the data**
  - Such as ambiguity in how *coincident stranding* is defined, uncertainty of the actual stranding date, the possibility of existing but unreported stranding events, or the presence of other (non–US Navy) sonar
- **Introduced a draft Stranding Correlation Analysis Playbook (SCAP)**
  - The SCAP serves as an inference flowchart for stepping though the battery of statistical tests developed for this study
  - Multiple pathways through this flowchart are possible, depending on individual stakeholder preferences and requirements

For this project, we refined existing methodology and developed a battery of new statistical tests that can be used to mutually confirm independent inferences. We also developed methods that take into account uncertainties in the underlying data, including the following:

1. The ambiguous definition of a *coincident stranding* (typically, a stranding is labeled "coincident" if it occurs within six days and 60 nmi of the last sonar use)
2. The uncertainty of the actual stranding date, which must be extrapolated from when the stranding was observed
3. The possibility that a given area of operations may include other unreported strandings
4. The presence of other non-US sonar (along with the more general ambiguity of specifying the requisite set of sonar events that may be correlated with strandings)

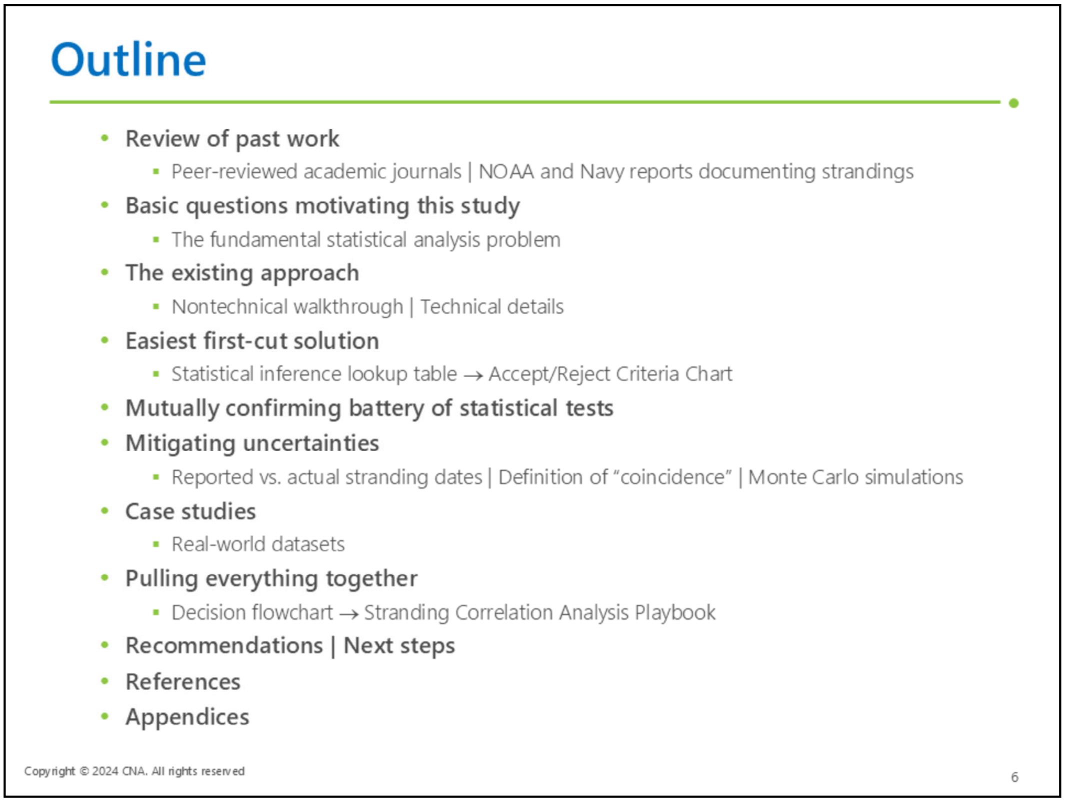
As the main result of the study, we developed a draft Stranding Correlation Analysis Playbook (SCAP), which is a visual flowchart for drawing inferences at varying levels of granularity and specificity. SCAP users may trace multiple pathways through the logic, depending on individual preferences and requirements. We provide the SCAP at the end of the slide deck, but we first introduce a suite of statistical tests and Monte Carlo simulations (MCS) as necessary context.



Recommendations

- Use both *significance* and *power* tests to reject the null hypothesis (i.e., that stranding events are not correlated with sonar) and not just significance alone, as is currently done
- Use Monte Carlo sampling to determine the robustness of single test inferences with respect to uncertainties in the data
- Follow the guidelines in the SCAP flowchart to apply a sufficient battery of tests to achieve the desired level of inferential veracity

Our overall recommendation is to use both P-Value and power tests to reject the null hypothesis (i.e., that stranding events are not correlated with sonar), and not just significance alone, as is current practice.



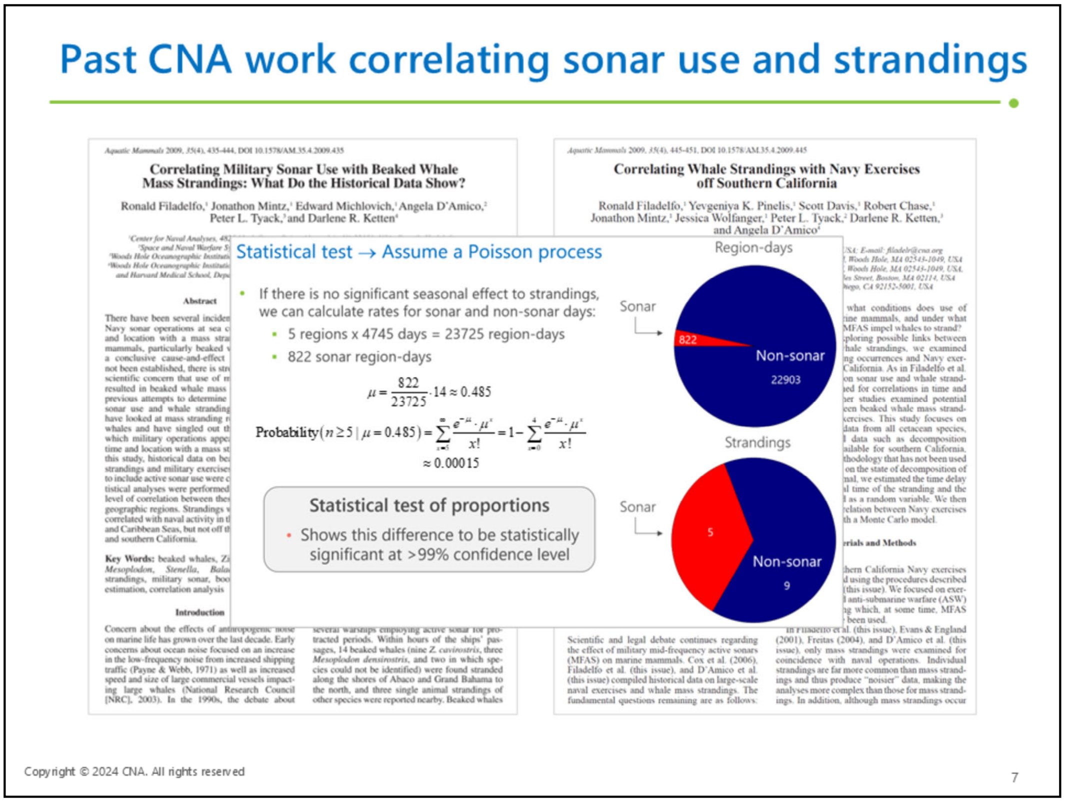
Outline

- Review of past work
  - Peer-reviewed academic journals | NOAA and Navy reports documenting strandings
- Basic questions motivating this study
  - The fundamental statistical analysis problem
- The existing approach
  - Nontechnical walkthrough | Technical details
- Easiest first-cut solution
  - Statistical inference lookup table → Accept/Reject Criteria Chart
- Mutually confirming battery of statistical tests
- Mitigating uncertainties
  - Reported vs. actual stranding dates | Definition of "coincidence" | Monte Carlo simulations
- Case studies
  - Real-world datasets
- Pulling everything together
  - Decision flowchart → Stranding Correlation Analysis Playbook
- Recommendations | Next steps
- References
- Appendices

This slide presents the outline of the slide deck.

Because the subject matter is inherently technical, the slides have been designed with two organizing principles in mind: (1) those parts of the narrative that may otherwise come across as excessively technical by certain readers are introduced by a nontechnical summary, and (2) the exposition of the most technically detailed slides relies more on simple visualizations of concepts that can be understood quickly and intuitively, rather than on traditional bullet-ridden, text-based explanations (which are mostly relegated to the supporting notes section of those slides).

The unannotated slide deck is included at the end of this document to allow the sponsor to use these slides in discussions with regulators and others, and to allow better readability when zooming in to see some of the mathematical detail.



Past CNA work correlating sonar use and strandings

Aquatic Mammals 2009, 35(4), 435-444, DOI 10.1578/AM.35.4.2009.435

Correlating Military Sonar Use with Beaked Whale Mass Strandings: What Do the Historical Data Show?

Ronald Filadelfo,[1] Jonathon Mintz,[1] Edward Michlovich,[1] Angela D'Amico,[2] Peter L. Tyack,[3] and Darlene R. Ketten[4]

Aquatic Mammals 2009, 35(4), 445-451, DOI 10.1578/AM.35.4.2009.445

Correlating Whale Strandings with Navy Exercises off Southern California

Ronald Filadelfo,[1] Yevgeniya K. Pinelis,[1] Scott Davis,[1] Robert Chase,[1] Jonathon Mintz,[1] Jessica Wolfanger,[1] Peter L. Tyack,[2] Darlene R. Ketten,[3] and Angela D'Amico[4]

Statistical test → Assume a Poisson process

- If there is no significant seasonal effect to strandings, we can calculate rates for sonar and non-sonar days:
  - 5 regions x 4745 days = 23725 region-days
  - 822 sonar region-days

$$\mu = \frac{822}{23725} \cdot 14 \approx 0.485$$

$$\text{Probability}(n \geq 5 \mid \mu = 0.485) = \sum_{x=5}^{\infty} \frac{e^{-\mu} \cdot \mu^{x}}{x!} = 1 - \sum_{x=0}^{4} \frac{e^{-\mu} \cdot \mu^{x}}{x!}$$
$$\approx 0.00015$$

Statistical test of proportions

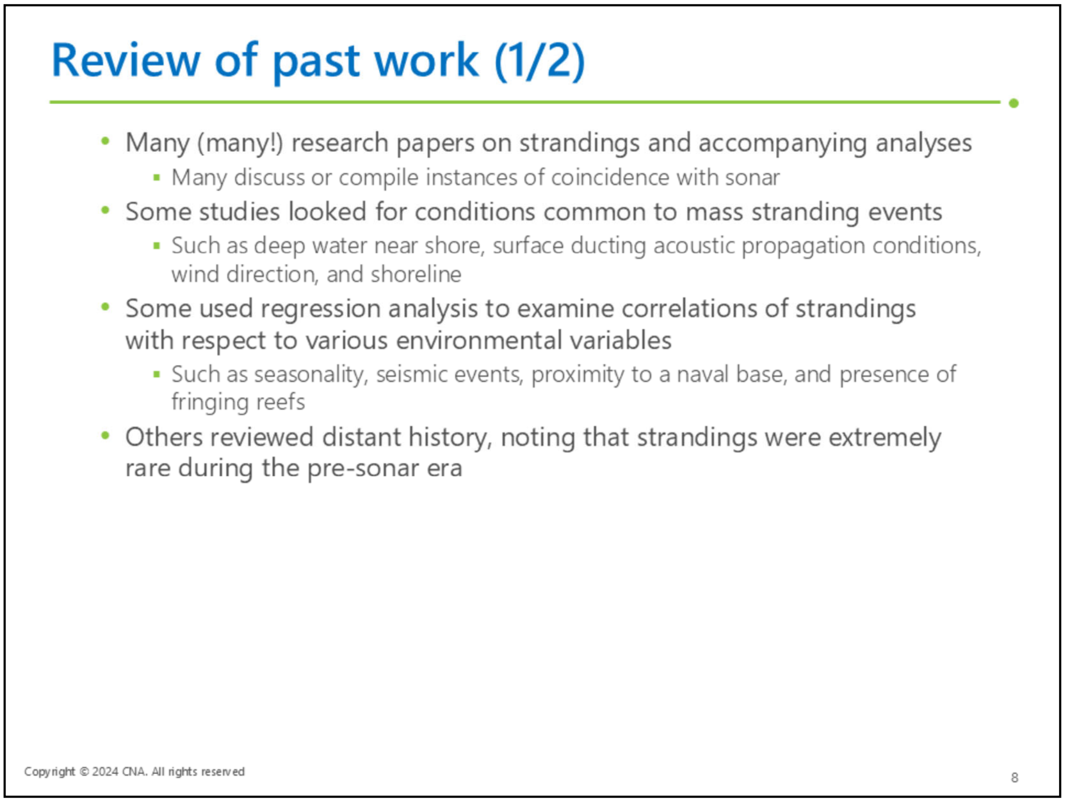
- Shows this difference to be statistically significant at >99% confidence level

7

When we began studying the correlation between naval sonar operations and whale strandings several years ago in studies performed for OPNAV, the Navy encouraged us to share our past work on sonar use and strandings with the academic research community. This slide shows the two academic journal articles we published [16,18].

Consider, for example, our analysis of Mediterranean beaked whale mass strandings, described in the article shown on the left. In addition to a bootstrap analysis, we performed a standard test of proportions on the difference in stranding rates between the times sonar activity was occurring and was not occurring. By dividing the Mediterranean into five regions, we obtained 23,725 (13 years x 365 days/year x 5 regions) region-days from 1992 to 2004. For the sonar periods, we observed five beaked whale mass strandings during the 822 region-days of sonar activity. For the non-sonar periods, we observed nine beaked whale mass strandings during the 22,903 region-days of non-sonar activity. Thus, we found a much greater stranding rate during the sonar periods (5 / 822 > 9 / 22,903). The pie charts on the right show this difference graphically: the fraction of beaked whale mass strandings that occurred during sonar periods was much greater than would be expected based on the fraction of time that sonar activity was occurring.

How significant is this difference in beaked whale mass stranding rates? A statistical test of proportions shows it to be significant at the 0.999 level, meaning there is less than a 1 in 1,000 chance that random (sampling) variability would have yielded a difference this big if there were no actual difference in the beaked whale mass stranding rates between the sonar and non-sonar periods.



Review of past work (1/2)

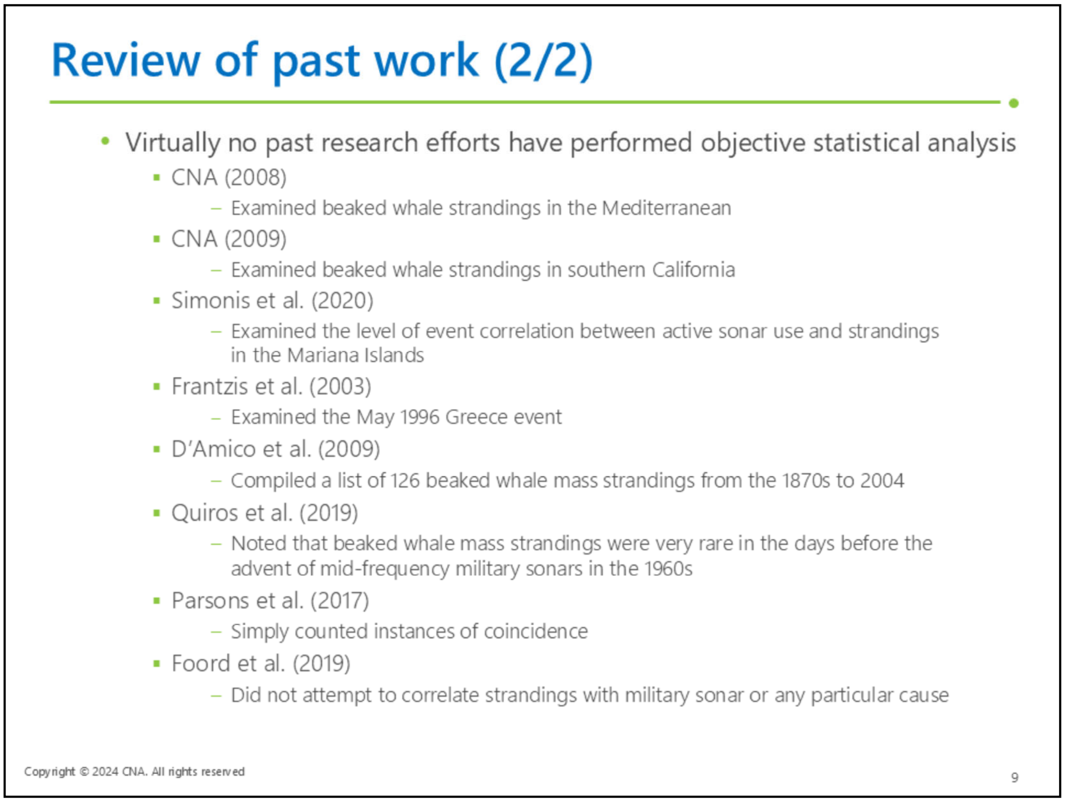
- Many (many!) research papers on strandings and accompanying analyses
  - Many discuss or compile instances of coincidence with sonar
- Some studies looked for conditions common to mass stranding events
  - Such as deep water near shore, surface ducting acoustic propagation conditions, wind direction, and shoreline
- Some used regression analysis to examine correlations of strandings with respect to various environmental variables
  - Such as seasonality, seismic events, proximity to a naval base, and presence of fringing reefs
- Others reviewed distant history, noting that strandings were extremely rare during the pre-sonar era

8

Scientific study of anthropogenic links to marine mammal strandings, particularly those potentially related to military active sonar, began in earnest following the stranding of beaked whales in Greece in May 1996. This stranding was coincident with the testing of low- and mid-frequency acoustic sources by the North Atlantic Treaty Organization (NATO) SACLANT ASW Research Center. Since that time, many research papers have been published on this subject.

This slide summarizes the various mathematical and statistical approaches of past research that sought to correlate military sonar and strandings. Overall, we found that few studies incorporated rigorous correlation analyses—probably because of the lack of robust data on strandings and on sonar use.

Several studies searched for conditions that seemed common to mass stranding events: deep water near shore [22, 29], surface ducting acoustic propagation conditions [29], wind direction and shoreline orientation [22, 28], seasonality [24], proximity to a naval base [26], and presence of fringing reefs [22].



Review of past work (2/2)

- Virtually no past research efforts have performed objective statistical analysis
  - CNA (2008)
    - Examined beaked whale strandings in the Mediterranean
  - CNA (2009)
    - Examined beaked whale strandings in southern California
  - Simonis et al. (2020)
    - Examined the level of event correlation between active sonar use and strandings in the Mariana Islands
  - Frantzis et al. (2003)
    - Examined the May 1996 Greece event
  - D’Amico et al. (2009)
    - Compiled a list of 126 beaked whale mass strandings from the 1870s to 2004
  - Quiros et al. (2019)
    - Noted that beaked whale mass strandings were very rare in the days before the advent of mid-frequency military sonars in the 1960s
  - Parsons et al. (2017)
    - Simply counted instances of coincidence
  - Foord et al. (2019)
    - Did not attempt to correlate strandings with military sonar or any particular cause

9

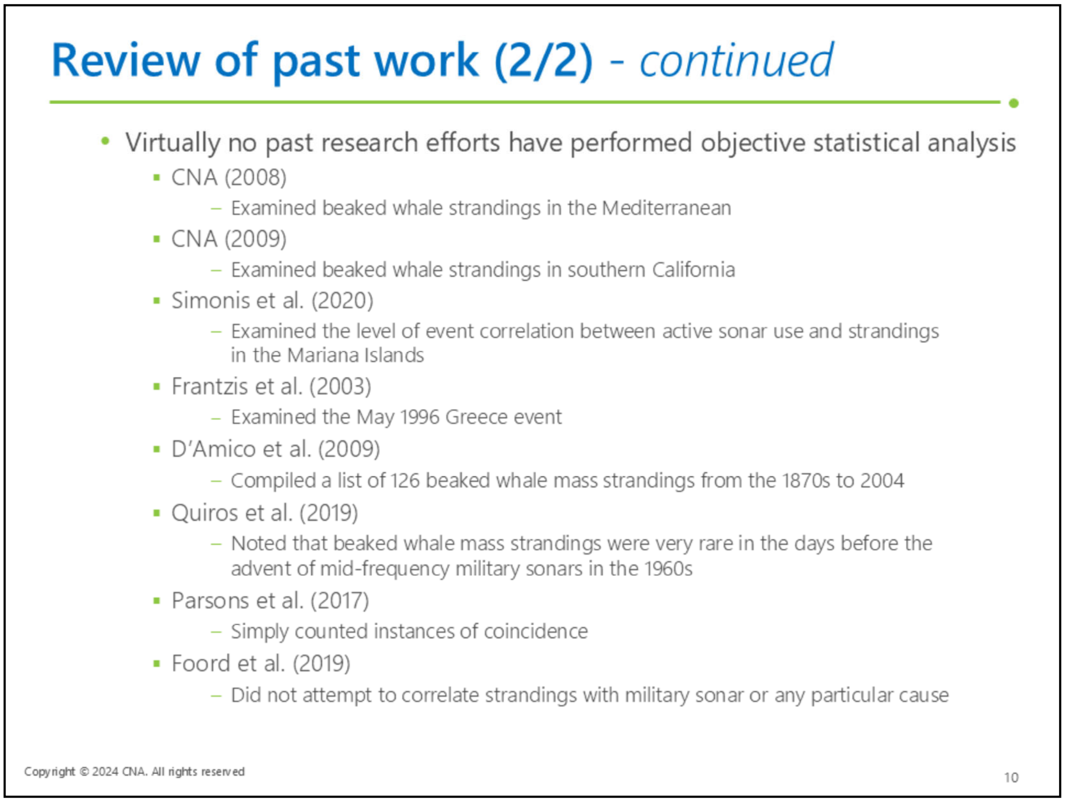
Virtually all studies have documented instances of time-space coincidence between sonar use and strandings, and some have documented injuries that could be consistent with acoustic trauma. These studies have made a compelling case for a sonar-stranding link, but few have performed objective statistical analyses that account for strandings that occur in the absence of sonar and for sonar use that results in no coincident strandings. Our 2005 examination of beaked whale strandings in the Mediterranean (published in 2009) [16] and our subsequent study for southern California in 2008 [17] were among the first to do this. More recently, Simonis et al. used our methods to examine the level of event correlation between active sonar use and strandings in the Mariana Islands [27].

In our literature search, we found the following studies that purportedly used a probabilistic approach:

- Frantzis, 2003 [1]: Frantzis examined the May 1996 Greece stranding event but somewhat arbitrarily selected a time period going back 16.5 years (6,026 days) before the stranding event, a period in which there were no beaked whale mass strandings in the Mediterranean. Frantzis noted that the sonar use occurred over a four-day period, so the odds of the stranding occurring by random chance during these four days was 4/6026, or less than 0.07 percent.

- D’Amico et al., 2009 [11]: D’Amico compiled a list of 126 beaked whale mass strandings from the 1870s to 2004, noting that the large majority of these occurred after mid-frequency military active sonars appeared in the 1950s. Given the incompleteness of the available data and the likely observational bias, D’Amico noted that no definitive quantitative statements concerning the level of sonar-stranding correlation could be made.

- Quiros et al., 2019 [25]: Like D’Amico, they compiled historical strandings from various open sources, noting that beaked whale mass strandings were very rare in the days before the advent of mid-frequency military sonars in the 1960s.



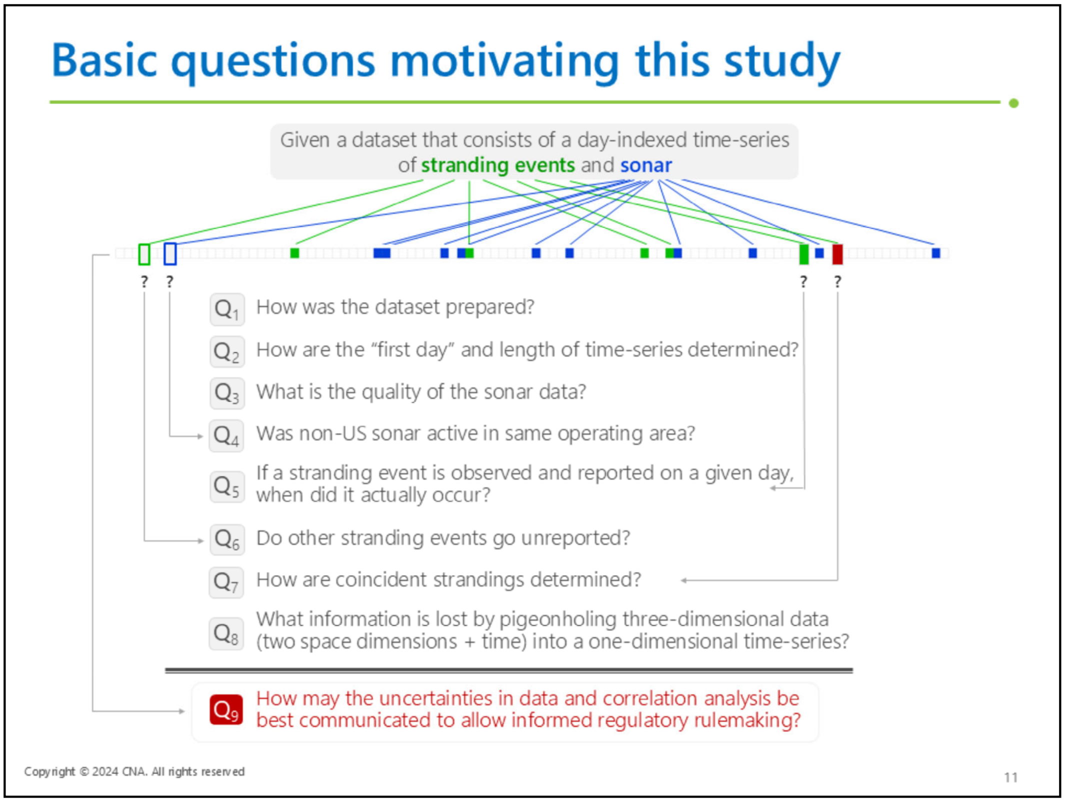
Review of past work (2/2) - *continued*

- Virtually no past research efforts have performed objective statistical analysis
  - CNA (2008)
    - Examined beaked whale strandings in the Mediterranean
  - CNA (2009)
    - Examined beaked whale strandings in southern California
  - Simonis et al. (2020)
    - Examined the level of event correlation between active sonar use and strandings in the Mariana Islands
  - Frantzis et al. (2003)
    - Examined the May 1996 Greece event
  - D'Amico et al. (2009)
    - Compiled a list of 126 beaked whale mass strandings from the 1870s to 2004
  - Quiros et al. (2019)
    - Noted that beaked whale mass strandings were very rare in the days before the advent of mid-frequency military sonars in the 1960s
  - Parsons et al. (2017)
    - Simply counted instances of coincidence
  - Foord et al. (2019)
    - Did not attempt to correlate strandings with military sonar or any particular cause

(Continued …)

- Parsons et al., 2017 [23]: They noted many instances of sonar-stranding coincidence and referenced our 2005 Mediterranean paper. However, in their discussion of correlation, they simply counted instances of coincidence. They also claimed that many more strandings occur without being observed, implying that they occur during sonar periods. However, if this claim is true, then there should also be many more unobserved strandings during non-sonar periods. They make a similar case in referencing the CNA Mediterranean paper, stating that the lack of good data on sonar use means there could very well have been more coincident strandings; however there could also have been many more non-stranding sonar events too.

- Foord et al., 2019 [19]: Although they performed statistical analyses to search for seasonal patterns of strandings in Australia, they did not attempt to correlate strandings with military sonar or any other any particular cause.



Basic questions motivating this study

Given a dataset that consists of a day-indexed time-series of **stranding events** and **sonar**

Q_1 How was the dataset prepared?

Q_2 How are the "first day" and length of time-series determined?

Q_3 What is the quality of the sonar data?

Q_4 Was non-US sonar active in same operating area?

Q_5 If a stranding event is observed and reported on a given day, when did it actually occur?

Q_6 Do other stranding events go unreported?

Q_7 How are coincident strandings determined?

Q_8 What information is lost by pigeonholing three-dimensional data (two space dimensions + time) into a one-dimensional time-series?

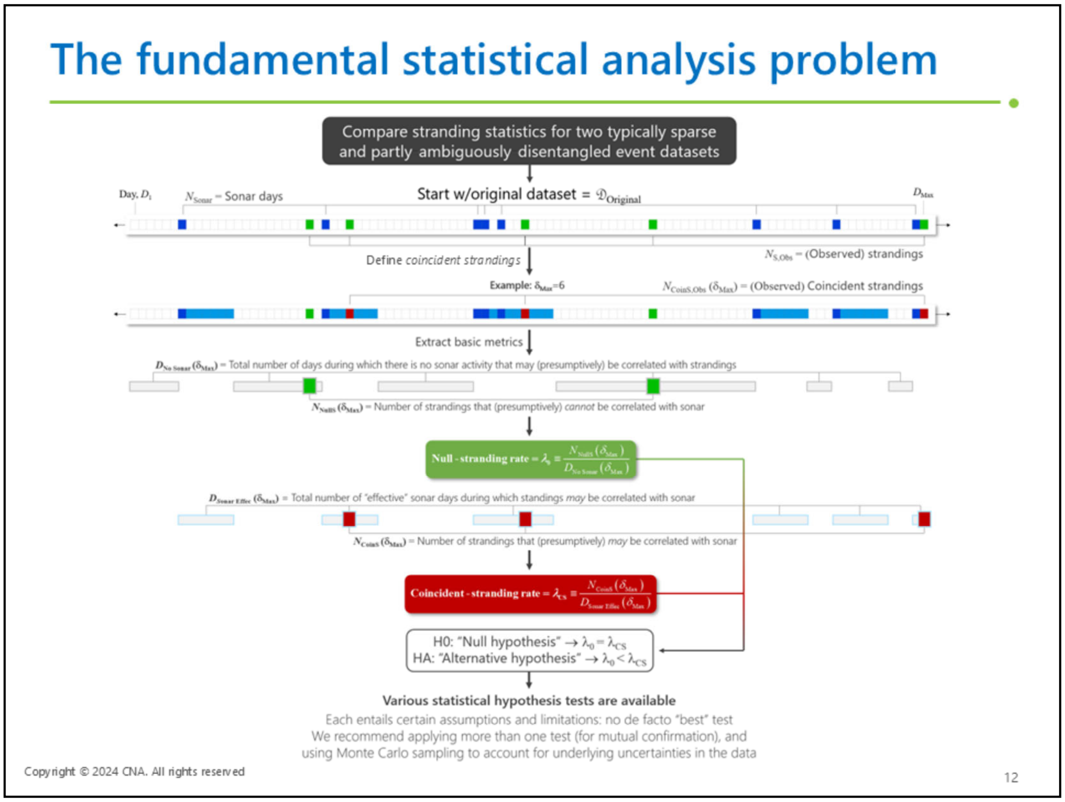
Q_9 How may the uncertainties in data and correlation analysis be best communicated to allow informed regulatory rulemaking?

11

Nine basic questions motivated and shaped this study, as shown on the slide.

To date, all statistical analyses of possible correlations between sonar use and whale strandings have been grounded on datasets in the form of a time-series of sonar activity and stranding events. These datasets are not always easy to acquire. Once such a dataset is created for an operating area of interest, the formal problem is to determine whether the number of strandings that are coincident with sonar is large enough compared to what is expected by chance to warrant inferring that strandings are correlated with sonar. However, the process of answering this problem is rife with ambiguities and uncertainties, as the long list of questions on this slide suggests (questions to which there are no immediately obvious answers).

We address each of these questions in this briefing, but the overarching goal of this study was to address the question highlighted in red. Specifically, we sought to develop a methodological framework to help inform regulatory rulemaking by explicitly accounting for and communicating the uncertainties and ambiguities inherent in all statistics-based analysis efforts to correlate sonar with strandings.



The fundamental statistical analysis problem

Compare stranding statistics for two typically sparse and partly ambiguously disentangled event datasets

Start w/original dataset = $\mathfrak{D}_{\text{Original}}$

Day, D_1
N_{Sonar} = Sonar days
D_{Max}
$N_{\text{S,Obs}}$ = (Observed) strandings

Define *coincident strandings*

Example: δ_{Max}=6
$N_{\text{CoinS,Obs}}$ (δ_{Max}) = (Observed) Coincident strandings

Extract basic metrics

$D_{\text{No Sonar}}$ (δ_{Max}) = Total number of days during which there is no sonar activity that may (presumptively) be correlated with strandings

N_{NullS} (δ_{Max}) = Number of strandings that (presumptively) *cannot* be correlated with sonar

$$\text{Null - stranding rate} = \lambda_0 \equiv \frac{N_{\text{NullS}}(\delta_{\text{Max}})}{D_{\text{No Sonar}}(\delta_{\text{Max}})}$$

$D_{\text{Sonar Effec}}$ (δ_{Max}) = Total number of "effective" sonar days during which standings *may* be correlated with sonar

N_{CoinS} (δ_{Max}) = Number of strandings that (presumptively) *may* be correlated with sonar

$$\text{Coincident - stranding rate} = \lambda_{\text{CS}} \equiv \frac{N_{\text{CoinS}}(\delta_{\text{Max}})}{D_{\text{Sonar Effec}}(\delta_{\text{Max}})}$$

H0: "Null hypothesis" → $\lambda_0 = \lambda_{\text{CS}}$
HA: "Alternative hypothesis" → $\lambda_0 < \lambda_{\text{CS}}$

Various statistical hypothesis tests are available

Each entails certain assumptions and limitations: no de facto "best" test
We recommend applying more than one test (for mutual confirmation), and
using Monte Carlo sampling to account for underlying uncertainties in the data

12

This slide illustrates how visual explanations may be used to convey ideas easily and intuitively that otherwise might be difficult to understand with a purely text-based pedagogy.

The underlying statistical problem is to compare two time-series of binary values: one consists of *stranding data* (in which each date is assigned a one if a stranding was observed and a zero if not), and the other consists of *sonar data* (in which each date is assigned a one if sonar was active and a zero if not). Specifically, the problem is to determine whether these two time-series are correlated beyond mere chance.

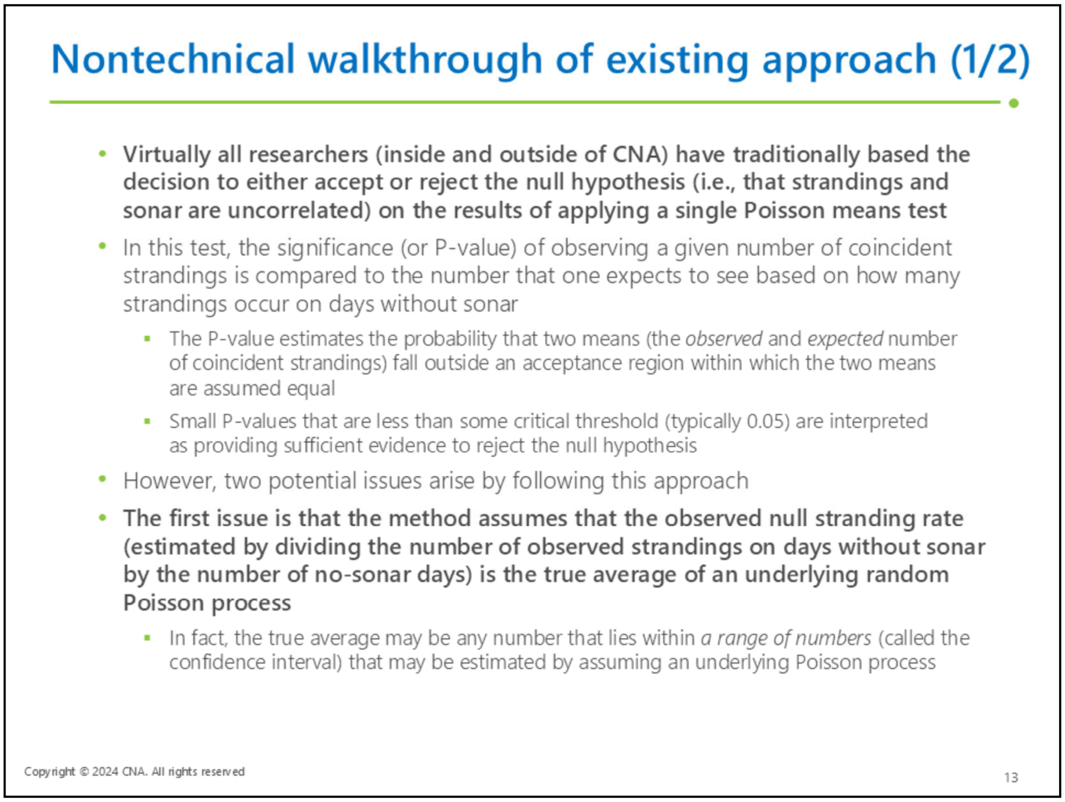
The basic parameters that must be extracted for the analysis are as follows: (1) D_{Max} = total number of days in the original dataset, (2) N_{Sonar} = total number of sonar days, and (3) $N_{\text{S,Obs}}$ = total number of observed strandings.

The null stranding rate, λ_0, is defined by dividing the number of stranding events observed on days when sonar was not active, $N_{\text{NullS}}(\delta_x)$, by the number of such days, $D_{\textit{No}\ \text{Sonar}}(\delta_x)$. The parameter δ_x is typically assigned the value of six days and represents the maximum number of days that can have passed since the last sonar day for a stranding to be considered coincident with sonar.

The expected coincident stranding rate, λ_{CS}, is defined by dividing the number of coincident stranding events, $N_{\text{CoinS}}(\delta_x)$, by the number of effective sonar days, $D_{\text{sonar Effec}}(\delta_x)$. The effective sonar days include days when sonar was active along with all days that were within δ_x days of the last sonar day.

The problem is to decide which of two alternative hypotheses is correct: the null hypothesis, H0, which is that $\lambda_0 = \lambda_{\text{CS}}$, or the alternative hypothesis, HA, which asserts that $\lambda_0 < \lambda_{\text{CS}}$.

The slides that follow discuss a variety of available statistical tests and how they and other methods (including MCS) can be used to account for underlying uncertainties in the data, such as the arbitrariness of assuming $\delta_x = 6$ (why not $\delta_x = 3$, or 8, or any other number?).



Nontechnical walkthrough of existing approach (1/2)

- **Virtually all researchers (inside and outside of CNA) have traditionally based the decision to either accept or reject the null hypothesis (i.e., that strandings and sonar are uncorrelated) on the results of applying a single Poisson means test**
- In this test, the significance (or P-value) of observing a given number of coincident strandings is compared to the number that one expects to see based on how many strandings occur on days without sonar
  - The P-value estimates the probability that two means (the *observed* and *expected* number of coincident strandings) fall outside an acceptance region within which the two means are assumed equal
  - Small P-values that are less than some critical threshold (typically 0.05) are interpreted as providing sufficient evidence to reject the null hypothesis
- However, two potential issues arise by following this approach
- **The first issue is that the method assumes that the observed null stranding rate (estimated by dividing the number of observed strandings on days without sonar by the number of no-sonar days) is the true average of an underlying random Poisson process**
  - In fact, the true average may be any number that lies within *a range of numbers* (called the confidence interval) that may be estimated by assuming an underlying Poisson process

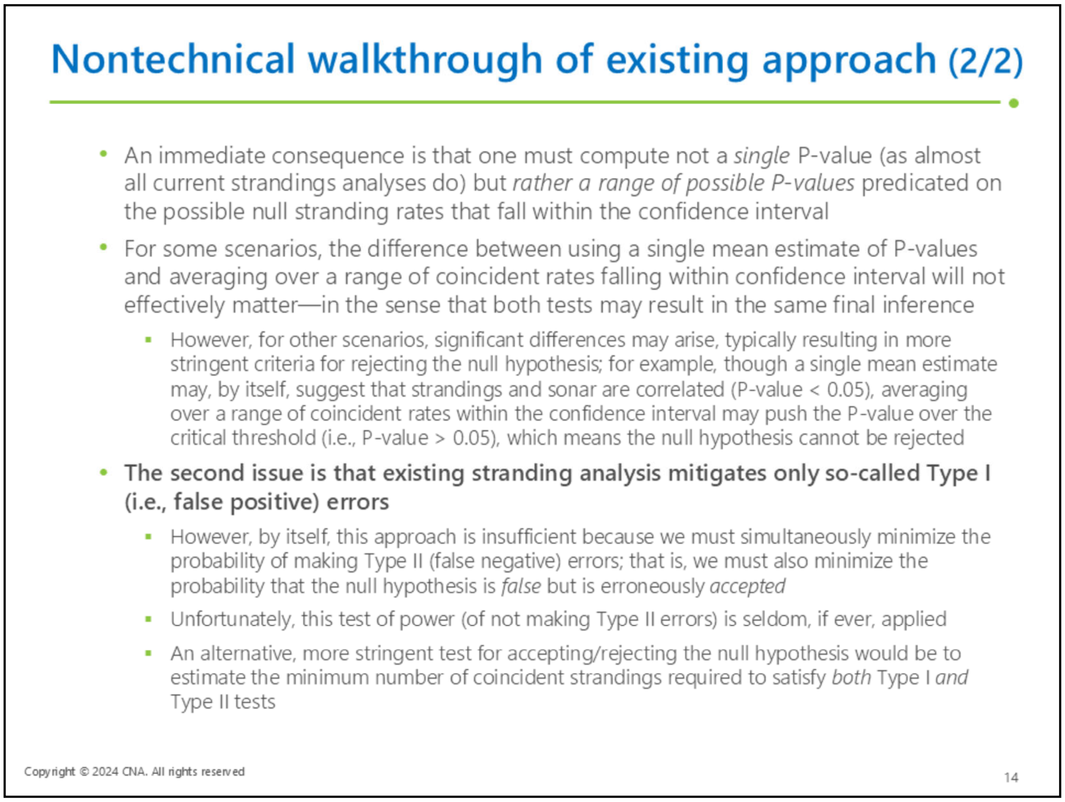
13

Slides 13 and 14 give a nontechnical walkthrough of the five deep-dive slides that follow (slides 14 to 19). Collectively, these seven slides lie at the heart of this study.

We briefly summarize the approach used in past CNA studies (and by other researchers outside CNA) to either accept or reject the null hypothesis, and then we discuss two critical issues that limit the general veracity of these statistical inferences.

The first issue is that the null stranding rate, $N_{\mathrm{NullS}}(\delta_x)$ / $D_{No\ \mathrm{Sonar}}(\delta_x)$, is not just an *approximate proxy* for the true average null stranding rate—it *is* the de facto null stranding rate.

But, assuming the null hypothesis, stranding events are distributed according to a random underlying Poisson process.



Nontechnical walkthrough of existing approach (2/2)

- An immediate consequence is that one must compute not a *single* P-value (as almost all current strandings analyses do) but *rather a range of possible P-values* predicated on the possible null stranding rates that fall within the confidence interval
- For some scenarios, the difference between using a single mean estimate of P-values and averaging over a range of coincident rates falling within confidence interval will not effectively matter—in the sense that both tests may result in the same final inference
  - However, for other scenarios, significant differences may arise, typically resulting in more stringent criteria for rejecting the null hypothesis; for example, though a single mean estimate may, by itself, suggest that strandings and sonar are correlated (P-value < 0.05), averaging over a range of coincident rates within the confidence interval may push the P-value over the critical threshold (i.e., P-value > 0.05), which means the null hypothesis cannot be rejected
- **The second issue is that existing stranding analysis mitigates only so-called Type I (i.e., false positive) errors**
  - However, by itself, this approach is insufficient because we must simultaneously minimize the probability of making Type II (false negative) errors; that is, we must also minimize the probability that the null hypothesis is *false* but is erroneously *accepted*
  - Unfortunately, this test of power (of not making Type II errors) is seldom, if ever, applied
  - An alternative, more stringent test for accepting/rejecting the null hypothesis would be to estimate the minimum number of coincident strandings required to satisfy *both* Type I *and* Type II tests

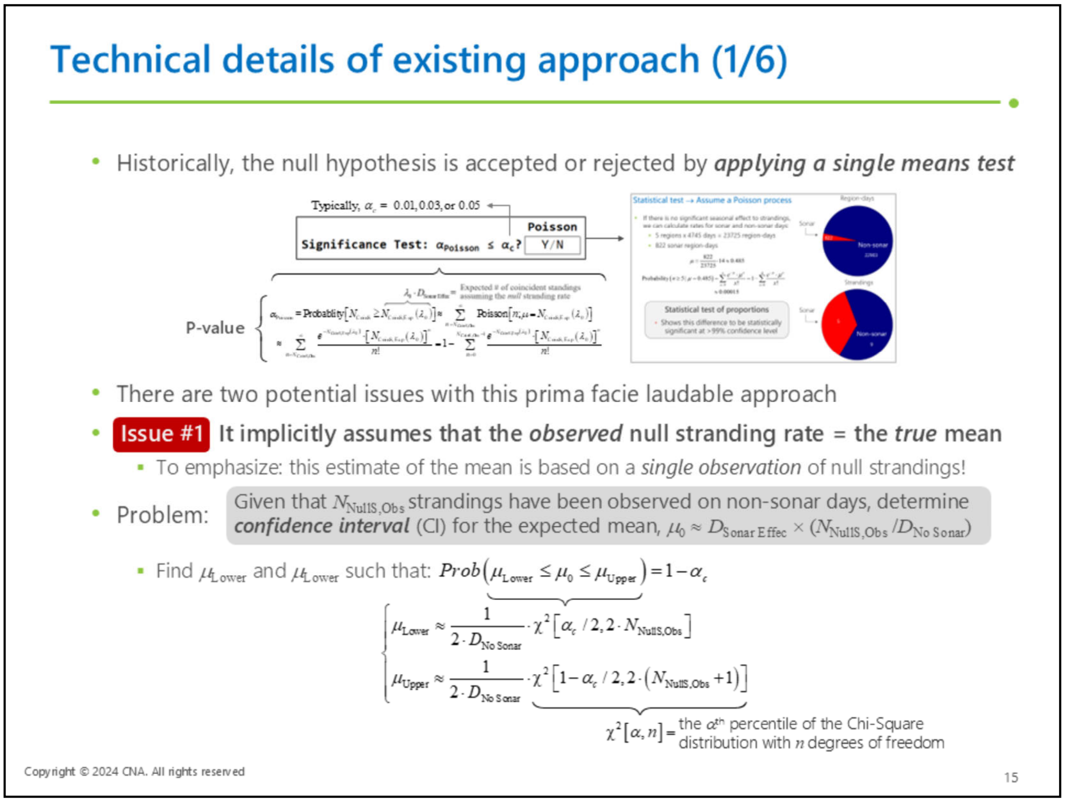
14

Assuming that stranding events are Poisson distributed, then the observed number of strandings represents only a *single sample* in a statistical distribution.

Imagine there are an infinite number of worlds governed by, and consistent with, Poisson-distributed strandings. In our world, we observe and record some specific number of strandings, N_S; in other words, it is our sole measurement and the only number we have. But our doppelgangers residing in other worlds observe a *range of stranding numbers*, some of which equal ours, some of which are smaller, and some of which are larger. This is the nature of probability, which lies at the core of all statistical analysis. In short, inferences cannot be drawn on the basis of a single observation of a set of randomly occurring events because random processes intrinsically entail multiple alternative probabilistic outcomes.

The immediate ramification is that it is insufficient to reject the null hypothesis on the basis of finding the P-value to be less than a critical threshold *if that P-value is estimated using the observed number of strandings as a single-valued proxy for the true mean of a random process*—which is what stranding analyses typically all do. Details appear on slides 15 and 16, but the salient nontechnical takeaway is that the observed number of strandings, N_S, is itself drawn from a random process for which the true mean lies somewhere between a lower bound, $N_{S,\text{Lower}}$, and upper bound, $N_{S,\text{Upper}}$: $N_{S,\text{Lower}} \leq N_S \leq N_{S,\text{Upper}}$.

The second issue limiting the veracity of statistical inferences is that stranding analyses typically mitigate only Type I (i.e., false positive) errors by estimating P-values, as discussed. However, by itself, this step is insufficient because we must simultaneously minimize the probability of making Type II (i.e., false negative) errors. That is, we must also minimize the probability of *erroneously accepting the null hypothesis when it is actually false*. Unfortunately, this test of power is seldom, if ever, made. As later slides show, this additional test effectively strengthens the criteria required to reject the null hypothesis.



Technical details of existing approach (1/6)

- Historically, the null hypothesis is accepted or rejected by ***applying a single means test***

Typically, α_c = 0.01, 0.03, or 0.05

Significance Test: $\alpha_{\text{Poisson}} \le \alpha_c$? Poisson Y/N

P-value

$$\begin{cases} \alpha_{\text{Poisson}} = \text{Probablity}\big[N_{\text{Coinc}} \ge \overbrace{N_{\text{Coinc,Exp}}(\lambda_0)}^{\lambda_0 \cdot D_{\text{Sonar Effec}}}\big] \approx \displaystyle\sum_{n=N_{\text{Coinc,Obs}}}^{\infty} \text{Poisson}\big[n;\mu = N_{\text{Coinc,Exp}}(\lambda_0)\big] \\ \approx \displaystyle\sum_{n=N_{\text{Coinc,Obs}}}^{\infty} \frac{e^{-N_{\text{Coinc,Exp}}(\lambda_0)} \cdot \big[N_{\text{Coinc,Exp}}(\lambda_0)\big]^n}{n!} = 1 - \sum_{n=0}^{N_{\text{Coinc,Obs}}-1} \frac{e^{-N_{\text{Coinc,Exp}}(\lambda_0)} \cdot \big[N_{\text{Coinc,Exp}}(\lambda_0)\big]^n}{n!} \end{cases}$$

$\lambda_0 \cdot D_{\text{Sonar Effec}}$ = Expected # of coincident standings assuming the *null* stranding rate

Statistical test → Assume a Poisson process

- If there is no significant seasonal effect to strandings, we can calculate rates for sonar and non-sonar days:
  - 5 regions x 4745 days = 23725 region-days
  - 822 sonar region-days

$$\mu = \frac{822}{23725} \cdot 14 \approx 0.485$$

$$\text{Probability}(n \ge 5 \mid \mu = 0.485) = \sum_{n=5}^{\infty} \frac{e^{-\mu} \cdot \mu^n}{n!} = 1 - \sum_{n=0}^{4} \frac{e^{-\mu} \cdot \mu^n}{n!} \approx 0.00015$$

Statistical test of proportions

- Shows this difference to be statistically significant at >99% confidence level

- There are two potential issues with this prima facie laudable approach
- **Issue #1** It implicitly assumes that the ***observed*** null stranding rate = the ***true*** mean
  - To emphasize: this estimate of the mean is based on a *single observation* of null strandings!
- Problem: Given that $N_{\text{NullS,Obs}}$ strandings have been observed on non-sonar days, determine ***confidence interval*** (CI) for the expected mean, $\mu_0 \approx D_{\text{Sonar Effec}} \times (N_{\text{NullS,Obs}} / D_{\text{No Sonar}})$
  - Find μ_{Lower} and μ_{Lower} such that: $\underbrace{Prob\left(\mu_{\text{Lower}} \le \mu_0 \le \mu_{\text{Upper}}\right)}_{} = 1 - \alpha_c$

$$\begin{cases} \mu_{\text{Lower}} \approx \dfrac{1}{2 \cdot D_{\text{No Sonar}}} \cdot \chi^2\left[\alpha_c / 2, 2 \cdot N_{\text{NullS,Obs}}\right] \\ \mu_{\text{Upper}} \approx \dfrac{1}{2 \cdot D_{\text{No Sonar}}} \cdot \underbrace{\chi^2\left[1 - \alpha_c / 2, 2 \cdot \left(N_{\text{NullS,Obs}} + 1\right)\right]}_{} \end{cases}$$

$\chi^2[\alpha, n]$ = the α^{th} percentile of the Chi-Square distribution with n degrees of freedom

Copyright © 2024 CNA. All rights reserved

15

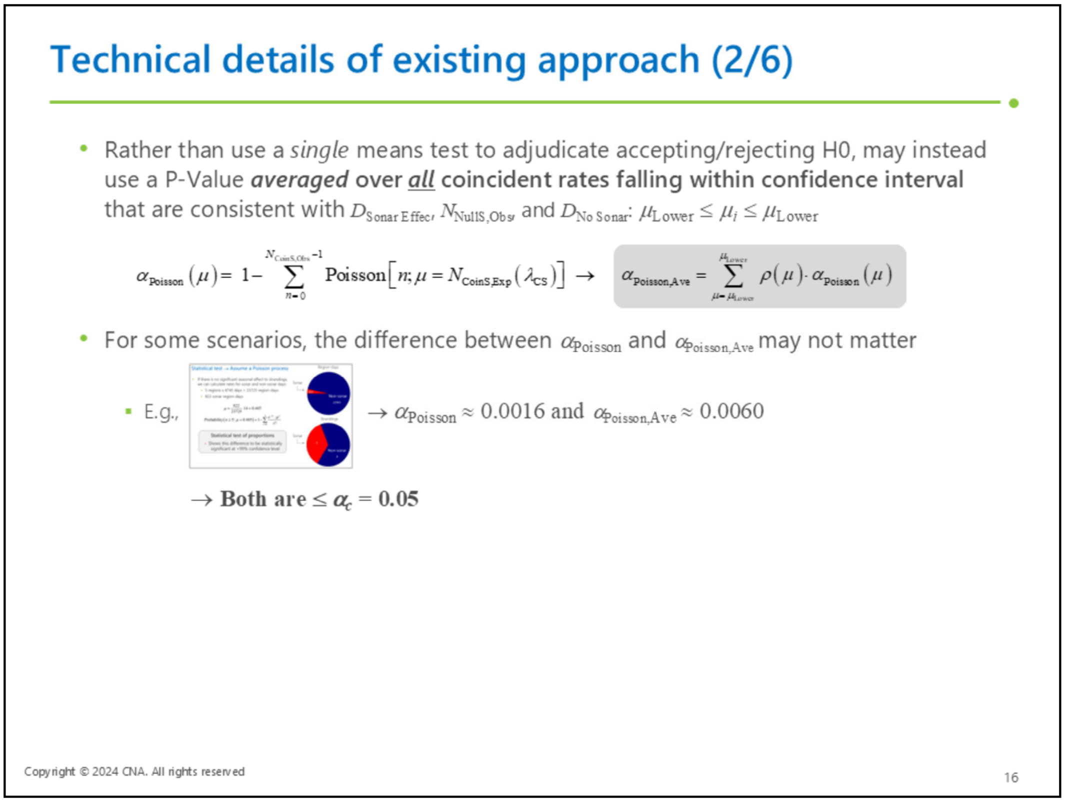
Slides 15-20 provide a deep-dive explanation of why the existing methodology is simultaneously correct and limited.

The method consists entirely of applying a *single-mean test* that estimates the P-value, or significance, of observing a given number of coincident strandings compared to the number that is expected (assuming that strandings and sonar are uncorrelated). A zoomed-in version of the equation at the top of the slide is given in **Appendix A**.

The point estimate for the Poisson mean is equal to the number of stranding events observed on days when sonar was not active, N_{NullS}, divided by the number of such days, $D_{\text{No Sonar}}$. And the P-value is the probability that at least the same number of coincident strandings as actually observed will result by applying the null stranding rate (= the Poisson mean, as just defined) to days with sonar. If the P-value is small (i.e., less than 0.05, meaning that, intuitively, it is extremely unlikely that the null stranding rate yields the observed number of coincident strandings), the conventional wisdom is that this finding provides sufficient statistical evidence to reject the null hypothesis that the null stranding rate is equal to the coincident stranding rate.

But several issues with this approach limit its veracity, anchored on the fact that what *ought* to be a statistical comparison between two means—mean #1 = *null stranding rate*, and mean #2 = *coincident stranding rate*—instead consists of equating mean #1 with the single observed value of null strandings, equating mean #2 with the number of observed coincident strandings, and comparing their respective Poisson statistics. The problem with this approach is that the only information we have to go on is a single observation of null and coincident strandings—in other words, the true means remain unknown.

The equations at the bottom of the slide use the χ^2 (i.e., "Chi-Squared") goodness-of-fit test for the Poisson distribution to estimate the lower, μ_{Lower}, and upper, μ_{Upper}, confidence limits of the true mean[9].



Technical details of existing approach (2/6)

- Rather than use a *single* means test to adjudicate accepting/rejecting H0, may instead use a P-Value ***averaged*** **over *__all__* coincident rates falling within confidence interval** that are consistent with $D_{\text{Sonar Effec}}$, $N_{\text{NullS,Obs}}$, and $D_{\text{No Sonar}}$: $\mu_{\text{Lower}} \le \mu_i \le \mu_{\text{Lower}}$

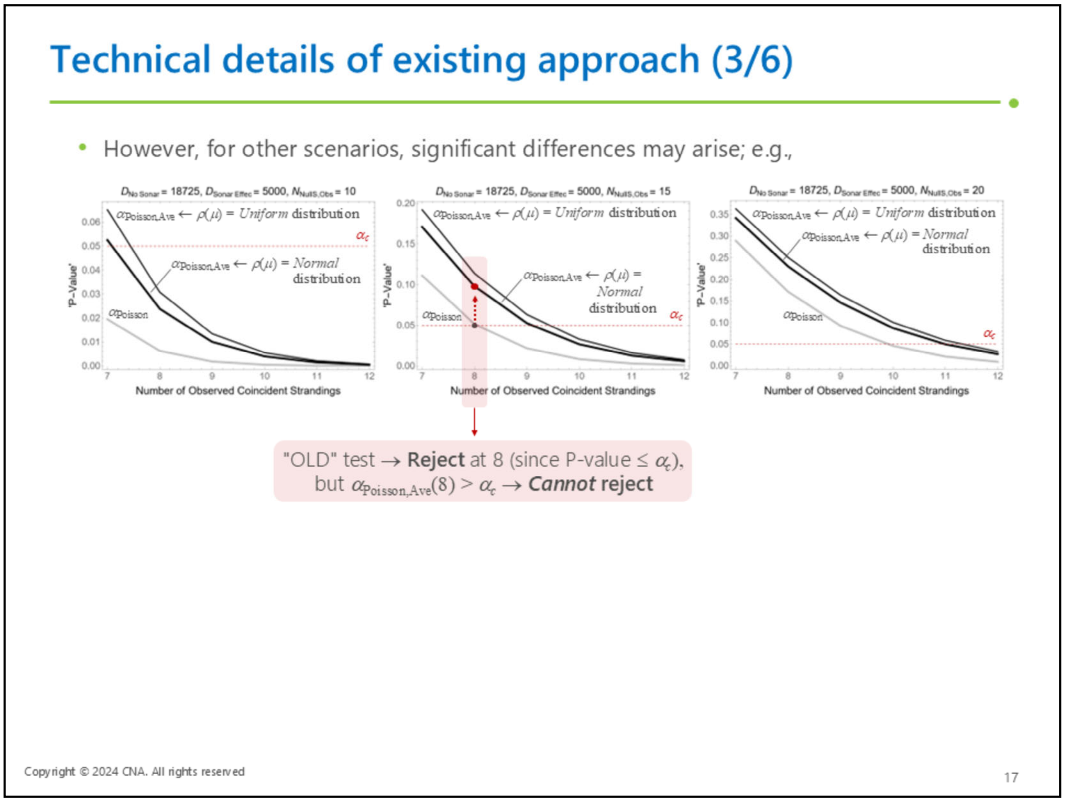
$$\alpha_{\text{Poisson}}(\mu) = 1 - \sum_{n=0}^{N_{\text{CoinS,Obs}}-1} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}(\lambda_{\text{CS}})\right] \rightarrow \alpha_{\text{Poisson,Ave}} = \sum_{\mu=\mu_{\text{Lower}}}^{\mu_{\text{Lower}}} \rho(\mu) \cdot \alpha_{\text{Poisson}}(\mu)$$

- For some scenarios, the difference between α_{Poisson} and $\alpha_{\text{Poisson,Ave}}$ may not matter
  - E.g., $\rightarrow \alpha_{\text{Poisson}} \approx 0.0016$ and $\alpha_{\text{Poisson,Ave}} \approx 0.0060$

$\rightarrow$ **Both are** $\le \alpha_c = 0.05$

Given that the value of the true null stranding rate, μ_{True}, lies somewhere within a confidence internal, $\mu_{\text{Lower}} \le \mu_{\text{True}} \le \mu_{\text{Upper}}$ (as explained on the previous slide), the single-mean test may be strengthened by averaging the P-value over all possible null stranding rates between μ_{Lower} and μ_{Upper}. The factor $\rho(\mu)$ that appears in the equation highlighted in gray (in the upper right side of the slide) represents the distribution of means around the "best guess" central value, $\mu_0 \approx D_{\text{sonar Effec}} \times (N_{\text{NullS}}/D_{\text{No Sonar}})$. For example, $\rho(\mu)$ may be uniformly distributed (although it represents a maximally conservative hypothesis because it is unlikely to be the case) or, more likely, normally distributed. Both hypotheses are considered in the bottom half of the slide. In either case, positing a range of possible null stranding rates effectively raises the bar on how much statistical evidence is required to reject the null hypothesis.

For some scenarios, this statistical refinement may not matter in a practical sense. For example, if *both* the existing (single-means) and strengthened (average-over-means-in-confidence-interval) P-value estimates are less than the critical threshold (or if both are greater), the final inference would remain the same—*accept* or *reject* the null hypothesis, respectively.



Technical details of existing approach (3/6)

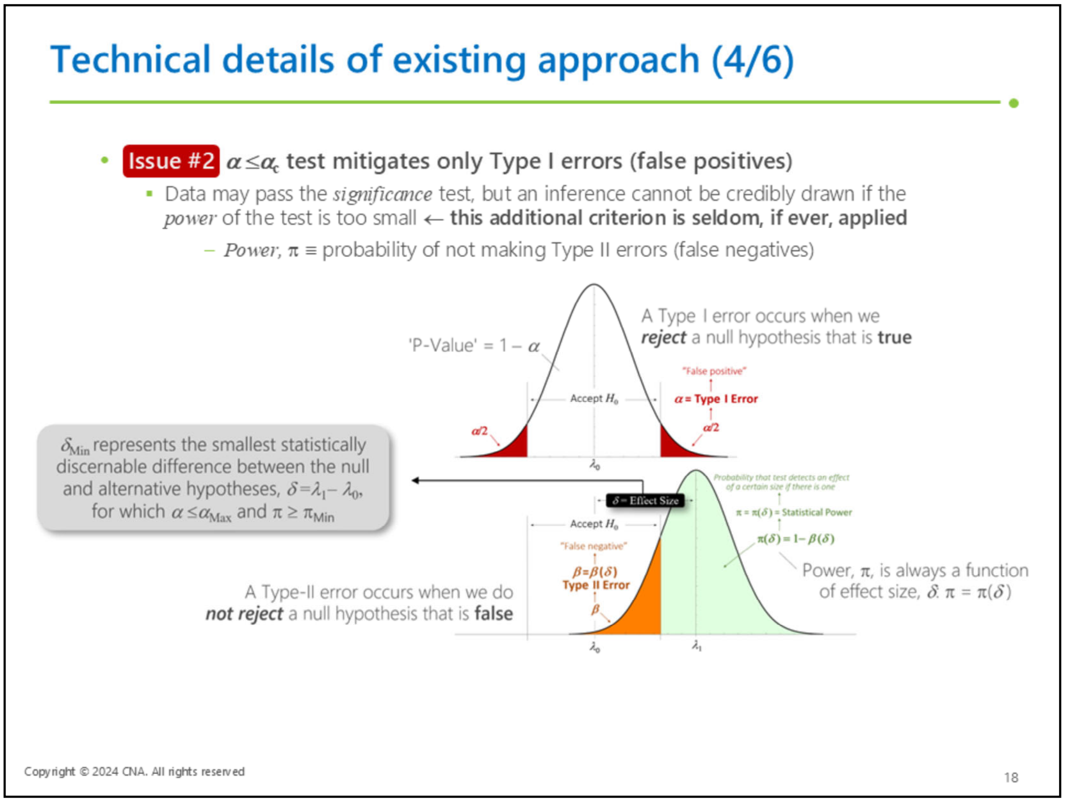
- However, for other scenarios, significant differences may arise; e.g.,

"OLD" test → **Reject** at 8 (since P-value ≤ α_c),
but $\alpha_{\text{Poisson,Ave}}(8) > \alpha_c$ → ***Cannot* reject**

17

However, for other scenarios, significant differences can arise, as illustrated by the figures on the bottom of the slide. The three plots show the P-value versus the number of observed coincident strandings for a notional dataset that contains 18,725 no sonar days, 500 effective sonar days (i.e., the total number of sonar days "padded" with days, $\delta_x = 6$ days), and—from left to right—10, 15, and 20 null strandings (i.e., strandings that occur on days without sonar), respectively. In each graph, the three curves—from bottom to top—denote the P-value as estimated using the single-mean Poisson test (α_{Poisson}), the average P-value using $\rho(\mu)$ = normal distribution, and the average P-value using $\rho(\mu)$ = uniform distribution, respectively.

The general case is illustrated by the portion of the second plot that is highlighted in red: if the observed number of coincident strandings is eight (or greater), the single-mean Poisson P-value falls below the critical value of 0.05, suggesting that the null hypothesis may be rejected. However, both P-values averaged over all possible null stranding rates within the confidence interval exceed the critical value and do *not* warrant rejection.



Technical details of existing approach (4/6)

- Issue #2 $\alpha \leq \alpha_c$ test mitigates only Type I errors (false positives)
  - Data may pass the *significance* test, but an inference cannot be credibly drawn if the *power* of the test is too small ← **this additional criterion is seldom, if ever, applied**
    - *Power*, $\pi \equiv$ probability of not making Type II errors (false negatives)

A Type I error occurs when we ***reject*** a null hypothesis that is **true**

δ_{Min} represents the smallest statistically discernable difference between the null and alternative hypotheses, $\delta = \lambda_1 - \lambda_0$, for which $\alpha \leq \alpha_{\text{Max}}$ and $\pi \geq \pi_{\text{Min}}$

A Type-II error occurs when we do ***not reject*** a null hypothesis that is **false**

Power, π, is always a function of effect size, δ: $\pi = \pi(\delta)$

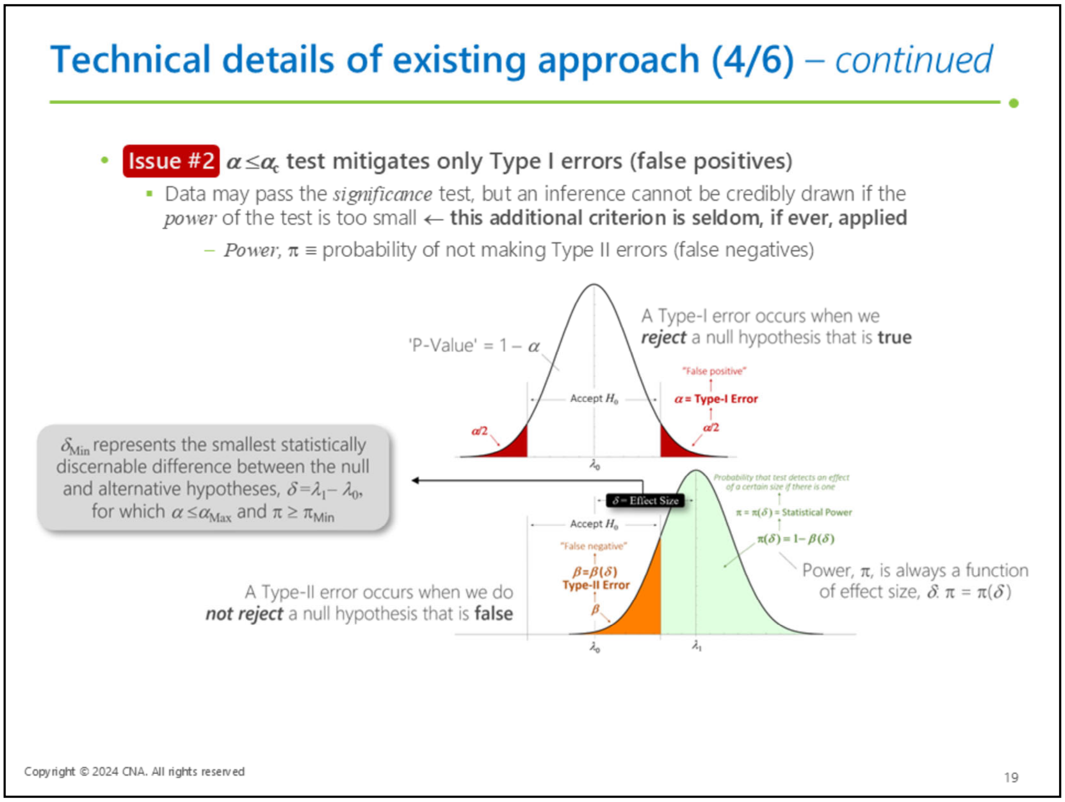
18

The second issue (limiting the veracity of statistical inferences) is that stranding analyses typically mitigate only so-called Type I (false positive) errors by estimating P-values. However, by itself, this approach is insufficient because we must simultaneously minimize the probability of making a Type II (false negative) error. That is, we must also minimize the probability of *erroneously accepting the null hypothesis when it is actually false*. Unfortunately, this test of power is seldom, if ever, performed. This slide illustrates graphically how these two types of statistical test errors are related.

A hypothesis test, $\mathfrak{T}$, consists of declaring two complementary assertions about the value of a specific parameter of interest and then testing to see which of the two hypotheses is best supported by the available data.

For example, if the goal is to determine whether the mean values of two probability distributions μ_1 and μ_0 are different, a typical choice for the first hypothesis is to assume that they are the same: $\mu_1 = \mu_0$ (which defines the null hypothesis, H0). The alternative hypothesis (HA) is then $\mu_1 \neq \mu_0$. And the obvious statistics to use in this case are the sample means, $<x_1>$ and $<x_0>$, which must be derived from the data.

$\mathfrak{T}$ is effectively a parametrized filter used to adjudicate what it means for $<x_1>$ and $<x_0>$ to be statistically close enough to warrant *rejecting the null hypothesis* (in favor of accepting HA). The standard practice is to use the parameter $\alpha \in [0,1]$ to define acceptance and rejection regions for H0. Specifically, $1 - \alpha$ is the probability that the $<x_1>$ and $<x_0>$ fall within the acceptance region of H0, and α is the probability that the means fall outside the acceptance region. A Type I error occurs when the null hypothesis, H0, is correct but is rejected. For this reason, the critical value that is chosen to be equal to α_c is typically small, such as $\alpha_c = 0.05$ or $\alpha_c = 0.01$.



Technical details of existing approach (4/6) – *continued*

- Issue #2 $\alpha \leq \alpha_c$ test mitigates only Type I errors (false positives)
  - Data may pass the *significance* test, but an inference cannot be credibly drawn if the *power* of the test is too small ← **this additional criterion is seldom, if ever, applied**
    - – *Power*, $\pi \equiv$ probability of not making Type II errors (false negatives)

A Type-I error occurs when we ***reject*** a null hypothesis that is **true**

δ_{Min} represents the smallest statistically discernable difference between the null and alternative hypotheses, $\delta = \lambda_1 - \lambda_0$, for which $\alpha \leq \alpha_{\text{Max}}$ and $\pi \geq \pi_{\text{Min}}$

A Type-II error occurs when we do ***not reject*** a null hypothesis that is **false**

Power, π, is always a function of effect size, δ: $\pi = \pi(\delta)$

19

(Continued …)

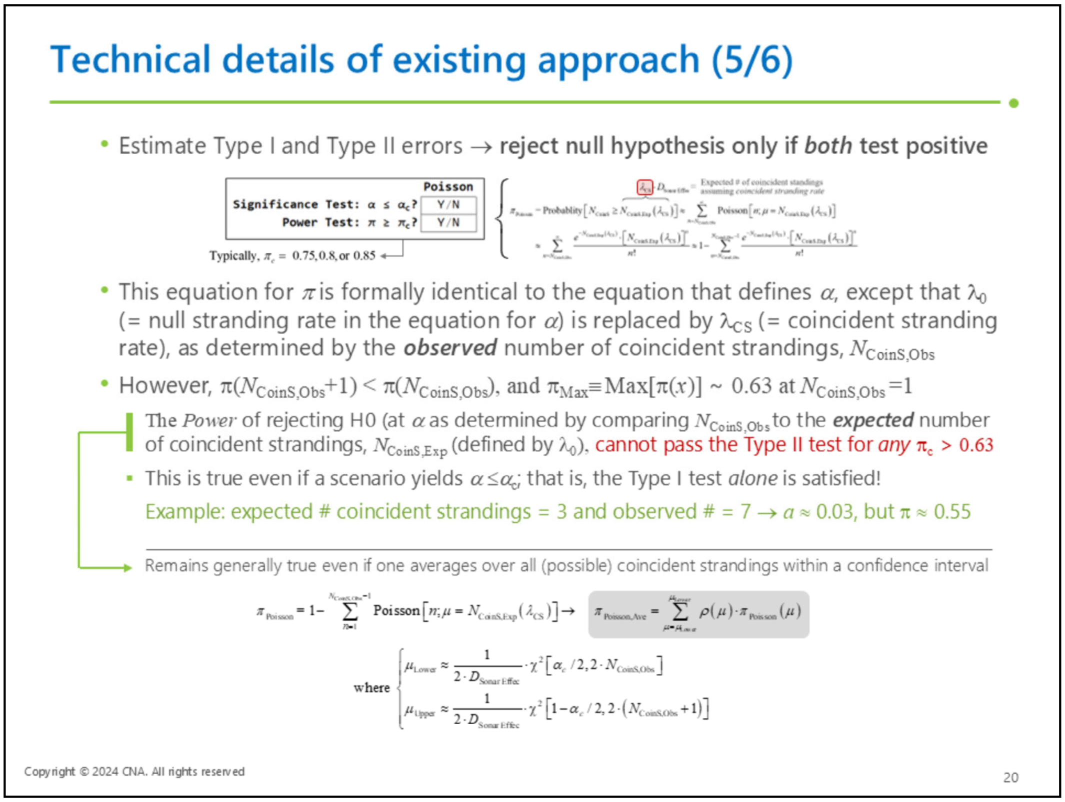
Now , assume that the H0 is false and consider the distribution of sample means under the HA: $\mu_1 \neq \mu_0$. Because the H0 is false, a credible test $\mathfrak{T}$ must reject it. But, as the slide illustrates, this "correct decision" will be made with the probability that $\pi = 1 - \beta$, where $\beta \in [0,1]$ is the Type II (false negative) error rate and π is the statistical power. A Type II error occurs when the null hypothesis is false but is erroneously accepted. Thus, the critical value of π is typically chosen to be large, such as π_c = 0.80, 0.85, or 0.9 (or even higher).

The salient points are as follows:

- Statistical power is always a function of the "effect size," $\pi = \pi(\delta)$, where $\delta = \mu_1 - \mu_{0.}$ Consequently, the desired critical value, π_c, effectively determines whether the hypothesis test, $\mathfrak{T}$, is capable of statistically resolving the difference between the two means.
- The veracity of $\mathfrak{T}$'s decision to accept (or reject) the null hypothesis depends on the degree to which *two criteria are both <u>simultaneously</u> satisfied*: $\alpha \leq \alpha_c$ and $\pi \leq \pi_c$.

Stated another way, although false positives are unlikely to occur with P-values of $\alpha \leq \alpha_c$—making it prima facie tempting to reject the null hypothesis (as is typically done in conventional stranding analyses)—the probability distributions describing the null and alternative distributions cannot be statistically discerned as being different unless the power is also sufficiently large.

*****Note***** While engaging in pedagogy of statistics is beyond the scope of this study, we would be remiss if we did not mention that "P-Values" are often misunderstood and/or misinterpreted. The P-Value denotes the probability that an observed difference between two distributions is due to chance, assuming the null hypothesis that the two distributions are the same. Small P-Values do not imply that there is a high probability that an observed difference is "correct"; **P-Values are not probability estimates for what the true values may be**. Similar, P-Values greater than or equal to a significance threshold do not demonstrate that two distributions are different, but only that there is a lack of statistical evidence to reject the null hypothesis: **large P-Values do not prove the null hypothesis**! Reference: N. Altman and M. Krzywinski, "P values and the search for significance," *Nature*, Vol. 14, 2017, https://www.nature.com/articles/nmeth.4120.



Technical details of existing approach (5/6)

- Estimate Type I and Type II errors → **reject null hypothesis only if *both* test positive**

	Poisson
Significance Test: $\alpha \le \alpha_c$?	Y/N
Power Test: $\pi \ge \pi_c$?	Y/N

Typically, $\pi_c = 0.75, 0.8, \text{or } 0.85$

$$\pi_{\text{Poisson}} = \text{Probablity}\left[N_{\text{CoinS}} \ge \underbrace{N_{\text{CoinS,Exp}}(\lambda_{\text{CS}})}_{\boxed{\lambda_{\text{CS}}} \cdot D_{\text{Sonar Effec}} = \text{Expected \# of coincident standings assuming coincident stranding rate}}\right] = \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}(\lambda_{\text{CS}})\right]$$

$$= \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \frac{e^{-N_{\text{CoinS,Exp}}(\lambda_{\text{CS}})} \cdot \left[N_{\text{CoinS,Exp}}(\lambda_{\text{CS}})\right]^n}{n!} = 1 - \sum_{n=N_{\text{CoinS,Obs}}}^{N_{\text{CoinS,Obs}}-1} \frac{e^{-N_{\text{CoinS,Exp}}(\lambda_{\text{CS}})} \cdot \left[N_{\text{CoinS,Exp}}(\lambda_{\text{CS}})\right]^n}{n!}$$

- This equation for π is formally identical to the equation that defines α, except that λ_0 (= null stranding rate in the equation for α) is replaced by λ_{CS} (= coincident stranding rate), as determined by the ***observed*** number of coincident strandings, $N_{\text{CoinS,Obs}}$
- However, $\pi(N_{\text{CoinS,Obs}}+1) < \pi(N_{\text{CoinS,Obs}})$, and $\pi_{\text{Max}} \equiv \text{Max}[\pi(x)] \sim 0.63$ at $N_{\text{CoinS,Obs}} = 1$
  - The *Power* of rejecting H0 (at α as determined by comparing $N_{\text{CoinS,Obs}}$ to the ***expected*** number of coincident strandings, $N_{\text{CoinS,Exp}}$ (defined by λ_0), cannot pass the Type II test for *any* $\pi_c > 0.63$
  - This is true even if a scenario yields $\alpha \le \alpha_c$; that is, the Type I test *alone* is satisfied!
  - Example: expected # coincident strandings = 3 and observed # = 7 → *a* ≈ 0.03, but π ≈ 0.55

Remains generally true even if one averages over all (possible) coincident strandings within a confidence interval

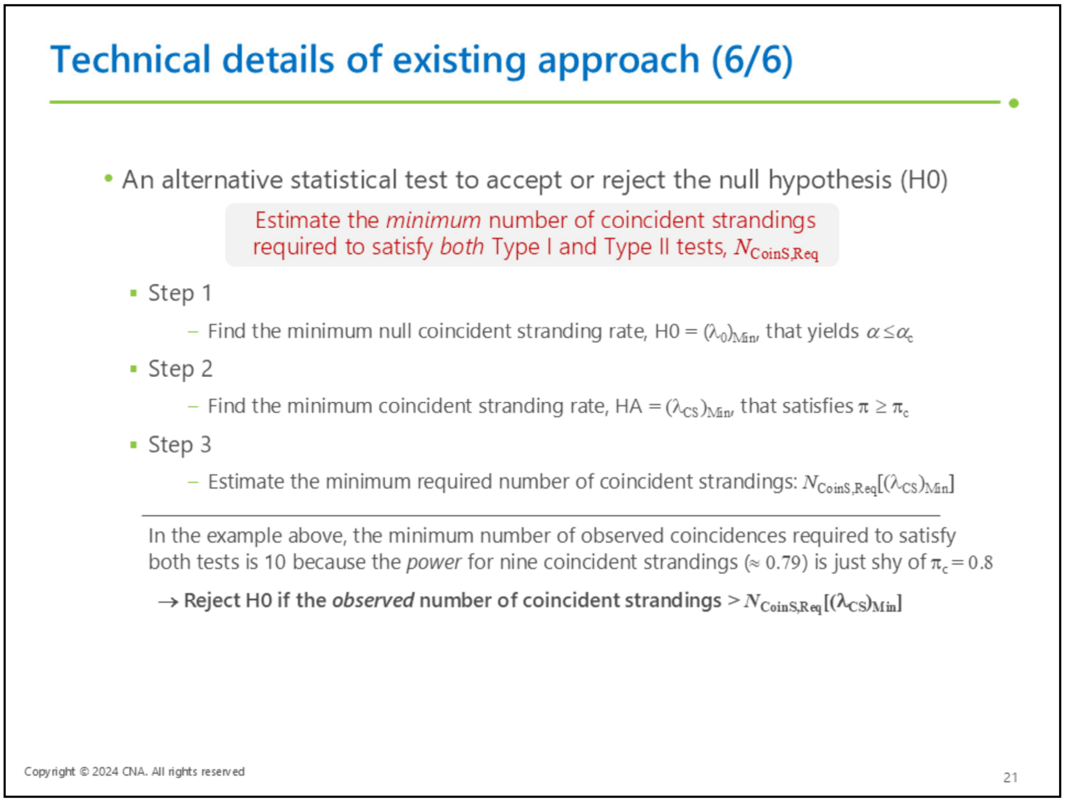
$$\pi_{\text{Poisson}} = 1 - \sum_{n=1}^{N_{\text{CoinS,Obs}}-1} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}(\lambda_{\text{CS}})\right] \rightarrow \pi_{\text{Poisson,Ave}} = \sum_{\mu=\mu_{\text{Lower}}}^{\mu_{\text{Lower}}} \rho(\mu) \cdot \pi_{\text{Poisson}}(\mu)$$

$$\text{where} \begin{cases} \mu_{\text{Lower}} \approx \dfrac{1}{2 \cdot D_{\text{Sonar Effec}}} \cdot \chi^2\left[\alpha_c/2, 2 \cdot N_{\text{CoinS,Obs}}\right] \\ \mu_{\text{Upper}} \approx \dfrac{1}{2 \cdot D_{\text{Sonar Effec}}} \cdot \chi^2\left[1-\alpha_c/2, 2 \cdot \left(N_{\text{CoinS,Obs}}+1\right)\right] \end{cases}$$

Summarizing the previous slide, the probability distributions describing the null and alternative hypotheses are statistically discernably different if and only if *both* the Type I and Type II error tests are satisfied.

Formally, the estimate for power, π, proceeds identically to how the P-value, α, is estimated, except that where the null stranding rate, λ_0, is used in the latter, the observed coincident stranding rate, λ_{CS}, is used in the former. Unfortunately, because of the nature of Poisson processes, the power to reject the null hypothesis (at the P-value estimated using the observed number of null strandings) will always be less than any critical (or desired) value greater than 0.63; in particular, if $\pi_c = 0.8$, as is typically assumed, then the Type II test will *always* fail. This is the case even if we average over all possible coincident strandings within the confidence interval (to mitigate using the single observation of null strandings as a proxy for an unknown true mean, as discussed earlier).

The next slide shows how both tests may be administered simultaneously.



Technical details of existing approach (6/6)

- An alternative statistical test to accept or reject the null hypothesis (H0)

Estimate the *minimum* number of coincident strandings required to satisfy *both* Type I and Type II tests, $N_{\mathrm{CoinS,Req}}$

- Step 1
  - Find the minimum null coincident stranding rate, H0 = $(\lambda_0)_{\mathrm{Min}}$, that yields $\alpha \leq \alpha_c$
- Step 2
  - Find the minimum coincident stranding rate, HA = $(\lambda_{\mathrm{CS}})_{\mathrm{Min}}$, that satisfies $\pi \geq \pi_c$
- Step 3
  - Estimate the minimum required number of coincident strandings: $N_{\mathrm{CoinS,Req}}[(\lambda_{\mathrm{CS}})_{\mathrm{Min}}]$

In the example above, the minimum number of observed coincidences required to satisfy both tests is 10 because the *power* for nine coincident strandings (≈ 0.79) is just shy of $\pi_c = 0.8$

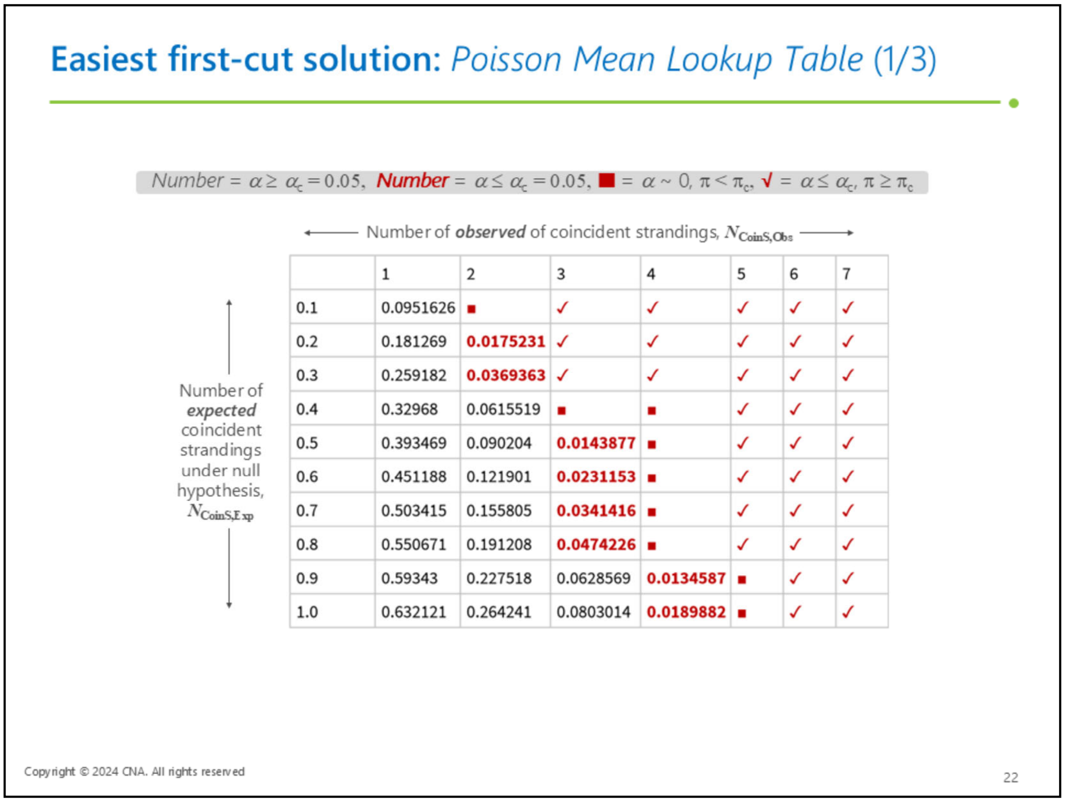
→ Reject H0 if the *observed* number of coincident strandings > $N_{\mathrm{CoinS,Req}}[(\lambda_{\mathrm{CS}})_{\mathrm{Min}}]$

21

This slide summarizes a three-step procedure to effectively administer both Type I and Type II tests simultaneously. The statistical inference—either to reject or accept the null hypothesis—is drawn not by directly estimating the P-value (and power, which, as shown on the last slide, cannot exceed 0.63 at a given P-value) but rather by comparing the minimal number of coincident strandings required to satisfy both tests to the observed number of coincident strandings.

An important caveat is that the observed number of coincident strandings is unclear because *coincident* is ambiguously defined. A given stranding (that occurs on, say, day $= D_{\mathrm{Stranding}}$) is typically labeled “coincident with sonar” if $D_{\mathrm{Stranding}} - D_{\mathrm{sonar,Last}} \leq 6$, where $D_{\mathrm{sonar,Last}}$ is the last day sonar was used prior to the stranding. But, the true stranding date must be used here, which is not always the observed date recorded in the dataset. Also, we emphasize that the six-day cutoff for associating a stranding with sonar is arbitrary. It could just as easily be seven or eight days. In later slides, we discuss how to mitigate ambiguities of this type and other uncertainties in the data.

The takeaway is that a better method could supplant the existing approach of rejecting the null hypothesis. Rather than rejecting the null hypothesis on the basis of a single-mean-derived P-value, analysts could instead estimate the minimal number of coincident strandings that must be observed to simultaneously satisfy both Type I and Type II tests. The null hypothesis would then be rejected if the observed number of coincident strandings is *greater* than the required number.



Easiest first-cut solution: *Poisson Mean Lookup Table* (1/3)

Number = $\alpha \geq \alpha_c = 0.05$, ***Number*** = $\alpha \leq \alpha_c = 0.05$, ■ = $\alpha \sim 0$, $\pi < \pi_c$, √ = $\alpha \leq \alpha_c$, $\pi \geq \pi_c$

← Number of ***observed*** of coincident strandings, $N_{\text{CoinS,Obs}}$ →

Number of ***expected*** coincident strandings under null hypothesis, $N_{\text{CoinS,Exp}}$

	1	2	3	4	5	6	7
0.1	0.0951626	■	✓	✓	✓	✓	✓
0.2	0.181269	**0.0175231**	✓	✓	✓	✓	✓
0.3	0.259182	**0.0369363**	✓	✓	✓	✓	✓
0.4	0.32968	0.0615519	■	■	✓	✓	✓
0.5	0.393469	0.090204	**0.0143877**	■	✓	✓	✓
0.6	0.451188	0.121901	**0.0231153**	■	✓	✓	✓
0.7	0.503415	0.155805	**0.0341416**	■	✓	✓	✓
0.8	0.550671	0.191208	**0.0474226**	■	✓	✓	✓
0.9	0.59343	0.227518	0.0628569	**0.0134587**	■	✓	✓
1.0	0.632121	0.264241	0.0803014	**0.0189882**	■	✓	✓

The next three slides leverage the aforementioned refinement of existing stranding analysis (that recommends administering both Type I and Type II tests simultaneously) to create a lookup table that can be used to draw at-a-glance statistical inferences.

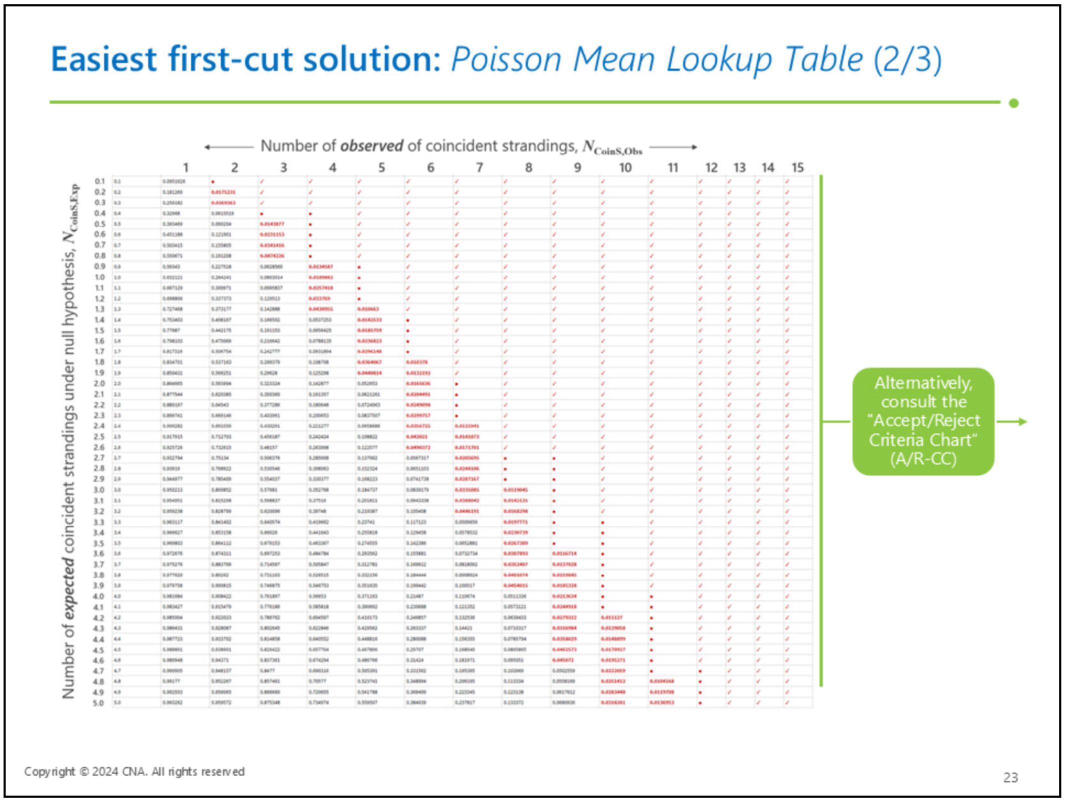
Given a dataset of interest, the only calculation that needs to be made is to estimate the expected number of coincident strandings, $N_{\text{CoinS,Exp}}$:

$$N_{\text{CoinS,Exp}} \approx \left(\frac{N_{\text{NullS}}}{D_{\text{No Sonar}}} \right) \cdot D_{\text{Sonar Effec}}$$

where N_{NullS} = the number of null strandings that occur on days without sonar, $D_{No\ \text{Sonar}}$ = the number of days when sonar was not active, and $D_{\text{sonar Effec}}$ = the number of effective sonar days (i.e., all days when sonar was active along with all days that were within δ_x days of the last sonar day, where δ_x is conventionally equal to six days).

Once the value of $N_{\text{CoinS,Exp}}$ is estimated, the user has to locate, roughly, the row (or space between rows) in which the value most closely resides and then the element in the column that corresponds to the number of observed coincident strandings. This element assumes one of four forms:

- A number highlighted in **black**, which means that the resulting P-value > α_c and thus that the null hypothesis *cannot* be rejected.
- A **red** number, which means that the resulting P-value ≤ α_c and thus that the null hypothesis *can* be rejected but *only if the Type I test is satisfied.*
- A red symbol, ■, which means that the resulting P-value is close to zero but that the power π < π_c; thus, the null hypothesis *can* be rejected but *only if the Type I test is satisfied.*
- A red check mark, √, which means the single null hypothesis *can* be rejected using the strictest criteria: *both Type I and Type II tests must be satisfied.*



Easiest first-cut solution: *Poisson Mean Lookup Table* (2/3)

← Number of ***observed*** of coincident strandings, $N_{\mathrm{CoinS,Obs}}$ →

Number of ***expected*** coincident strandings under null hypothesis, $N_{\mathrm{CoinS,Exp}}$

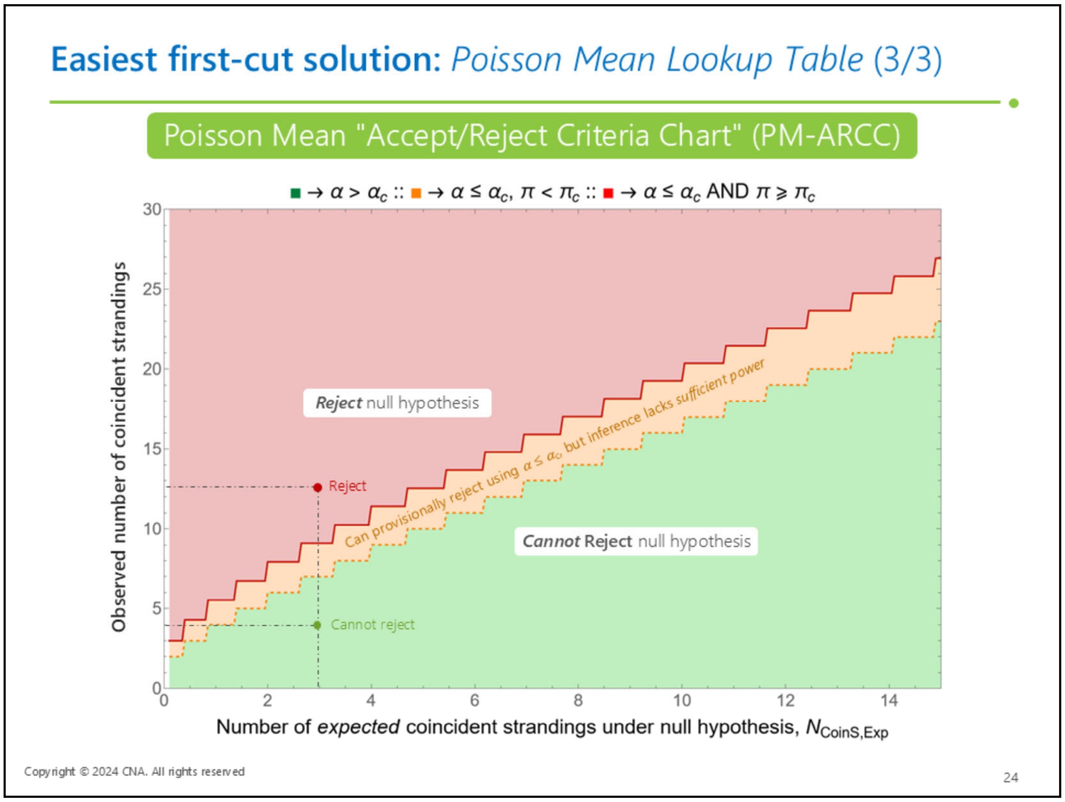
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.1	0.1	0.0951626	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.2	0.2	0.181269	0.0175231	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.3	0.3	0.259182	0.0369363	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.4	0.4	0.32968	0.0615519	●	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.5	0.5	0.393469	0.090204	0.0143877	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.6	0.6	0.451188	0.121901	0.0231153	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.7	0.7	0.503415	0.155805	0.0341416	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.8	0.8	0.550671	0.191208	0.0474226	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.9	0.9	0.59343	0.227518	0.0628569	0.0134587	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.0	1.0	0.632121	0.264241	0.0803014	0.0189882	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.1	1.1	0.667129	0.300971	0.0995837	0.0257418	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.2	1.2	0.698806	0.337373	0.120513	0.033769	●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.3	1.3	0.727468	0.373177	0.142888	0.0430955	0.010663	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.4	1.4	0.753403	0.408167	0.166502	0.0537253	0.0142533	●	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.5	1.5	0.77687	0.442175	0.191153	0.0656425	0.0185759	●	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.6	1.6	0.798103	0.475069	0.216642	0.0788135	0.0236823	●	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.7	1.7	0.817316	0.506754	0.242777	0.0931894	0.0296148	●	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.8	1.8	0.834701	0.537163	0.269379	0.108708	0.0364067	0.010378	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.9	1.9	0.850431	0.566251	0.29628	0.125298	0.0440814	0.0132192	✓	✓	✓	✓	✓	✓	✓	✓	✓
2.0	2.0	0.864665	0.593994	0.323324	0.142877	0.052653	0.0165636	●	✓	✓	✓	✓	✓	✓	✓	✓
2.1	2.1	0.877544	0.620385	0.350369	0.161357	0.0621261	0.0204491	●	✓	✓	✓	✓	✓	✓	✓	✓
2.2	2.2	0.889197	0.64543	0.377286	0.180648	0.0724963	0.0249098	●	✓	✓	✓	✓	✓	✓	✓	✓
2.3	2.3	0.899741	0.669146	0.403961	0.200653	0.0837507	0.0299757	●	✓	✓	✓	✓	✓	✓	✓	✓
2.4	2.4	0.909282	0.691559	0.430291	0.221277	0.0958686	0.0356725	0.0115941	✓	✓	✓	✓	✓	✓	✓	✓
2.5	2.5	0.917915	0.712703	0.456187	0.242424	0.108822	0.042021	0.0141873	✓	✓	✓	✓	✓	✓	✓	✓
2.6	2.6	0.925726	0.732615	0.48157	0.263998	0.122577	0.0490373	0.0171701	✓	✓	✓	✓	✓	✓	✓	✓
2.7	2.7	0.932794	0.75134	0.506376	0.285908	0.137092	0.0567317	0.0205695	●	●	✓	✓	✓	✓	✓	✓
2.8	2.8	0.93919	0.768922	0.530546	0.308063	0.152324	0.0651103	0.0244106	●	●	✓	✓	✓	✓	✓	✓
2.9	2.9	0.944977	0.785409	0.554037	0.330377	0.168223	0.0741738	0.0287167	●	●	✓	✓	✓	✓	✓	✓
3.0	3.0	0.950213	0.800852	0.57681	0.352768	0.184737	0.0839179	0.0335085	0.0119045	●	✓	✓	✓	✓	✓	✓
3.1	3.1	0.954951	0.815298	0.598837	0.37516	0.201811	0.0943338	0.0388042	0.0142125	●	✓	✓	✓	✓	✓	✓
3.2	3.2	0.959238	0.828799	0.620096	0.39748	0.219387	0.105408	0.0446191	0.0168298	●	✓	✓	✓	✓	✓	✓
3.3	3.3	0.963117	0.841402	0.640574	0.419662	0.23741	0.117123	0.0509658	0.0197771	●	●	✓	✓	✓	✓	✓
3.4	3.4	0.966627	0.853158	0.66026	0.441643	0.255818	0.129458	0.0578532	0.0230739	●	●	✓	✓	✓	✓	✓
3.5	3.5	0.969803	0.864112	0.679153	0.463367	0.274555	0.142386	0.0652881	0.0267389	●	●	✓	✓	✓	✓	✓
3.6	3.6	0.972676	0.874311	0.697253	0.484784	0.293562	0.155881	0.0732734	0.0307893	0.0116714	●	✓	✓	✓	✓	✓
3.7	3.7	0.975276	0.883799	0.714567	0.505847	0.312781	0.169912	0.0818092	0.0352407	0.0137028	●	✓	✓	✓	✓	✓
3.8	3.8	0.977629	0.89262	0.731103	0.526515	0.332156	0.184444	0.0908924	0.0401074	0.0159845	●	✓	✓	✓	✓	✓
3.9	3.9	0.979758	0.900815	0.746875	0.546753	0.351635	0.199442	0.100517	0.0454015	0.0185328	●	✓	✓	✓	✓	✓
4.0	4.0	0.981684	0.908422	0.761897	0.56653	0.371163	0.21487	0.110674	0.0511336	0.0213634	●	●	✓	✓	✓	✓
4.1	4.1	0.983427	0.915479	0.776186	0.585818	0.390692	0.230688	0.121352	0.0573121	0.0244918	●	●	✓	✓	✓	✓
4.2	4.2	0.985004	0.922023	0.789762	0.604597	0.410173	0.246857	0.132536	0.0639433	0.0279322	0.011127	●	✓	✓	✓	✓
4.3	4.3	0.986431	0.928087	0.802645	0.622846	0.429562	0.263337	0.14421	0.0710317	0.0316984	0.0129058	●	✓	✓	✓	✓
4.4	4.4	0.987723	0.933702	0.814858	0.640552	0.448816	0.280088	0.156355	0.0785794	0.0358029	0.0148899	●	✓	✓	✓	✓
4.5	4.5	0.988891	0.938901	0.826422	0.657704	0.467896	0.29707	0.168949	0.0865865	0.0402573	0.0170927	●	✓	✓	✓	✓
4.6	4.6	0.989948	0.94371	0.837361	0.674294	0.486766	0.31424	0.181971	0.095051	0.045072	0.0195271	●	✓	✓	✓	✓
4.7	4.7	0.990905	0.948157	0.8477	0.690316	0.505391	0.331562	0.195395	0.103969	0.0502559	0.0222059	●	●	✓	✓	✓
4.8	4.8	0.99177	0.952267	0.857461	0.70577	0.523741	0.348994	0.209195	0.113334	0.0558169	0.0251412	0.0104168	●	✓	✓	✓
4.9	4.9	0.992553	0.956065	0.866669	0.720655	0.541788	0.366499	0.223345	0.123138	0.0617612	0.0283448	0.0119708	●	✓	✓	✓
5.0	5.0	0.993262	0.959572	0.875348	0.734974	0.559507	0.384039	0.237817	0.133372	0.0680936	0.0318281	0.0136953	●	✓	✓	✓

Alternatively, consult the "Accept/Reject Criteria Chart" (A/R-CC)

This slide shows an expanded view of the lookup table introduced on the preceding slide.

Finer resolution tables may be generated in a few seconds using the Mathematica source code developed for this study (see **Appendix H**).

The next slide shows an even simpler purely graphical chart based on this lookup table.



Easiest first-cut solution: *Poisson Mean Lookup Table* (3/3)

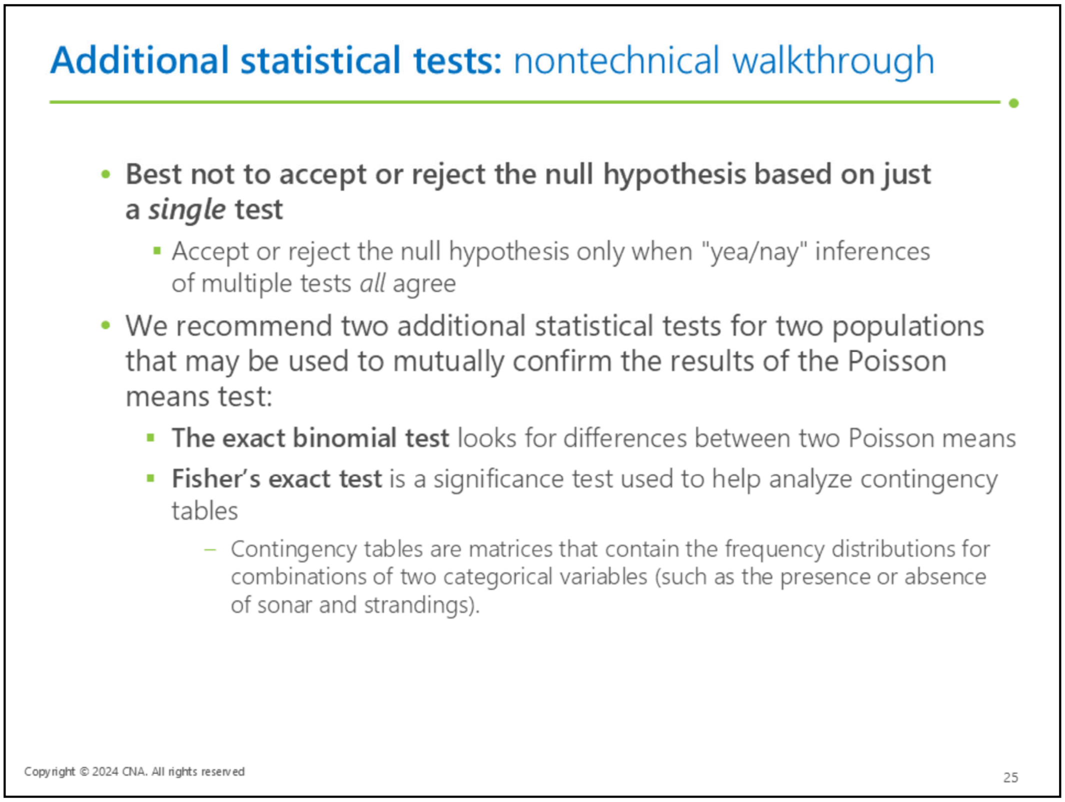
Poisson Mean "Accept/Reject Criteria Chart" (PM-ARCC)

24

The Poisson mean "Accept/Reject Criteria Chart" pulls together all of the methodological refinements discussed so far and displays them graphically.

The only calculation the user needs to make is to estimate the expected number of coincident strandings, $N_{\text{CoinS,Exp}}$, as explained on slide 19. Once this value is found, the color on the part of the chart that corresponds to where the vertical line (anchored on $x = N_{\text{CoinS,Exp}}$) intersects the horizontal line (y = observed number of coincident strandings) determines the statistical inference:

- **Green** means that the null hypothesis *cannot* be rejected.
- **Orange** means that the null hypothesis can be *provisionally* rejected (on the basis of the significance, or Type I test, alone), with the caveat that such an inference lacks sufficient statistical power.
- **Red** means that the null hypothesis *may* be rejected (after passing both Type I and Type II error tests).



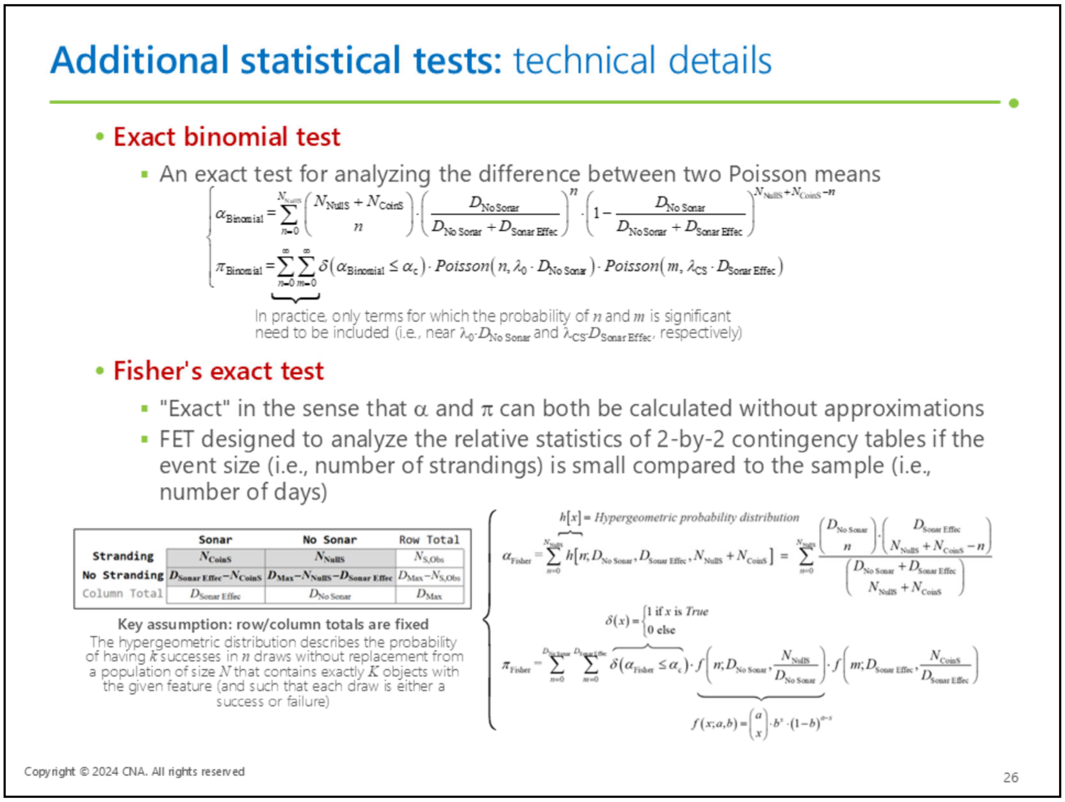
Additional statistical tests: nontechnical walkthrough

- **Best not to accept or reject the null hypothesis based on just a *single* test**
  - Accept or reject the null hypothesis only when "yea/nay" inferences of multiple tests *all* agree
- We recommend two additional statistical tests for two populations that may be used to mutually confirm the results of the Poisson means test:
  - **The exact binomial test** looks for differences between two Poisson means
  - **Fisher's exact test** is a significance test used to help analyze contingency tables
    - Contingency tables are matrices that contain the frequency distributions for combinations of two categorical variables (such as the presence or absence of sonar and strandings).

25

This slide provides a brief, nontechnical overview of two additional statistical tests that are introduced on the next slide. Because all statistical analyses are inevitably accompanied (and are often plagued) by methodological and interpretative caveats (e.g., approximation, assumptions, incomplete or ambiguous data, small sample sizes) [1], it is standard practice to apply more than one test whenever possible. Although different statistical tests typically yield similar results, any disagreements among properly administered tests may indicate potential inconsistencies, errors, or erroneous interpretations of the underlying data. A battery of tests may therefore be necessary to mutually confirm the consistency and veracity of the final inference.

There are far too many statistical tests available to even enumerate, much less describe and apply for this study. The texts by Fagerland [5] and Mathews [8] provide cogent pedagogical discussions. For illustrative purposes, we use two well-known tests: Fisher's exact test (FET) and the exact binomial test (EBT). Details appear on the following slide.



Additional statistical tests: technical details

- **Exact binomial test**
  - An exact test for analyzing the difference between two Poisson means

$$\begin{cases} \alpha_{\text{Binomial}} = \sum_{n=0}^{N_{\text{NullS}}} \binom{N_{\text{NullS}} + N_{\text{CoinS}}}{n} \cdot \left(\frac{D_{\text{No Sonar}}}{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}} \right)^{n} \cdot \left(1 - \frac{D_{\text{No Sonar}}}{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}} \right)^{N_{\text{NullS}} + N_{\text{CoinS}} - n} \\ \pi_{\text{Binomial}} = \underbrace{\sum_{n=0}^{\infty} \sum_{m=0}^{\infty}} \delta\left(\alpha_{\text{Binomial}} \le \alpha_c\right) \cdot Poisson\left(n, \lambda_0 \cdot D_{\text{No Sonar}}\right) \cdot Poisson\left(m, \lambda_{\text{CS}} \cdot D_{\text{Sonar Effec}}\right) \end{cases}$$

In practice, only terms for which the probability of n and m is significant need to be included (i.e., near $\lambda_0 \cdot D_{\text{No Sonar}}$ and $\lambda_{\text{CS}} \cdot D_{\text{Sonar Effec}}$, respectively)

- **Fisher's exact test**
  - "Exact" in the sense that α and π can both be calculated without approximations
  - FET designed to analyze the relative statistics of 2-by-2 contingency tables if the event size (i.e., number of strandings) is small compared to the sample (i.e., number of days)

	Sonar	No Sonar	Row Total
Stranding	N_{CoinS}	N_{NullS}	$N_{\text{S,Obs}}$
No Stranding	$D_{\text{Sonar Effec}} - N_{\text{CoinS}}$	$D_{\text{Max}} - N_{\text{NullS}} - D_{\text{Sonar Effec}}$	$D_{\text{Max}} - N_{\text{S,Obs}}$
Column Total	$D_{\text{Sonar Effec}}$	$D_{\text{No Sonar}}$	D_{Max}

Key assumption: row/column totals are fixed

The hypergeometric distribution describes the probability of having k successes in n draws without replacement from a population of size N that contains exactly K objects with the given feature (and such that each draw is either a success or failure)

$$\begin{cases} \alpha_{\text{Fisher}} = \sum_{n=0}^{N_{\text{NullS}}} \overbrace{h}^{h[x] = \textit{Hypergeometric probability distribution}}\left[n, D_{\text{No Sonar}}, D_{\text{Sonar Effec}}, N_{\text{NullS}} + N_{\text{CoinS}}\right] = \sum_{n=0}^{N_{\text{NullS}}} \frac{\binom{D_{\text{No Sonar}}}{n} \cdot \binom{D_{\text{Sonar Effec}}}{N_{\text{NullS}} + N_{\text{CoinS}} - n}}{\binom{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}}{N_{\text{NullS}} + N_{\text{CoinS}}}} \\ \pi_{\text{Fisher}} = \sum_{n=0}^{D_{\text{No Sonar}}} \sum_{m=0}^{D_{\text{Sonar Effec}}} \overbrace{\delta\left(\alpha_{\text{Fisher}} \le \alpha_c\right)}^{\delta(x) = \begin{cases} 1 \text{ if } x \text{ is } \textit{True} \\ 0 \text{ else} \end{cases}} \cdot \underbrace{f\left(n; D_{\text{No Sonar}}, \frac{N_{\text{NullS}}}{D_{\text{No Sonar}}}\right)}_{f(x;a,b) = \binom{a}{x} \cdot b^x \cdot (1-b)^{a-x}} \cdot f\left(m; D_{\text{Sonar Effec}}, \frac{N_{\text{CoinS}}}{D_{\text{Sonar Effec}}}\right) \end{cases}$$

Copyright © 2024 CNA. All rights reserved

26

This slide gives the details of the EBT and FET. They are both exact in the sense that both significance, α, and power, π, can be calculated exactly, without resorting to approximations.

The binomial test is based on noting that the conditional distribution of x given $x + y = k$ follows a binomial distribution [7]. Specifically (and using single-letter notional variables for simplicity), we are testing the null hypothesis , $\lambda_0 = \lambda_A$, in comparison with the alternative hypothesis, $\lambda_0 < \lambda_A$, where $\lambda_0 = x_0 / n_0$ and $\lambda_A = x_A / n_A$ are the Poisson rates, x_0 and x_A are the number of null strandings (on non-sonar days) and coincident strandings (or the count under the alternative hypothesis), and n_0 and n_A are the number of null (non-sonar) and alternative (sonar) days. Then x_0 and x_A are Poisson distributed with means $\mu_1 = n_1 \times \lambda_1$ and $\mu_2 = n_2 \times \lambda_2$, respectively. The P-value is then given by the binomial probability:

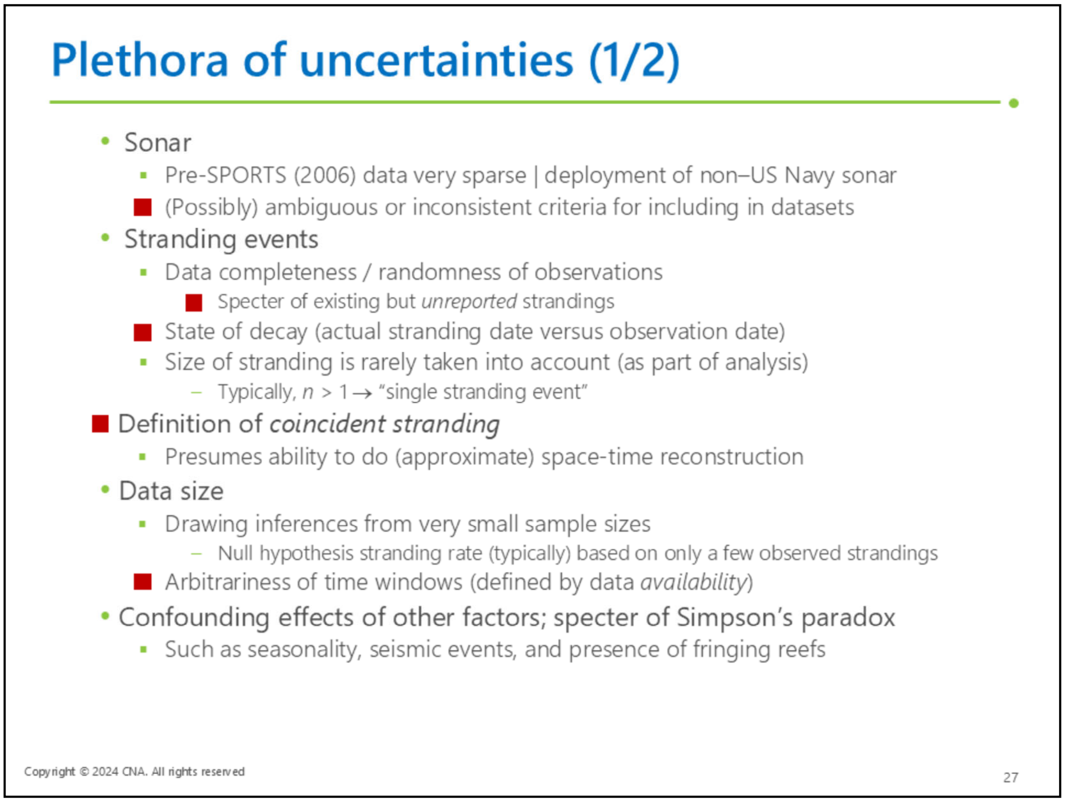
$$\text{Prob}\left(0 \le x \le x_1; k = x_1 + x_2; p = \frac{n_1}{n_1 + n_2}\right) = \sum_{x=0}^{x_1} \binom{k}{x} \cdot p^x \cdot (1-p)^{k-x} = \sum_{x=0}^{x_1} \binom{x_1 + x_2}{x} \cdot \left(\frac{n_1}{n_1 + n_2}\right)^x \cdot \left(1 - \frac{n_1}{n_1 + n_2}\right)^{x_1 + x_2 - x}$$

where $\binom{a}{b} = a! \ /[b!\ (a - b)]$, and $a! = a \times (\text{a-1}) \times \ldots \times 2 \times 1$.

In the expression for power (=π_{Binomial}), $Poisson(x,\mu) = \mu^x \cdot e^{-\mu}/x!$

The equations (for significance and power) of FET are given in the lower half of the slide. The only assumption made to derive them is that the binary data are all independent [6].

Based on the 2-by-2 contingency table that appears at bottom left, the test consists of calculating the probability of obtaining the expected results assuming the null hypothesis is true, using all possible 2-by-2 tables that could have been observed for combinations of matrix elements. The sum of rows and columns—known as “marginal” totals—are both fixed by the observed data.



Plethora of uncertainties (1/2)

- Sonar
  - Pre-SPORTS (2006) data very sparse | deployment of non–US Navy sonar
  - 🟥 (Possibly) ambiguous or inconsistent criteria for including in datasets
- Stranding events
  - Data completeness / randomness of observations
    - 🟥 Specter of existing but *unreported* strandings
  - 🟥 State of decay (actual stranding date versus observation date)
  - Size of stranding is rarely taken into account (as part of analysis)
    - Typically, $n > 1 \rightarrow$ "single stranding event"
- 🟥 Definition of *coincident stranding*
  - Presumes ability to do (approximate) space-time reconstruction
- Data size
  - Drawing inferences from very small sample sizes
    - Null hypothesis stranding rate (typically) based on only a few observed strandings
  - 🟥 Arbitrariness of time windows (defined by data *availability*)
- Confounding effects of other factors; specter of Simpson's paradox
  - Such as seasonality, seismic events, and presence of fringing reefs

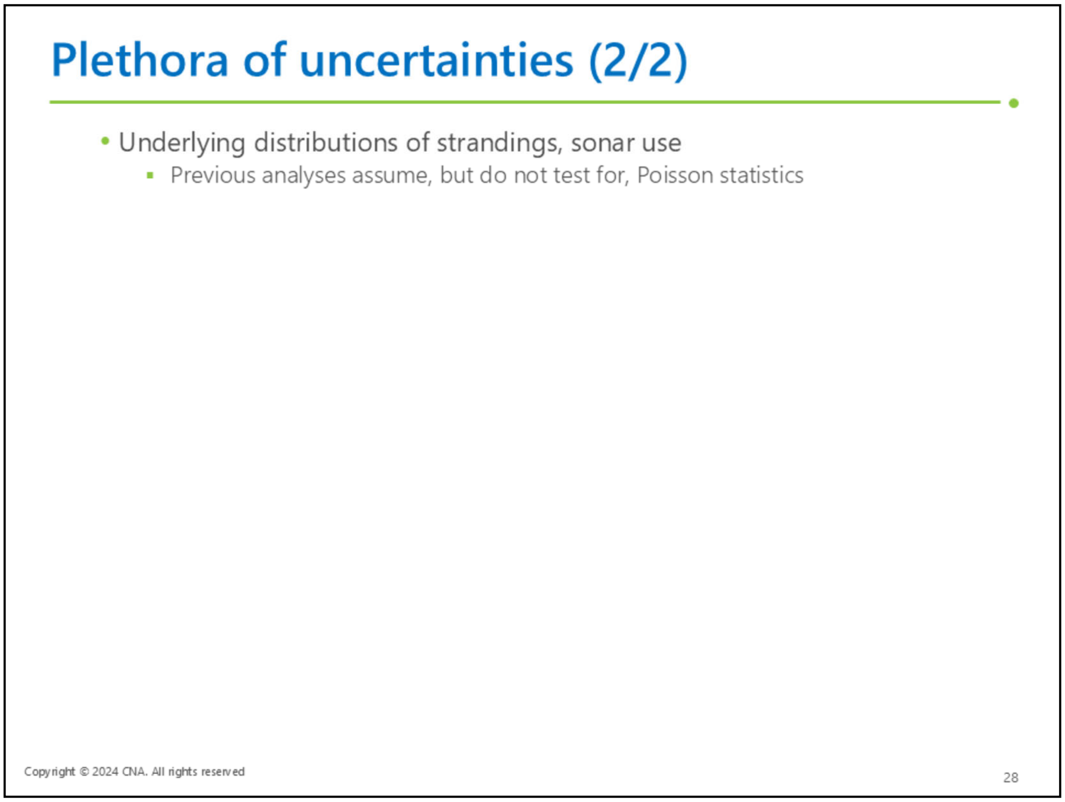
27

This slide summarizes key sources of uncertainty. However, we emphasize that whatever the degree to which these uncertainties (individually or collectively) may curtail the veracity of stranding analysis, they only *compound* the limitations inherent in all statistics-based methods. (Uncertainties marked with 🟥 are discussed in detail on later slides.) Sources of uncertainty include the following:

- Paucity of reliable sonar data prior to 2006 and the variable quality of data contained in the (ostensibly more complete) Sonar Positional Reporting System (SPORTS) system[1]
- Possibly ambiguous or inconsistent criteria used to measure sonar days in datasets
- The possible existence of unreported strandings
- Lack of an unambiguous definition of *coincident strandings* (the rule-of-thumb is that a stranding is coincident with sonar if it occurs within six days and 60 nmi of the last sonar use, but these numbers are largely arbitrary)
- The confounding effects of other factors that are typically unaccounted for in sonar-stranding time-series datasets, such as seasonality, seismic events, and the presence of fringing reefs. Simpson's paradox is particularly relevant here, if only as a heuristic reminder that statistics alone cannot tell the whole story. Simpson's paradox refers to a phenomenon whereby a correlation between two variables in a statistical population appears, disappears, or even "reverses if the population is divided into subpopulations ."[2] We can use batting averages as an example. A given player might have a higher batting average than another player each year for three years in a row. Yet, the other player might have a higher average for the three years as a whole. The reason is that the number of "at bats" changes from year to year, so this variable must be taken into account.

[1] J. Mintz, R. Filadelfo, and L. Bell, "Analysis of mid-frequency active (MFA) sonar use in Navy exercises using SPORTS," CNA, Research Memorandum, D0017310.A4, Jan 2008.

[2] "Simpson's Paradox," *Stanford Encyclopedia of Philosophy*, 24 March 2021, https://plato.Stanford.edu/entries/paradox-Simpson.



Plethora of uncertainties (2/2)

- Underlying distributions of strandings, sonar use
  - Previous analyses assume, but do not test for, Poisson statistics

28

Another source of uncertainty (albeit one that is arguably more of a technical limitation), is that past analyses typically assume, but do not explicitly test for, Poisson statistics. Although Poisson statistics are commonly used to describe count data generated by measuring the number of discrete events (such as the number of strandings) over a period of time, strictly speaking they are to be used only if a certain set of assumptions hold. Specifically, the process that generates the events must be homogenous in time (i.e., is memoryless), and the times between events must be independent and exponentially distributed.[1]

A simple test to see whether the underlying process is consistent with Poisson statistics is to estimate the values of the mean, μ, and variance, σ^2. For a Poisson process, it is easy to show that $\sigma^2 = \mu$. However, both *underdispersion* (wherein the observed variance, $\sigma^2 < \mu$, is significantly smaller than the expected variance) and *overdispersion* (when $\sigma^2 > \mu$) are possible in count data. In either case, the count data are inconsistent with Poisson statistics, and analysis requires alternative statistics. The negative binomial distribution is the most common model used to mitigate overdispersed data,[2] while the Conway-Maxwell Poisson (CMP)[3] and generalized Poisson (GP)[4] distributions may be used to model both underdispersed and overdispersed data.

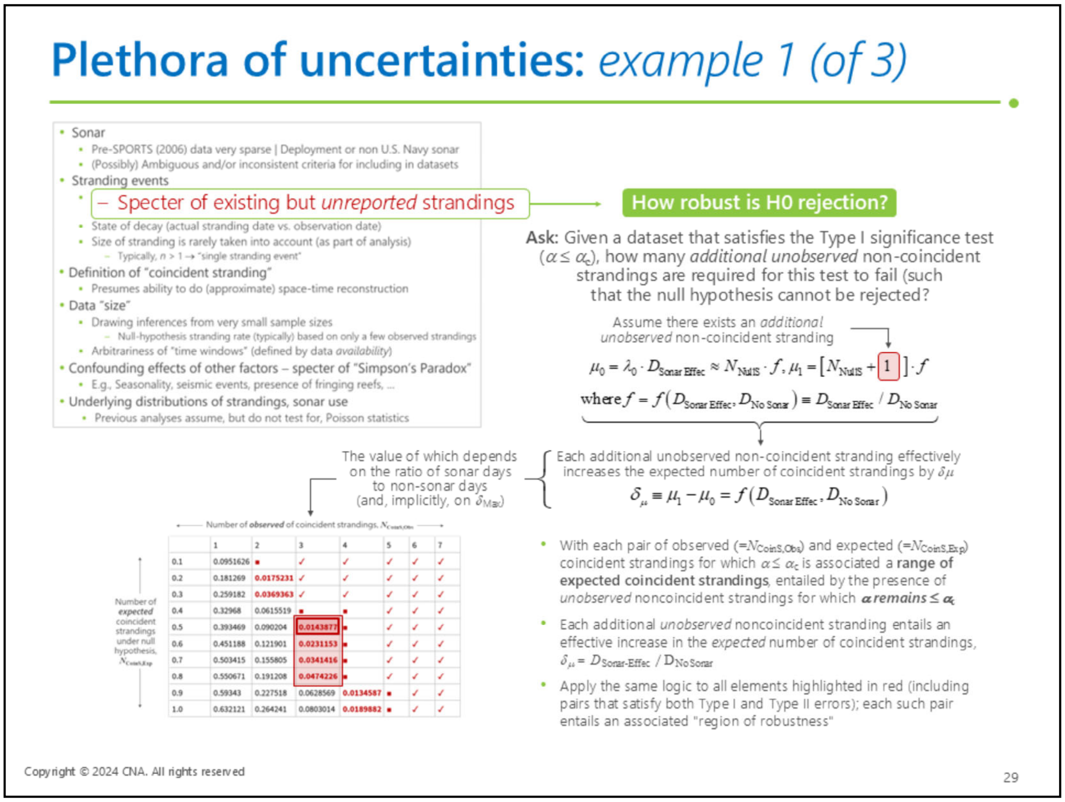
Because the stranding times in most of the real-world datasets used as case studies for the methods explored in this study are approximately exponentially distributed (as expected for a Poisson process), we did not pursue the analysis of alternative statistics tests. This being said, both underdispersed and overdispersed count data are certain to arise in future scenarios, for which a more comprehensive analysis will be needed.

[1] Sheldon M. Ross, *Introduction to Probability Models*, 13th Edition, Academic Press, 2023.

[2] E. Weisstein, "Negative Binomial Distribution," https://mathworld.wolfram.com/NegativeBinomialDistribution.html.

[3] Fraser Daly and R. E. Gaunt, “The Conway-Maxwell-Poisson distribution: distributional theory and approximation,” arXiv:1503.07012v2 [math.PR], 8 July 2016, https://arxiv.org/abs/1503.07012.

[4] Paulo C. Hubert, M. Lauretto, and J. Stern, “FBST for Generalized Poisson Distribution,” AIP Conference Proceedings, Vol. 1193, No. 210, 2009, https://philarchive.org/archive/STEFFA-5.



Plethora of uncertainties: *example 1 (of 3)*

- Sonar
  - Pre-SPORTS (2006) data very sparse | Deployment or non U.S. Navy sonar
  - (Possibly) Ambiguous and/or inconsistent criteria for including in datasets
- Stranding events
  - – Specter of existing but *unreported* strandings
  - State of decay (actual stranding date vs. observation date)
  - Size of stranding is rarely taken into account (as part of analysis)
    - Typically, $n > 1 \rightarrow$ "single stranding event"
- Definition of "coincident stranding"
  - Presumes ability to do (approximate) space-time reconstruction
- Data "size"
  - Drawing inferences from very small sample sizes
    - Null-hypothesis stranding rate (typically) based on only a few observed strandings
  - Arbitrariness of "time windows" (defined by data *availability*)
- Confounding effects of other factors – specter of "Simpson's Paradox"
  - E.g., Seasonality, seismic events, presence of fringing reefs, ...
- Underlying distributions of strandings, sonar use
  - Previous analyses assume, but do not test for, Poisson statistics

How robust is H0 rejection?

Ask: Given a dataset that satisfies the Type I significance test ($\alpha \leq \alpha_c$), how many *additional unobserved* non-coincident strandings are required for this test to fail (such that the null hypothesis cannot be rejected?

Assume there exists an *additional unobserved* non-coincident stranding

$$\underbrace{\mu_0 = \lambda_0 \cdot D_{\text{Sonar Effec}} \approx N_{\text{NullS}} \cdot f,\ \mu_1 = \left[N_{\text{NullS}} + \boxed{1}\,\right] \cdot f \quad \text{where } f = f\left(D_{\text{Sonar Effec}}, D_{\text{No Sonar}}\right) \equiv D_{\text{Sonar Effec}} / D_{\text{No Sonar}}}$$

The value of which depends on the ratio of sonar days to non-sonar days (and, implicitly, on δ_{Max})

Each additional unobserved non-coincident stranding effectively increases the expected number of coincident strandings by $\delta\mu$

$$\delta_\mu \equiv \mu_1 - \mu_0 = f\left(D_{\text{Sonar Effec}}, D_{\text{No Sonar}}\right)$$

Number of *observed* of coincident strandings, $N_{\text{CoinS,Obs}}$

Number of *expected* coincident strandings under null hypothesis, $N_{\text{CoinS,Exp}}$

	1	2	3	4	5	6	7
0.1	0.0951626	■	✓	✓	✓	✓	✓
0.2	0.181269	0.0175231	✓	✓	✓	✓	✓
0.3	0.259182	0.0369363	✓	✓	✓	✓	✓
0.4	0.32968	0.0615519	■	■	✓	✓	✓
0.5	0.393469	0.090204	0.0143877	■	✓	✓	✓
0.6	0.451188	0.121901	0.0231153	■	✓	✓	✓
0.7	0.503415	0.155805	0.0341416	■	✓	✓	✓
0.8	0.550671	0.191208	0.0474226	■	✓	✓	✓
0.9	0.59343	0.227518	0.0628569	0.0134587	■	✓	✓
1.0	0.632121	0.264241	0.0803014	0.0189882	■	✓	✓

- With each pair of observed (=$N_{\text{CoinS,Obs}}$) and expected (=$N_{\text{CoinS,Exp}}$) coincident strandings for which $\alpha \leq \alpha_c$ is associated a **range of expected coincident strandings**, entailed by the presence of *unobserved* noncoincident strandings for which ***α remains*** $\leq \alpha_c$
- Each additional *unobserved* noncoincident stranding entails an effective increase in the *expected* number of coincident strandings, $\delta_\mu = D_{\text{Sonar-Effec}} / \text{D}_{\text{No Sonar}}$
- Apply the same logic to all elements highlighted in red (including pairs that satisfy both Type I and Type II errors); each such pair entails an associated "region of robustness"

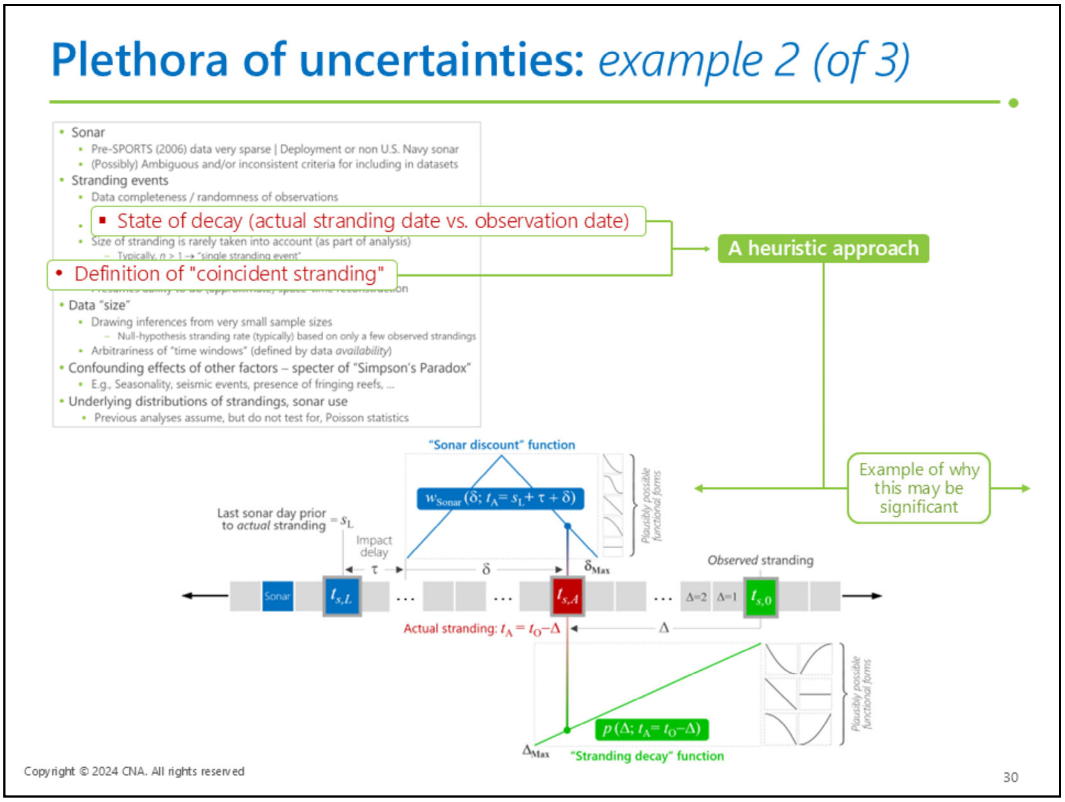
29

This slide takes a deeper look at the possible ramifications of one of the uncertainties described briefly on the preceding slide. Specifically, the specter of existing but unreported strandings.

If an unreported stranding was coincident with sonar, that can only strengthen the statistical evidence for correlation. But, what if there was an unreported *non-coincident* stranding? Mathematically, if additional null strandings exist, their presence would effectively increase the expected number of coincident strandings, which in turn would increase the minimal number of coincident strandings that would have to be observed to satisfy Type I and Type II tests (as discussed earlier).

Assume that for a given scenario, we have determined there is sufficient statistical evidence to reject the null hypothesis. Absent knowing whether any unobserved strandings occurred, we can still ask, How *robust* is the evidence used to reject the null hypothesis? That is, would we still reject the null hypothesis had an additional null stranding been included? What about an additional two null strandings?

The statistical evidence used to support rejecting the null hypothesis is strengthened if it remains robust to additional hypothetical null strandings.



Plethora of uncertainties: *example 2 (of 3)*

- Sonar
  - Pre-SPORTS (2006) data very sparse | Deployment or non U.S. Navy sonar
  - (Possibly) Ambiguous and/or inconsistent criteria for including in datasets
- Stranding events
  - Data completeness / randomness of observations
  - ▪ State of decay (actual stranding date vs. observation date)
  - Size of stranding is rarely taken into account (as part of analysis)
    - Typically, $n > 1 \rightarrow$ "single stranding event"
- • Definition of "coincident stranding"
  - Presumes ability to do (approximate) space-time reconstruction
- Data "size"
  - Drawing inferences from very small sample sizes
    - Null-hypothesis stranding rate (typically) based on only a few observed strandings
  - Arbitrariness of "time windows" (defined by data *availability*)
- Confounding effects of other factors – specter of "Simpson's Paradox"
  - E.g., Seasonality, seismic events, presence of fringing reefs, ...
- Underlying distributions of strandings, sonar use
  - Previous analyses assume, but do not test for, Poisson statistics

A heuristic approach

Example of why this may be significant

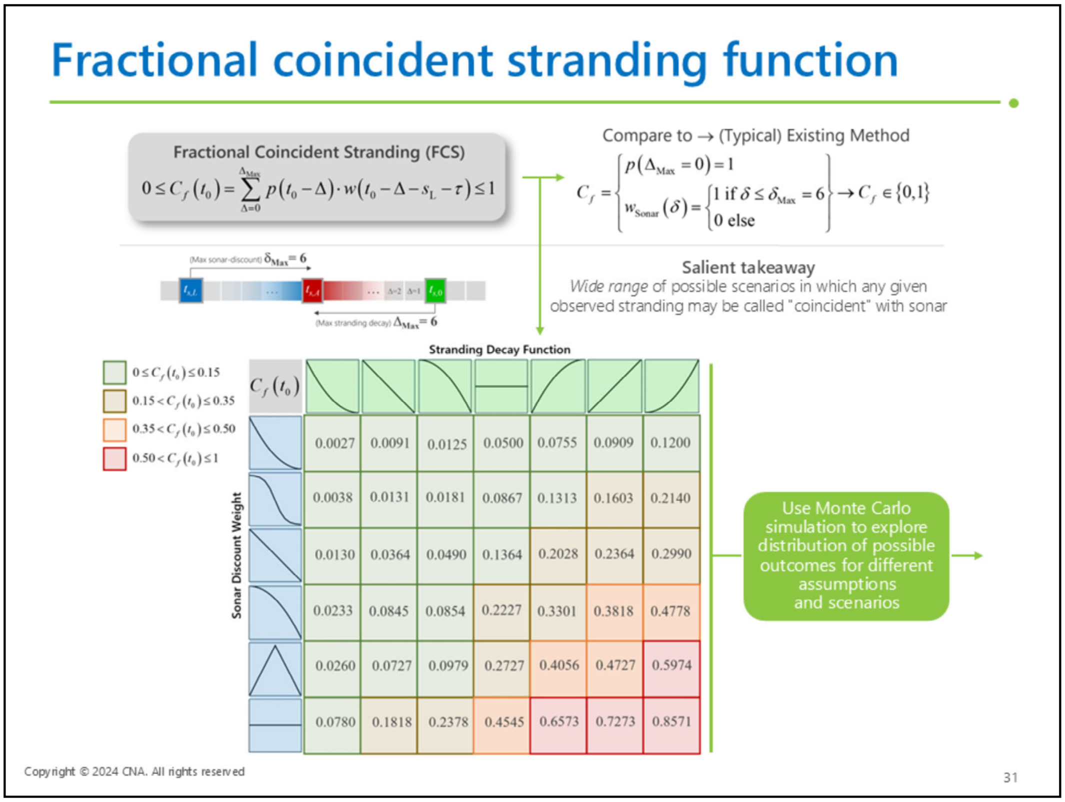
"Sonar discount" function
$w_{\text{Sonar}}(\delta; t_A = s_L + \tau + \delta)$
Plausibly possible functional forms
Last sonar day prior to *actual* stranding $= s_L$
Impact delay
τ
δ
δ_{Max}
Observed stranding
Sonar
$t_{s,L}$
$t_{s,A}$
Δ=2
Δ=1
$t_{s,0}$
Actual stranding: $t_A = t_O - \Delta$
Δ
$p(\Delta; t_A = t_O - \Delta)$
Δ_{Max}
"Stranding decay" function
Plausibly possible functional forms

30

Two immediate uncertainties associated with strandings are as follows: (1) not knowing *when* the actual stranding occurred, considering that we typically know only when a given stranding was *observed* (or reported), and (2) the imprecise (and, hence, ambiguous) manner in which *coincidence* is defined. This slide illustrates schematically how these two classes of uncertainty may be parameterized. Later slides will show how this formalism may be used to refine estimates of significance and statistical power.

Standard practice is to call a stranding a "coincident stranding" if two conditions are met: (1) no more than six days have elapsed since the last day sonar was used prior to the stranding, and (2) the distance between where that last sonar was used and where the stranding occurred was no more than 60 nmi. Four variables are in play:

1. The time delay, δ_x, between the last sonar day and the stranding, which is nominally set to six days but may, in principle, assume a range of other reasonable values.
2. The time delay between the actual versus observed (or reported) stranding dates, Δ_x. If necropsy data are available (which is not typical), we may assume that Δ_x will obviously depend on the state of decay of the stranded mammal (see **Appendix F**).
3. The functional form of a *stranding decay function* that (loosely speaking) represents the probability that a stranding observed on day $t_{s,0}$ actually occurred on day $t_{s,A}$.
4. The functional form of a *sonar discount function* that (loosely speaking) represents the probability that a stranding on day $t_{s,A}$ is a coincident stranding, given that the last sonar day occurred on day $t_{s,L}$. The slide also shows a few possible functional forms for these latter two functions.



Fractional coincident stranding function

Fractional Coincident Stranding (FCS)

$$0 \leq C_f(t_0) = \sum_{\Delta=0}^{\Delta_{\text{Max}}} p(t_0 - \Delta) \cdot w(t_0 - \Delta - s_{\text{L}} - \tau) \leq 1$$

Compare to → (Typical) Existing Method

$$C_f = \left\{ \begin{matrix} p(\Delta_{\text{Max}} = 0) = 1 \\ w_{\text{Sonar}}(\delta) = \begin{cases} 1 \text{ if } \delta \leq \delta_{\text{Max}} = 6 \\ 0 \text{ else} \end{cases} \end{matrix} \right\} \rightarrow C_f \in \{0,1\}$$

(Max sonar-discount) $\delta_{\text{Max}} = 6$
$t_{s,L}$
…
$t_{s,A}$
…
Δ=2
Δ=1
$t_{s,0}$
(Max stranding decay) $\Delta_{\text{Max}} = 6$

Salient takeaway

Wide range of possible scenarios in which any given observed stranding may be called "coincident" with sonar

🟩 $0 \leq C_f(t_0) \leq 0.15$

🟫 $0.15 < C_f(t_0) \leq 0.35$

🟧 $0.35 < C_f(t_0) \leq 0.50$

🟥 $0.50 < C_f(t_0) \leq 1$

Stranding Decay Function

Sonar Discount Weight

$C_f(t_0)$							
	0.0027	0.0091	0.0125	0.0500	0.0755	0.0909	0.1200
	0.0038	0.0131	0.0181	0.0867	0.1313	0.1603	0.2140
	0.0130	0.0364	0.0490	0.1364	0.2028	0.2364	0.2990
	0.0233	0.0845	0.0854	0.2227	0.3301	0.3818	0.4778
	0.0260	0.0727	0.0979	0.2727	0.4056	0.4727	0.5974
	0.0780	0.1818	0.2378	0.4545	0.6573	0.7273	0.8571

Use Monte Carlo simulation to explore distribution of possible outcomes for different assumptions and scenarios

31

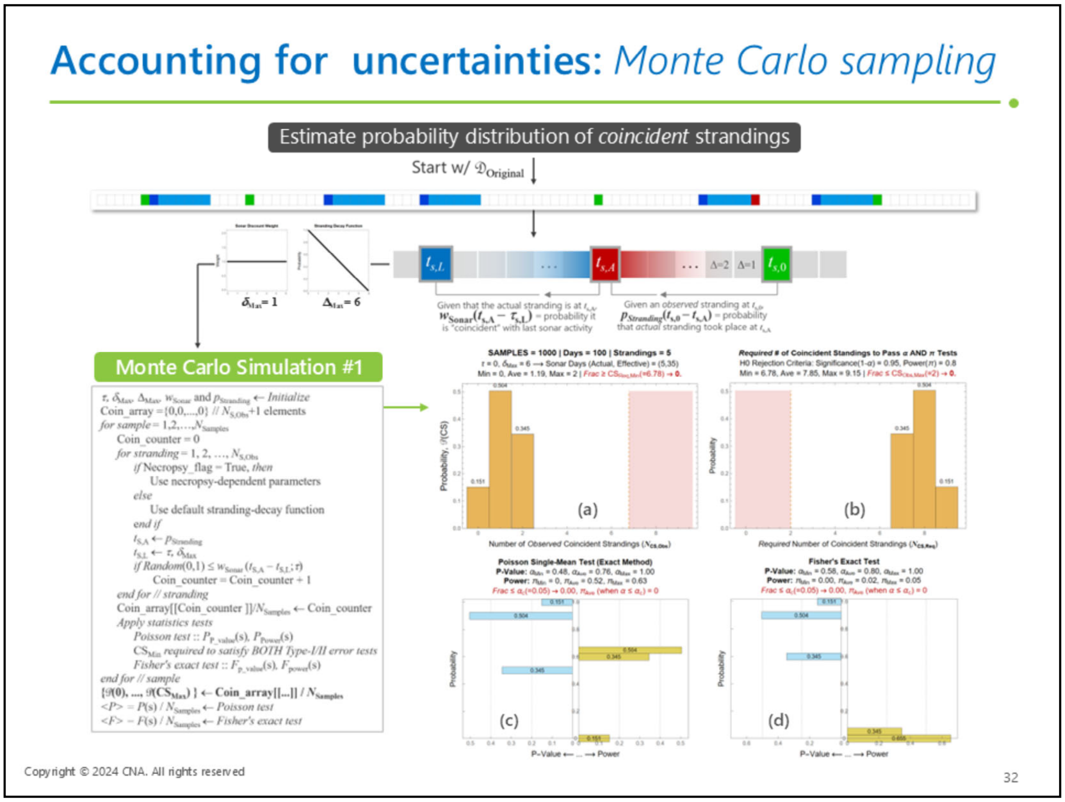
This slide is designed to motivate a statistical method (introduced on the following two slides) that directly accommodates uncertainties associated with actual versus observed stranding dates and ambiguity in tagging a given stranding as "coincident with sonar."

The fractional coincident stranding (FCS), $C_f(t_0)$, of a stranding observed at time t_0 uses the formalism introduced on the previous slide to generalize the conventional binary-valued interpretation of coincidence (i.e., as either coincident with sonar or not) to include fractional values. The idea is to elicit a sense of how strongly these two classes of uncertainties may influence statistical analyses.

The matrix contains FCS values for an illustrative scenario in which δ_x and Δ_x both equal six days. The columns and rows denote different functional forms that may be used to represent the stranding decay and sonar discount functions, respectively. Individual entries are color coded according to the legend that appears at the top left of the matrix.

The takeaway is that *there is a wide spectrum of FCS values*, ranging from values that are close to one (which is the *only* possible value for strandings identified as coincident using conventional methods) to those that are close to zero. Of course, we do not know which, if any, of the possible functional forms most closely match reality, nor do we know which values for δ_x and Δ_x are best to use. But this is the salient point. Absent such knowledge, the range of FCS values deduced from a set of plausible functional forms and parameter values gives a measure of uncertainty we expect to find in our statistical inferences (as drawn under specific assumptions). The tighter the range of FCS values, the more robust we expect our inferences to be (at least with respect to this particular class of uncertainties).

The next two slides introduce a MCS approach that leverages these ideas to explicitly account for date uncertainties and *coincident* definition ambiguities.



Accounting for uncertainties: *Monte Carlo sampling*

Estimate probability distribution of *coincident* strandings

Start w/ $\mathfrak{D}_{\text{Original}}$

Given that the actual stranding is at $t_{s,A}$, $w_{\text{Sonar}}(t_{s,A} - \tau_{s,L})$ = probability it is "coincident" with last sonar activity

Given an *observed* stranding at $t_{s,0}$, $p_{Stranding}(t_{s,0} - t_{s,A})$ = probability that *actual* stranding took place at $t_{s,A}$

Monte Carlo Simulation #1

```
$\tau$, $\delta_{\text{Max}}$, $\Delta_{\text{Max}}$, $w_{\text{Sonar}}$ and $p_{\text{Stranding}}$ ← Initialize
Coin_array = {0,0,...,0} // $N_{\text{S,Obs}}$+1 elements
for sample = 1,2,…,$N_{\text{Samples}}$
    Coin_counter = 0
    for stranding = 1, 2, …, $N_{\text{S,Obs}}$
        if Necropsy_flag = True, then
            Use necropsy-dependent parameters
        else
            Use default stranding-decay function
        end if
        $t_{\text{S,A}} \leftarrow p_{\text{Stranding}}$
        $t_{\text{S,L}} \leftarrow \tau, \delta_{\text{Max}}$
        if Random(0,1) $\le w_{\text{Sonar}}(t_{\text{S,A}} - t_{\text{S,L}}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}}$ ← Coin_counter
    Apply statistics tests
        Poisson test :: $P_{\text{P\_value}}$(s), $P_{\text{Power}}$(s)
        $\text{CS}_{\text{Min}}$ required to satisfy BOTH Type-I/II error tests
        Fisher's exact test :: $F_{\text{P\_value}}$(s), $F_{\text{power}}$(s)
end for // sample
{$\mathcal{P}$(0), ..., $\mathcal{P}$($\text{CS}_{\text{Max}}$) } ← Coin_array[[...]] / $N_{\text{Samples}}$
<P> = P(s) / $N_{\text{Samples}}$ ← Poisson test
<F> = F(s) / $N_{\text{Samples}}$ ← Fisher's exact test
```

SAMPLES = 1000 | Days = 100 | Strandings = 5
τ = 0, δ_{Max} = 6 ⟶ Sonar Days (Actual, Effective) = (5,35)
Min = 0, Ave = 1.19, Max = 2 | *Frac ≥ $\text{CS}_{\text{Req,Min}}$(=6.78) → 0.*

Required # of Coincident Standings to Pass α AND π Tests
H0 Rejection Criteria: Significance(1-α) = 0.95, Power(π) = 0.8
Min = 6.78, Ave = 7.85, Max = 9.15 | *Frac ≤ $\text{CS}_{\text{Obs,Max}}$(=2) → 0.*

Poisson Single-Mean Test (Exact Method)
P-Value: α_{Min} = 0.48, α_{Ave} = 0.76, α_{Max} = 1.00
Power: π_{Min} = 0, π_{Ave} = 0.52, π_{Max} = 0.63
Frac ≤ α_c(=0.05) → 0.00, π_{Ave} (when $\alpha \le \alpha_c$) = 0

Fisher's Exact Test
P-Value: α_{Min} = 0.58, α_{Ave} = 0.80, α_{Max} = 1.00
Power: π_{Min} = 0.00, π_{Ave} = 0.02, π_{Max} = 0.05
Frac ≤ α_c(=0.05) → 0.00, π_{Ave} (when $\alpha \le \alpha_c$) = 0

32

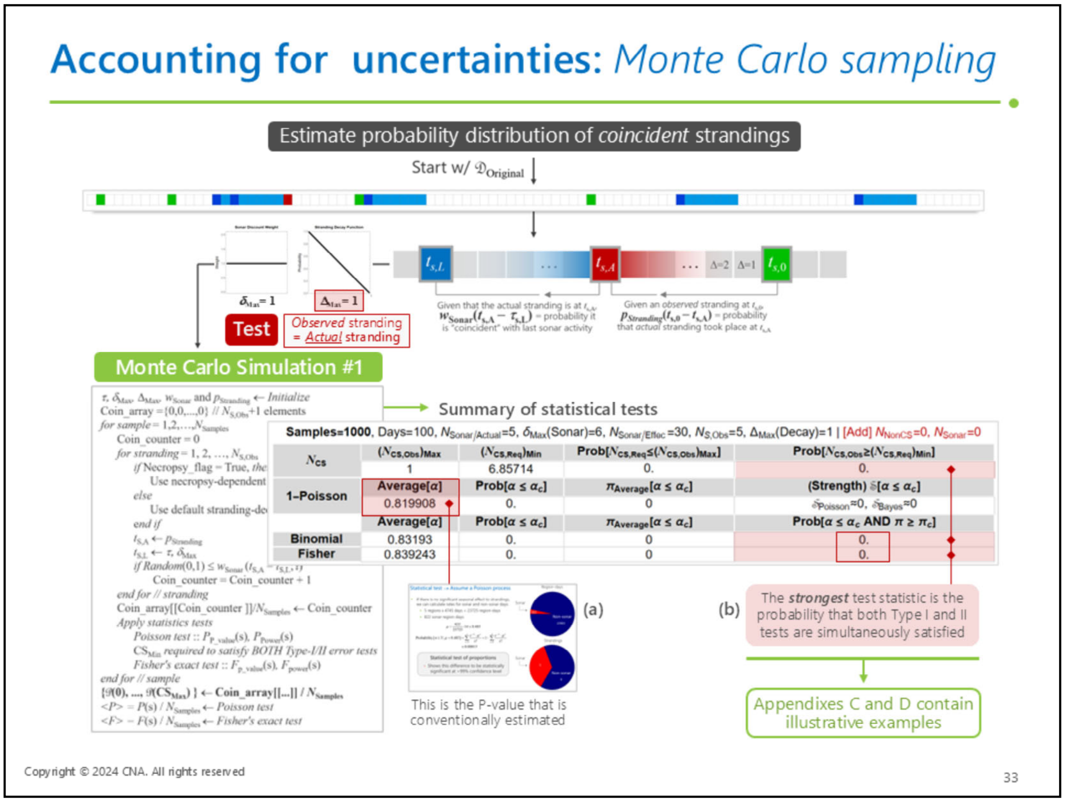
This slide introduces the first of several MCS we developed for this study. The general idea was to average the inferential results of a battery of statistical tests (e.g., the single-mean Poisson test, FET, and EBT) over a large sample of datasets that were randomized over various classes of uncertainty.

The pseudocode for the first MCS appears on the left of the slide. Runs are initialized with the original dataset, $\mathfrak{D}_{\text{Original}}$, as depicted graphically at the top of the slide. The user selects values for δ_x and Δ_x, along with the functional forms for the stranding decay and sonar discount functions. The parameters that define the stranding decay function may be based on the state of decay of stranded animals if necropsy data are available (see **Appendix F**). The simulation loops through N_{Samples}, where for each sample, an actual stranding date (as fixed in $\mathfrak{D}_{\text{Original}}$) is randomly determined using the stranding decay function and is labeled as "coincident" with sonar or "not coincident" probabilistically, according to the sonar discount function. A battery of statistical tests is applied to each randomized dataset sample, the results of which are averaged over all samples after the run is complete and then archived.

Multiple types of outputs are available. This slide shows a graphical output in which histograms are used to display the distributions of results: (a) frequency of the number of observed coincident strandings, (b) frequency of the required number of coincident strandings (to satisfy Type I and Type II tests; see slides 18 to 22), (c) frequency of the P-value and power of the single-mean Poisson test, and (d) frequency of the P-value and power of FET (a histogram of the results of applying the EBT is optional).

A set of relevant summary statistics appears along the top of each histogram. Key statistics are highlighted in **red** at the top of each histogram: (a) fraction of runs in which the observed number of coincident strandings is at least as great as the required number, (b) fraction of runs in which the required number of coincident strandings is less than or equal to the observed number, (c) average power when $\alpha \le \alpha_c$ for the single-mean Poisson test, and (d) average power when $\alpha \le \alpha_c$ for the FET.

*Additional details and code fragments appear in **Appendix I**.*



Accounting for uncertainties: *Monte Carlo sampling*

Estimate probability distribution of *coincident* strandings

Start w/ $\mathfrak{D}_{\text{Original}}$
$\delta_{\text{Max}} = 1$
$\Delta_{\text{Max}} = 1$
$t_{s,L}$
$t_{s,A}$
Δ=2
Δ=1
$t_{s,0}$

Given that the actual stranding is at $t_{s,A}$, $w_{\text{Sonar}}(t_{s,A} - \tau_{s,L})$ = probability it is "coincident" with last sonar activity

Given an *observed* stranding at $t_{s,0}$, $p_{\text{Stranding}}(t_{s,0} - t_{s,A})$ = probability that *actual* stranding took place at $t_{s,A}$

Test

Observed stranding = *Actual* stranding

Monte Carlo Simulation #1

```
$\tau$, $\delta_{\text{Max}}$, $\Delta_{\text{Max}}$, $w_{\text{Sonar}}$ and $p_{\text{Stranding}}$ ← Initialize
Coin_array ={0,0,...,0} // $N_{\text{S,Obs}}$+1 elements
for sample = 1,2,…,$N_{\text{Samples}}$
    Coin_counter = 0
    for stranding = 1, 2, …, $N_{\text{S,Obs}}$
        if Necropsy_flag = True, the
            Use necropsy-dependent
        else
            Use default stranding-de
        end if
        $t_{\text{S,A}}$ ← $p_{\text{Stranding}}$
        $t_{\text{S,L}}$ ← $\tau$, $\delta_{\text{Max}}$
        if Random(0,1) ≤ $w_{\text{Sonar}}(t_{\text{S,A}} - \tau_{\text{S,L}}, t)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}}$ ← Coin_counter
    Apply statistics tests
        Poisson test :: $P_{\text{P\_value}}(s)$, $P_{\text{Power}}(s)$
        $CS_{\text{Min}}$ required to satisfy BOTH Type-I/II error tests
        Fisher's exact test :: $F_{\text{p\_value}}(s)$, $F_{\text{power}}(s)$
end for // sample
{$\mathcal{P}(0)$, ..., $\mathcal{P}(CS_{\text{Max}})$ } ← Coin_array[[...]] / $N_{\text{Samples}}$
<P> = P(s) / $N_{\text{Samples}}$ ← Poisson test
<F> = F(s) / $N_{\text{Samples}}$ ← Fisher's exact test
```

Summary of statistical tests

Samples=1000, Days=100, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$=30, $N_{\text{S,Obs}}$=5, Δ_{Max}(Decay)=1 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \leq (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \geq (N_{\text{CS,Req}})_{\text{Min}}$]
	1	6.85714	0.	0.
1–Poisson	Average[α]	Prob[$\alpha \leq \alpha_c$]	$\pi_{\text{Average}}[\alpha \leq \alpha_c]$	(Strength) $\mathbb{S}[\alpha \leq \alpha_c]$
	0.819908	0.	0	$\mathbb{S}_{\text{Poisson}} \approx 0$, $\mathbb{S}_{\text{Bayes}} \approx 0$
	Average[α]	Prob[$\alpha \leq \alpha_c$]	$\pi_{\text{Average}}[\alpha \leq \alpha_c]$	Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]
Binomial	0.83193	0.	0	0.
Fisher	0.839243	0.	0	0.

(a)

This is the P-value that is conventionally estimated

(b)

The ***strongest*** test statistic is the probability that both Type I and II tests are simultaneously satisfied

Appendixes C and D contain illustrative examples

33

This slide shows an optional text-based summary of MCS runs.

The top row summarizes key features of the original dataset: the number of Monte Carlo samples, total number of days, actual ($N_{\text{Sonar/Actual}}$) and effective ($N_{\text{sonar/Effective}}$) number of days with sonar (the latter is a function of the maximum sonar decay range (δ_{Max}), and total number of observed strandings ($N_{\text{S,Obs}}$).

The first row (labeled "N_{CS}") contains the average value for the maximum number of observed coincident strandings $(N_{\text{CS,Obs}})_{\text{Max}}$, the average value of the required minimum number of coincident strandings to simultaneously satisfy both Type I and Type II tests $(N_{\text{CS,Req}})_{\text{Min}}$, the probability that the required number of coincident strandings to satisfy both Type I and Type II tests is less than or equal to the maximum observed number (Prob[$N_{\text{CS,Req}} \leq (N_{\text{CS,Obs}})_{\text{Max}}$]), and the probability that the observed number of coincident strandings is less than or equal to the minimum required number to satisfy both Type I and Type II tests (Prob[$N_{\text{CS,Obs}} \geq (N_{\text{CS,Req}})_{\text{Min}}$]).

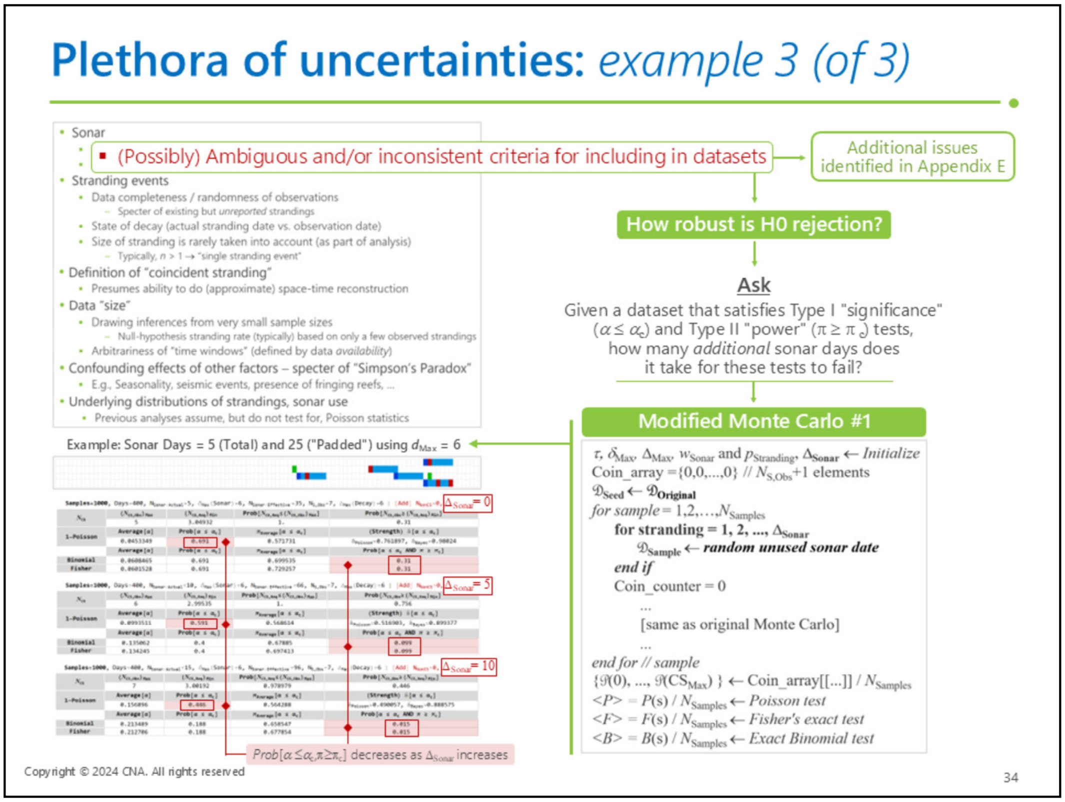
The second row, labeled "1-Poisson," contains the results of applying the single-mean Poisson test: the average value of α, the probability that $\alpha \leq \alpha_c$, the average power when $\alpha \leq \alpha_c$, and the strength of runs when $\alpha \leq \alpha_c$ (*strength* is defined in **Appendix C**).

The third and fourth rows, labeled "binomial" and "Fisher," summarize the results of applying the EBT and FET, respectively. Columns one to three are the same as for the single-mean Poisson test, but the last column, labeled "Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]," gives the probability that both Type I and Type II tests are satisfied for each of the two tests.

The area marked with "(a)" shows the field that contains the P-value that is estimated in most studies, and the area marked with "(b)" shows arguably the strongest test statistics that represent the probability that both Type I and Type II tests will be simultaneously satisfied.

Additional details about specific elements are given in **Appendix C.**

.



Plethora of uncertainties: *example 3 (of 3)*

- Sonar
  - ▪ (Possibly) Ambiguous and/or inconsistent criteria for including in datasets
- Stranding events
  - Data completeness / randomness of observations
    - Specter of existing but *unreported* strandings
  - State of decay (actual stranding date vs. observation date)
  - Size of stranding is rarely taken into account (as part of analysis)
    - Typically, $n > 1 \rightarrow$ "single stranding event"
- Definition of "coincident stranding"
  - Presumes ability to do (approximate) space-time reconstruction
- Data "size"
  - Drawing inferences from very small sample sizes
    - Null-hypothesis stranding rate (typically) based on only a few observed strandings
  - Arbitrariness of "time windows" (defined by data *availability*)
- Confounding effects of other factors – specter of "Simpson's Paradox"
  - E.g., Seasonality, seismic events, presence of fringing reefs, ...
- Underlying distributions of strandings, sonar use
  - Previous analyses assume, but do not test for, Poisson statistics

Additional issues identified in Appendix E

How robust is H0 rejection?

Ask

Given a dataset that satisfies Type I "significance" ($\alpha \leq \alpha_c$) and Type II "power" ($\pi \geq \pi_c$) tests, how many *additional* sonar days does it take for these tests to fail?

Modified Monte Carlo #1

```
$\tau, \delta_{Max}, \Delta_{Max}, w_{Sonar}$ and $p_{Stranding}, \Delta_{Sonar}$ ← Initialize
Coin_array = {0,0,...,0} // $N_{S,Obs}$+1 elements
$\mathfrak{D}_{Seed} \leftarrow \mathfrak{D}_{Original}$
for sample = 1,2,…,$N_{Samples}$
    for stranding = 1, 2, ..., $\Delta_{Sonar}$
        $\mathfrak{D}_{Sample}$ ← random unused sonar date
    end if
    Coin_counter = 0
        ...
        [same as original Monte Carlo]
        ...
end for // sample
{$\mathfrak{P}(0)$, ..., $\mathfrak{P}(CS_{Max})$ } ← Coin_array[[...]] / $N_{Samples}$
<P> = P(s) / $N_{Samples}$ ← Poisson test
<F> = F(s) / $N_{Samples}$ ← Fisher's exact test
<B> = B(s) / $N_{Samples}$ ← Exact Binomial test
```

Example: Sonar Days = 5 (Total) and 25 ("Padded") using $d_{Max} = 6$

Samples=1000, Days=400, $N_{Sonar\text{-}Actual}$=5, d_{Max}(Sonar)=6, $N_{Sonar\text{-}Effective}$=35, $N_{S,Obs}$=7, d_{Max}(Decay)=6 | [Add] N_{Sonar}=0, $\Delta_{Sonar} = 0$

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Avg})_{Min}$	Prob$[N_{CS,Avg} \leq (N_{CS,Obs})_{Max}]$	Prob$[N_{CS,Obs} \geq (N_{CS,Avg})_{Min}]$
	5	3.04932	1.	0.31
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}[\alpha \leq \alpha_c]$	(Strength) $\delta[\alpha \leq \alpha_c]$
	0.0453349	0.691	0.571731	$\delta_{Poisson}$=0.761897, δ_{Bayes}=0.98024
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}[\alpha \leq \alpha_c]$	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.0608465	0.691	0.699535	0.31
Fisher	0.0601528	0.691	0.729257	0.31

Samples=1000, Days=400, $N_{Sonar\text{-}Actual}$=10, d_{Max}(Sonar)=6, $N_{Sonar\text{-}Effective}$=66, $N_{S,Obs}$=7, d_{Max}(Decay)=6 | [Add] N_{Sonar}=0, $\Delta_{Sonar} = 5$

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Avg})_{Min}$	Prob$[N_{CS,Avg} \leq (N_{CS,Obs})_{Max}]$	Prob$[N_{CS,Obs} \geq (N_{CS,Avg})_{Min}]$
	6	2.99535	1.	0.756
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}[\alpha \leq \alpha_c]$	(Strength) $\delta[\alpha \leq \alpha_c]$
	0.0993511	0.591	0.568614	$\delta_{Poisson}$=0.516903, δ_{Bayes}=0.899377
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}[\alpha \leq \alpha_c]$	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.135062	0.4	0.67885	0.099
Fisher	0.134245	0.4	0.697413	0.099

Samples=1000, Days=400, $N_{Sonar\text{-}Actual}$=15, d_{Max}(Sonar)=6, $N_{Sonar\text{-}Effective}$=96, $N_{S,Obs}$=7, d_{Max}(Decay)=6 | [Add] N_{Sonar}=0, $\Delta_{Sonar} = 10$

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Avg})_{Min}$	Prob$[N_{CS,Avg} \leq (N_{CS,Obs})_{Max}]$	Prob$[N_{CS,Obs} \geq (N_{CS,Avg})_{Min}]$
	7	3.00192	0.978979	0.446
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}[\alpha \leq \alpha_c]$	(Strength) $\delta[\alpha \leq \alpha_c]$
	0.156896	0.486	0.564288	$\delta_{Poisson}$=0.490057, δ_{Bayes}=0.888575
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}[\alpha \leq \alpha_c]$	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.213489	0.188	0.658547	0.015
Fisher	0.212706	0.188	0.677854	0.015

$Prob[\alpha \leq \alpha_c, \pi \geq \pi_c]$ decreases as Δ_{Sonar} increases

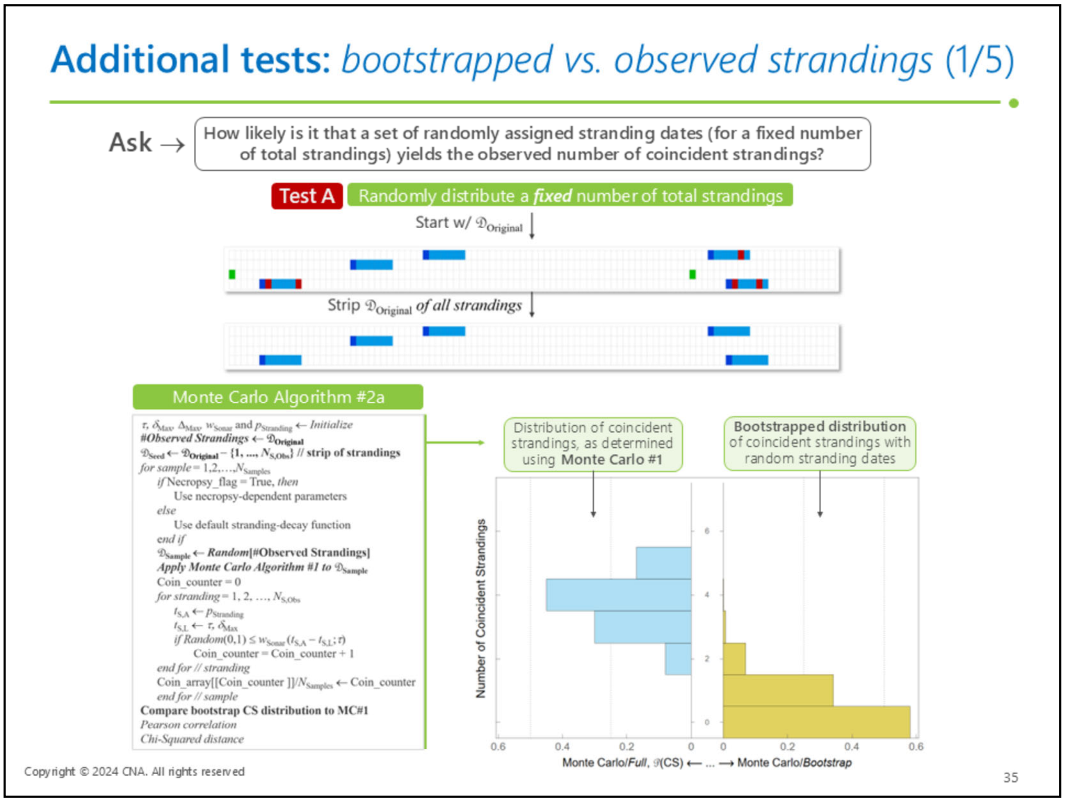
34

This slide takes a deeper look at how MCS may be used to mitigate another type of uncertainty: the ambiguous or inconsistent criteria that are often used to include or exclude sonar days from a given dataset. In order to properly prepare a dataset, the analyst must include all known sonar activity that could, in principle, be correlated with observed strandings. It is this requirement that the 60 nmi and six day space and time windows are designed to (loosely) capture. But, what if additional sonar days are left out of the dataset (and, thus, are unaccounted for)? What if non-US sonar was active near the same operating area?

One possible approach to mitigating the uncertainty introduced by having to estimate answers to such questions is to use a MCS (or, more precisely, a modified form of the MCS introduced on the previous two slides) to test how robust baseline "reject null hypothesis" inferences are to additional unaccounted for sonar days.

Heuristically, as the number of sonar days increases, the likelihood of observing a given number of *expected* coincident strandings (as defined by the null stranding rate, which remains fixed because we are probing only the effect of adding days during which sonar is active, not when it is inactive) also increases, which in turn decreases the statistical evidence sufficient to reject the null hypothesis. A rejection is *robust* if it remains unchanged when a certain test number of sonar days are randomly inserted into the original dataset prior to applying the battery of statistical tests.

Appendix E identifies and discusses additional (subtler) issues that would take too long to discuss in the main narrative but that may help refine the way that *null* and *coincident* stranding rates are defined in future studies.



Additional tests: *bootstrapped vs. observed strandings* (1/5)

Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test A Randomly distribute a ***fixed*** number of total strandings

Start w/ $\mathfrak{D}_{\text{Original}}$

Strip $\mathfrak{D}_{\text{Original}}$ *of all strandings*

Monte Carlo Algorithm #2a

```
$\tau$, $\delta_{Max}$, $\Delta_{Max}$, $w_{Sonar}$ and $p_{Stranding}$ ← Initialize
#Observed Strandings ← $\mathfrak{D}_{Original}$
$\mathfrak{D}_{Seed}$ ← $\mathfrak{D}_{Original}$ − {1, ..., $N_{S,Obs}$} // strip of strandings
for sample = 1,2,…,$N_{Samples}$
    if Necropsy_flag = True, then
        Use necropsy-dependent parameters
    else
        Use default stranding-decay function
    end if
    $\mathfrak{D}_{Sample}$ ← Random[#Observed Strandings]
    Apply Monte Carlo Algorithm #1 to $\mathfrak{D}_{Sample}$
    Coin_counter = 0
    for stranding = 1, 2, …, $N_{S,Obs}$
        $t_{S,A}$ ← $p_{Stranding}$
        $t_{S,L}$ ← $\tau$, $\delta_{Max}$
        if Random(0,1) ≤ $w_{Sonar}(t_{S,A} - t_{S,L}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{Samples}$ ← Coin_counter
    end for // sample
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

Distribution of coincident strandings, as determined using **Monte Carlo #1**

Bootstrapped distribution of coincident strandings with random stranding dates

Number of Coincident Strandings
6
4
2
0
0.6
0.4
0.2
0
0
0.2
0.4
0.6
Monte Carlo/*Full*, 𝒫(CS) ⟵ ... ⟶ Monte Carlo/*Bootstrap*

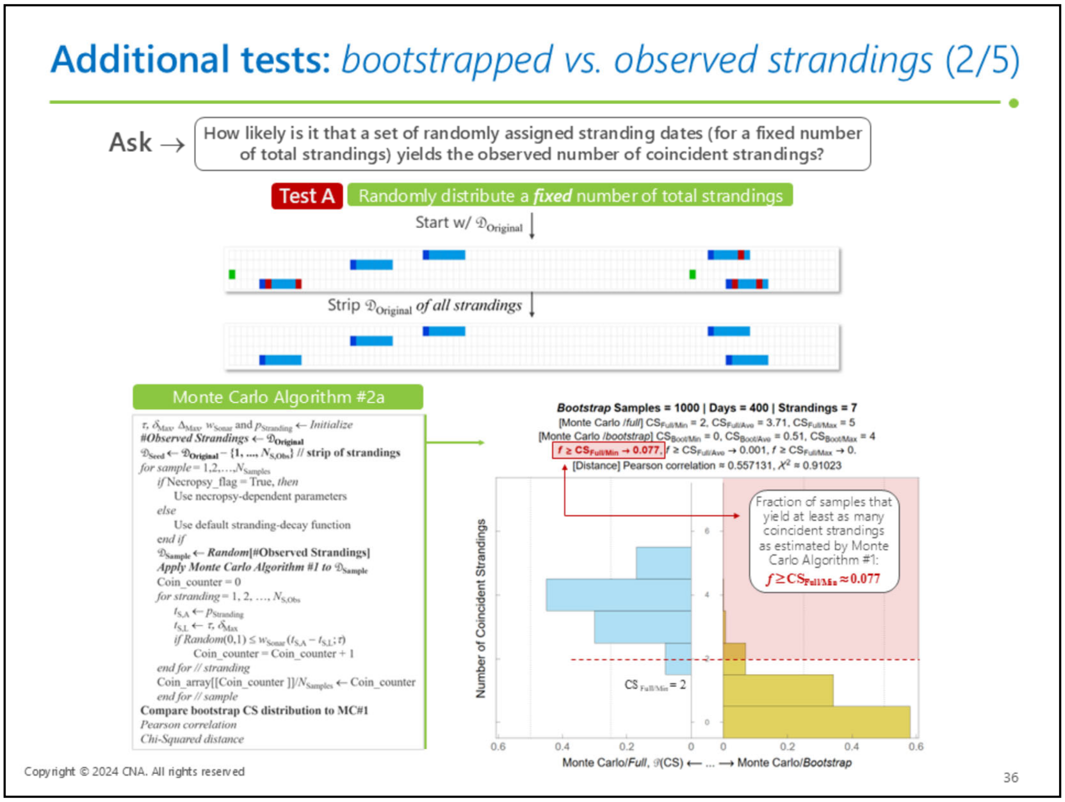
35

We next discuss two additional tests that may be used to determine the likelihood of observing a given number of coincident strandings (as expected from the null stranding rate). These are not statistical tests per se (that is, they do not directly support rejecting the null hypothesis); rather, they are simple MCS that estimate the probability that a random sample of a given number of total strandings includes (at least) the same number of coincident strandings as were actually observed. This approach is motivated by the recent study by Simonis et al., in which just such a comparison is made to illustrate (in their Mariana Islands, Western Pacific scenario) the "small probability of any stranding events occurring within the [coincidence] window" [27].

For Test A, we follow Simonis et al. and modify MCS 1 as described on slide 30 according to the lines highlighted in bold text in the pseudocode that appears on the bottom left of the slide above. The key change involves stripping the original dataset, $\mathfrak{D}_{\text{Original}}$, of all strandings and then running MCS 1 on the stripped dataset (that retains only the fixed set of sonar days but is otherwise empty) to which the total number of strandings (of any kind, as they appear in $\mathfrak{D}_{\text{Original}}$) are assigned random dates.

The output consists of a comparison of histograms (illustrated on the lower right of the slide). The histogram on the left—highlighted in **aqua**—shows the distribution of coincident strandings as determined using MCS 1 (wherein randomization is introduced strictly by applying the *stranding decay* and *sonar discount* functions to an otherwise fixed $\mathfrak{D}_{\text{original}}$). The histogram on the right—highlighted in **gold**—shows the distribution of coincident strandings as determined using the modified form of MCS 1 (namely, MCS 2a).

The next slide shows the complete form of the output that includes overlays of statistics that help quantify what (at first inspection) is a purely qualitative comparison between two distributions.



Additional tests: *bootstrapped vs. observed strandings* (2/5)

Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test A Randomly distribute a ***fixed*** number of total strandings

Start w/ $\mathfrak{D}_{\text{Original}}$ ↓

Strip $\mathfrak{D}_{\text{Original}}$ *of all strandings* ↓

Monte Carlo Algorithm #2a

```
$\tau, \delta_{\text{Max}}, \Delta_{\text{Max}}, w_{\text{Sonar}}$ and $p_{\text{Stranding}} \leftarrow$ Initialize
#Observed Strandings $\leftarrow \mathfrak{D}_{\text{Original}}$
$\mathfrak{D}_{\text{Seed}} \leftarrow \mathfrak{D}_{\text{Original}} - \{1, \ldots, N_{\text{S,Obs}}\}$ // strip of strandings
for sample = $1, 2, \ldots, N_{\text{Samples}}$
    if Necropsy_flag = True, then
        Use necropsy-dependent parameters
    else
        Use default stranding-decay function
    end if
    $\mathfrak{D}_{\text{Sample}} \leftarrow$ Random[#Observed Strandings]
    Apply Monte Carlo Algorithm #1 to $\mathfrak{D}_{\text{Sample}}$
    Coin_counter = 0
    for stranding = $1, 2, \ldots, N_{\text{S,Obs}}$
        $t_{\text{S,A}} \leftarrow p_{\text{Stranding}}$
        $t_{\text{S,L}} \leftarrow \tau, \delta_{\text{Max}}$
        if Random(0,1) $\le w_{\text{Sonar}}(t_{\text{S,A}} - t_{\text{S,L}}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}} \leftarrow$ Coin_counter
    end for // sample
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

Bootstrap **Samples = 1000 | Days = 400 | Strandings = 7**

[Monte Carlo /*full*] $CS_{Full/Min} = 2$, $CS_{Full/Ave} = 3.71$, $CS_{Full/Max} = 5$

[Monte Carlo /*bootstrap*] $CS_{Boot/Min} = 0$, $CS_{Boot/Ave} = 0.51$, $CS_{Boot/Max} = 4$

$f \ge CS_{Full/Min} \to 0.077$, $f \ge CS_{Full/Ave} \to 0.001$, $f \ge CS_{Full/Max} \to 0.$

[Distance] Pearson correlation ≈ 0.557131, $X^2 \approx 0.91023$

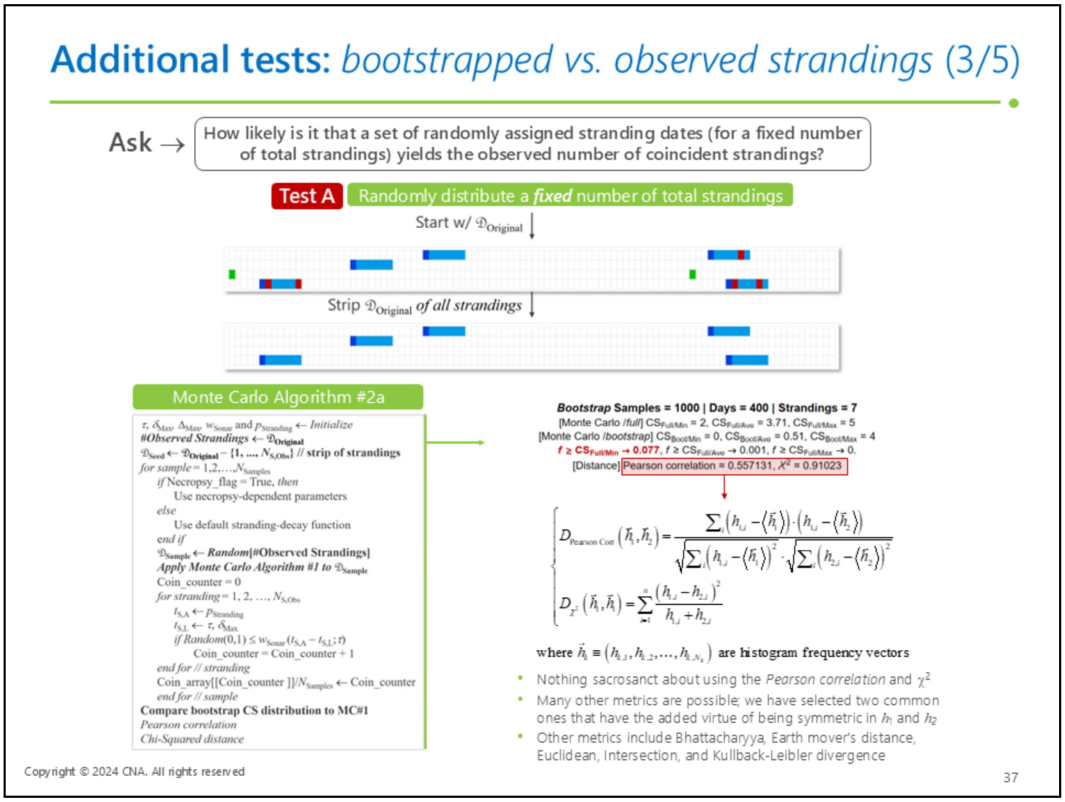
Fraction of samples that yield at least as many coincident strandings as estimated by Monte Carlo Algorithm #1: $f \ge CS_{Full/Min} \approx 0.077$

Number of Coincident Strandings
0
2
4
6
$CS_{Full/Min} = 2$
0.6
0.4
0.2
0
0
0.2
0.4
0.6
Monte Carlo/*Full*, $\mathscr{P}$(CS) ⟵ ... ⟶ Monte Carlo/*Bootstrap*

36

This slide shows a screenshot of the complete output of MCS 2a that includes both histograms and overlays of summary statistics.

The **dotted red line** depicts the minimum number of coincident strandings, as determined by MCS 1 (= $CS_{Full/Min}$). The region highlighted in light red in the right histogram represents the total area of the distribution from MCS 2a for which the number of coincident strandings is greater than or equal to $CS_{Full/Min}$. Heuristically—and in the sense used by Simonis et al. [27]—the smaller the fraction of samples that yield at least as many coincident strandings as MCS 1, the less likely that the number of coincident strandings that were actually observed arose purely because of chance.



Additional tests: *bootstrapped vs. observed strandings* (3/5)

Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test A Randomly distribute a ***fixed*** number of total strandings

Start w/ $\mathfrak{D}_{\text{Original}}$ ↓

Strip $\mathfrak{D}_{\text{Original}}$ *of all strandings* ↓

Monte Carlo Algorithm #2a

```
$\tau, \delta_{\text{Max}}, \Delta_{\text{Max}}, w_{\text{Sonar}}$ and $p_{\text{Stranding}} \leftarrow$ Initialize
#Observed Strandings ← $\mathfrak{D}_{\text{Original}}$
$\mathfrak{D}_{\text{Seed}} \leftarrow \mathfrak{D}_{\text{Original}} - \{1, \ldots, N_{\text{S,Obs}}\}$ // strip of strandings
for sample = $1, 2, \ldots, N_{\text{Samples}}$
    if Necropsy_flag = True, then
        Use necropsy-dependent parameters
    else
        Use default stranding-decay function
    end if
    $\mathfrak{D}_{\text{Sample}} \leftarrow$ Random[#Observed Strandings]
    Apply Monte Carlo Algorithm #1 to $\mathfrak{D}_{\text{Sample}}$
    Coin_counter = 0
    for stranding = $1, 2, \ldots, N_{\text{S,Obs}}$
        $t_{\text{S,A}} \leftarrow p_{\text{Stranding}}$
        $t_{\text{S,L}} \leftarrow \tau, \delta_{\text{Max}}$
        if $Random(0,1) \le w_{\text{Sonar}}(t_{\text{S,A}} - t_{\text{S,L}}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}}$ ← Coin_counter
    end for // sample
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

Bootstrap **Samples = 1000 | Days = 400 | Strandings = 7**

[Monte Carlo /*full*] $CS_{\text{Full/Min}} = 2$, $CS_{\text{Full/Ave}} = 3.71$, $CS_{\text{Full/Max}} = 5$

[Monte Carlo /*bootstrap*] $CS_{\text{Boot/Min}} = 0$, $CS_{\text{Boot/Ave}} = 0.51$, $CS_{\text{Boot/Max}} = 4$

$f \ge CS_{\text{Full/Min}} \rightarrow 0.077$, $f \ge CS_{\text{Full/Ave}} \rightarrow 0.001$, $f \ge CS_{\text{Full/Max}} \rightarrow 0.$

[Distance] Pearson correlation ≈ 0.557131, $X^2 \approx 0.91023$

$$\begin{cases} D_{\text{Pearson Corr}}\left(\vec{h}_1, \vec{h}_2\right) = \dfrac{\sum_i \left(h_{1,i} - \langle \vec{h}_1 \rangle\right) \cdot \left(h_{1,i} - \langle \vec{h}_2 \rangle\right)}{\sqrt{\sum_i \left(h_{1,i} - \langle \vec{h}_1 \rangle\right)^2} \cdot \sqrt{\sum_i \left(h_{2,i} - \langle \vec{h}_2 \rangle\right)^2}} \\ D_{\chi^2}\left(\vec{h}_1, \vec{h}_1\right) = \sum_{i=1}^{n} \dfrac{\left(h_{1,i} - h_{2,i}\right)^2}{h_{1,i} + h_{2,i}} \end{cases}$$

where $\vec{h}_k \equiv \left(h_{k,1}, h_{k,2}, \ldots, h_{k,N_k}\right)$ are histogram frequency vectors

- Nothing sacrosanct about using the *Pearson correlation* and χ^2
- Many other metrics are possible; we have selected two common ones that have the added virtue of being symmetric in h_1 and h_2
- Other metrics include Bhattacharyya, Earth mover's distance, Euclidean, Intersection, and Kullback-Leibler divergence

37

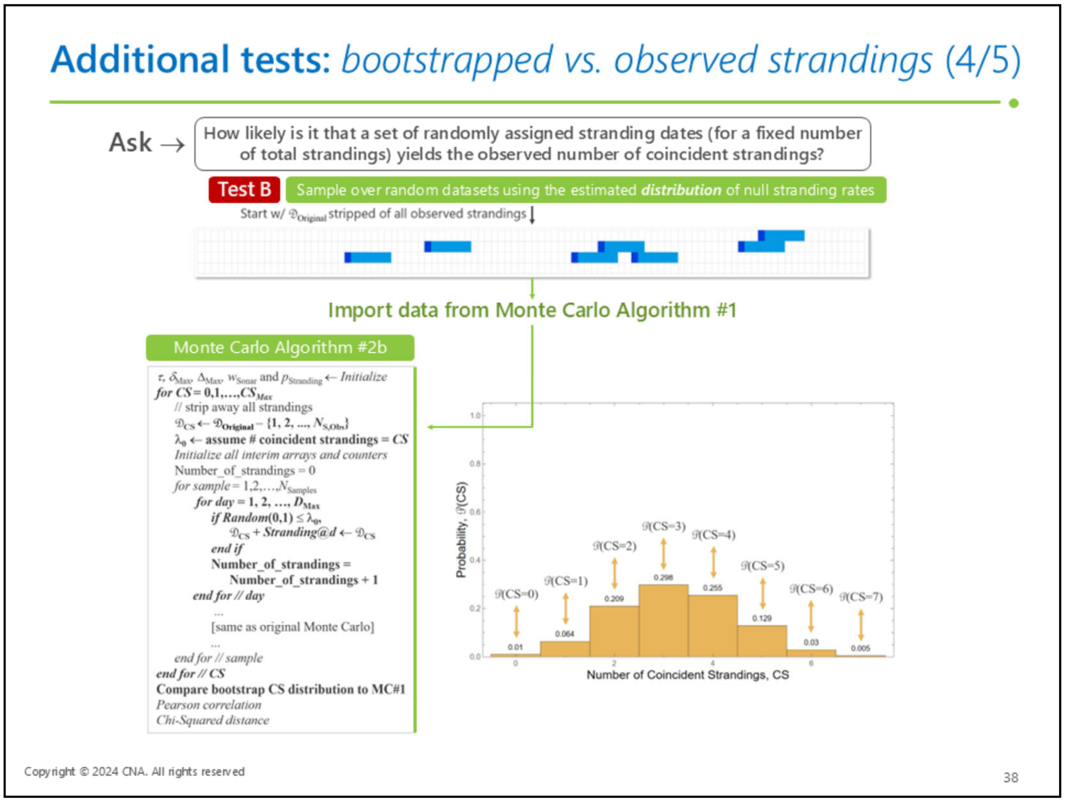
This slide highlights the two metrics that MCS 2a uses to quantify the difference between the two histograms: the Pearson correlation[1] and chi-squared[2] metrics.

As indicated on the slide, there is nothing sacrosanct about using these two particular metrics. Other metrics are possible.[3] These two were chosen for illustrative purposes and because they share the virtue of being symmetric in the two histograms.

[1] *Pearson Correlation*, SPSS Tutorials, Kent State University, https://libguides.library.kent.edu/SPSS/PearsonCorr

[2] Eric W. Weisstein, "Chi-Squared Test," *MathWorld: A Wolfram Web Resource*, https://mathworld.wolfram.com/Chi-SquaredTest.htmlSPSS Tutorials

[2] M. G. Forero, et al., "Analytical Comparison of Histogram Distance Measures," in R. Vera-Rodriguez and A. Morales, editors, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Springer-Verlag, 2019, https://link.springer.com/chapter/10.1007/978-3-030-13469-3_10



Additional tests: *bootstrapped vs. observed strandings* (4/5)

Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test B | Sample over random datasets using the estimated ***distribution*** of null stranding rates

Start w/ $\mathfrak{D}_{\text{Original}}$ stripped of all observed strandings ↓

Import data from Monte Carlo Algorithm #1

Monte Carlo Algorithm #2b

```
$\tau$, $\delta_{\text{Max}}$, $\Delta_{\text{Max}}$, $w_{\text{Sonar}}$ and $p_{\text{Stranding}}$ ← Initialize
for $CS = 0,1,\ldots,CS_{Max}$
    // strip away all strandings
    $\mathfrak{D}_{\text{CS}} \leftarrow \mathfrak{D}_{\text{Original}} - \{1, 2, \ldots, N_{\text{S,Obs}}\}$
    $\lambda_0$ ← assume # coincident strandings = $CS$
    Initialize all interim arrays and counters
    Number_of_strandings = 0
    for sample = $1,2,\ldots,N_{\text{Samples}}$
        for day = $1, 2, \ldots, D_{\text{Max}}$
            if $Random(0,1) \leq \lambda_0$,
                $\mathfrak{D}_{\text{CS}}$ + Stranding@d ← $\mathfrak{D}_{\text{CS}}$
            end if
            Number_of_strandings =
                Number_of_strandings + 1
        end for // day
            ...
            [same as original Monte Carlo]
            ...
    end for // sample
end for // CS
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

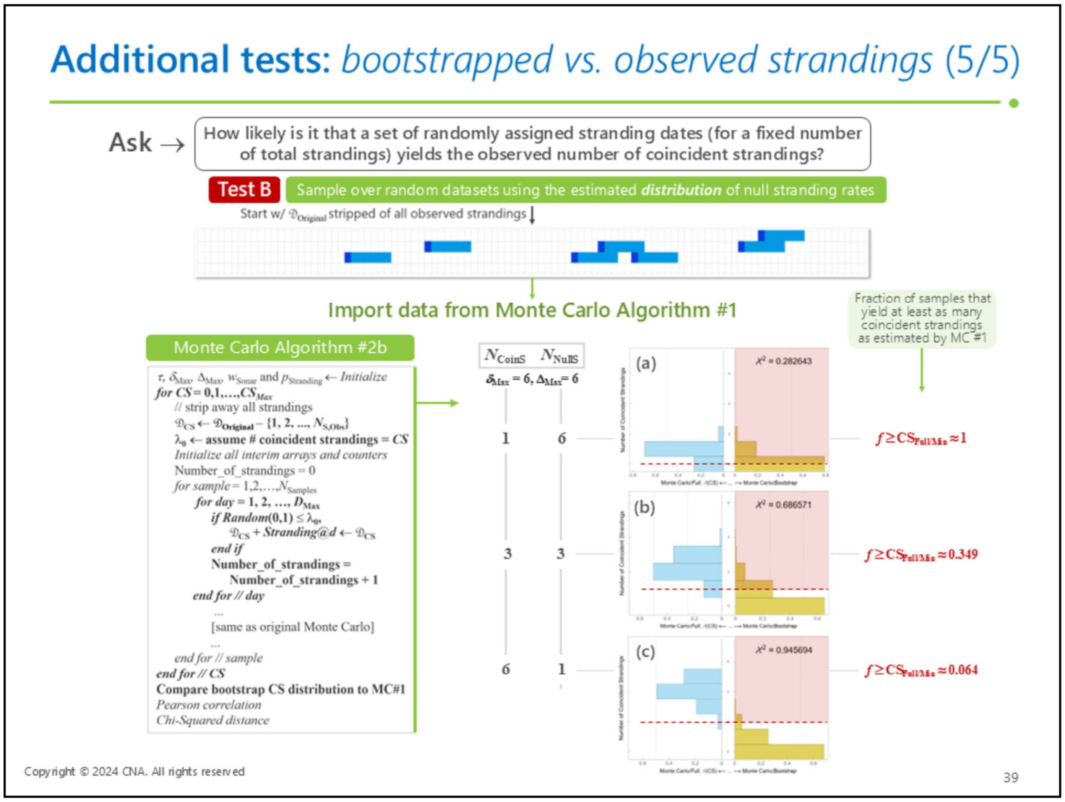
38

Slide 37 indicated that we developed two additional statistical tests (beyond the ones introduced in earlier parts of the slide deck) to determine the probability of observing a given number of coincident strandings (compared to the number we expected to observe from the null stranding rate).

This slide introduces the second of these two tests: Test B = *Monte Carlo Algorithm #2b* (MCS 2b). Unlike Test A, which modifies the main MCS 1 by randomly distributing a fixed number of total strandings (as observed in the original dataset, $\mathfrak{D}_{\text{original}}$), MCS/2b sample over the space of datasets that include a range of possible stranding numbers consistent with the estimated distribution of null stranding rates as determined by MCS 1.

Specifically, after $\mathfrak{D}_{\text{original}}$ is stripped of all strandings (as is done in MCS 2a), MCS/2b proceeds through two loops. The first—outer loop—steps through the set of coincident strandings (i.e., $\text{CS} = 0, 1, \ldots, \text{CS}_{\text{Max}}$) as determined by MCS 1. This meta-data must be imported from MCS 1 prior to running MCS 2b. A representative distribution appears in the bottom right of the slide. The second—outer loop—steps through each day in succession (i.e., $day = 1, \ldots, D_{\text{Max}}$) and associates a stranding with a given day if a pseudo-randomly generated number between 0 and 1 is less than or equal to the null stranding rate, as determined for the value of CS in the outer loop (it is assumed that the total number of strandings remains fixed, so the number of null strandings is always the total number minus the pseudo value of CS).

Notional results are presented on the next slide.



Additional tests: *bootstrapped vs. observed strandings* (5/5)

Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test B Sample over random datasets using the estimated ***distribution*** of null stranding rates

Start w/ $\mathfrak{D}_{Original}$ stripped of all observed strandings ↓

Import data from Monte Carlo Algorithm #1

Monte Carlo Algorithm #2b

```
$\tau$, $\delta_{Max}$, $\Delta_{Max}$, $w_{Sonar}$ and $p_{Stranding}$ ← Initialize
for $CS = 0,1,\ldots,CS_{Max}$
    // strip away all strandings
    $\mathfrak{D}_{CS} \leftarrow \mathfrak{D}_{Original} - \{1, 2, \ldots, N_{S,Obs}\}$
    $\lambda_0$ ← assume # coincident strandings = CS
    Initialize all interim arrays and counters
    Number_of_strandings = 0
    for sample = $1,2,\ldots,N_{Samples}$
        for day = $1, 2, \ldots, D_{Max}$
            if Random(0,1) $\leq \lambda_0$,
                $\mathfrak{D}_{CS}$ + Stranding@d ← $\mathfrak{D}_{CS}$
            end if
            Number_of_strandings =
                Number_of_strandings + 1
        end for // day
            ...
            [same as original Monte Carlo]
            ...
    end for // sample
end for // CS
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

N_{CoinS}	N_{NullS}
$\delta_{Max} = 6, \Delta_{Max} = 6$	
1	6
3	3
6	1

Fraction of samples that yield at least as many coincident strandings as estimated by MC #1

$f \geq CS_{Full/Min} \approx 1$

$f \geq CS_{Full/Min} \approx 0.349$

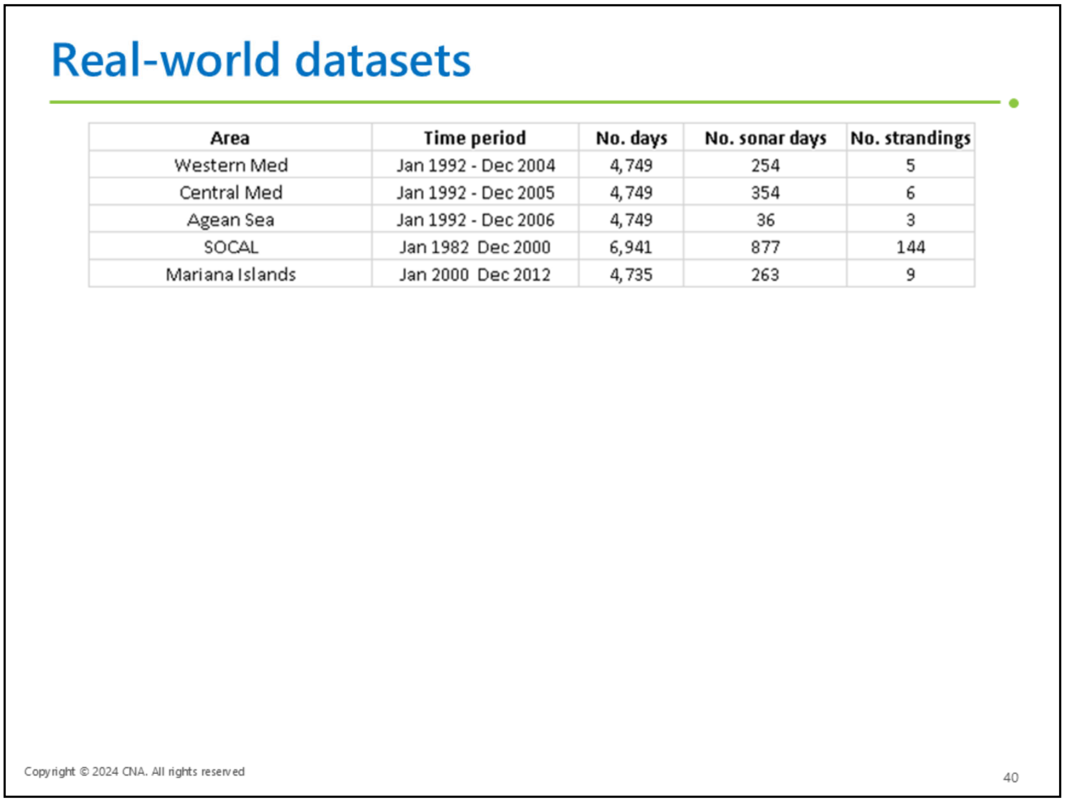
$f \geq CS_{Full/Min} \approx 0.064$

(a)
$X^2 = 0.282643$
(b)
$X^2 = 0.686571$
(c)
$X^2 = 0.945694$
Number of Coincident Strandings
Monte Carlo/Full, f(CS) ← ... → Monte Carlo/Bootstrap

39

This slide shows the histograms for three illustrative runs of MCS 2b. The original (notional) dataset that is stripped of all strandings appears on top. The values of δ_x and Δ_x are both set at six days. The three runs show that for a fixed number of total strandings (= 7), the probability of observing at least the number of coincident strandings as estimated by MCS 1 gets smaller and smaller as the number of observed coincident strandings (i.e., those appearing in $\mathfrak{D}_{\text{original}}$) increases—from one in (a) to three in (b) and six in (c).

To emphasize, neither Test A nor Test B is meant to replace bona fide statistical tests (such as the single-mean Poisson test, FET, or EBT discussed earlier). Rather, as used by Simonis et al. [27], they may be used to lend additional credence to the results of formal tests.



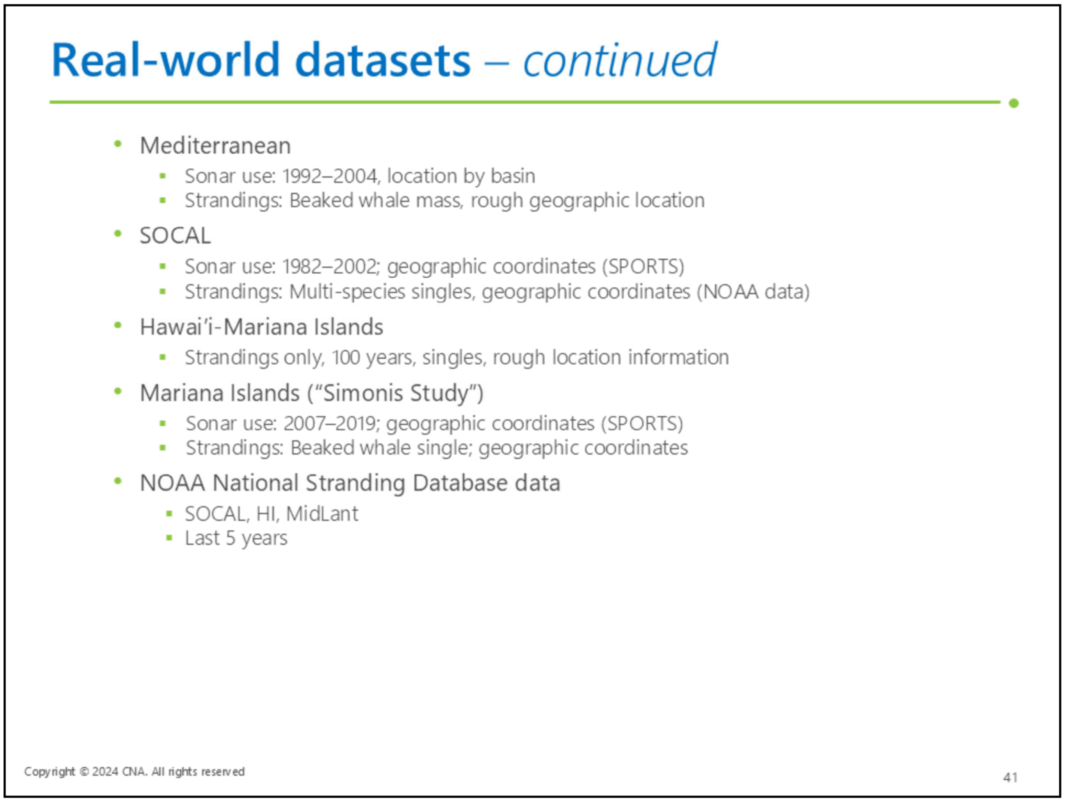
Real-world datasets

Area	Time period	No. days	No. sonar days	No. strandings
Western Med	Jan 1992 - Dec 2004	4,749	254	5
Central Med	Jan 1992 - Dec 2005	4,749	354	6
Agean Sea	Jan 1992 - Dec 2006	4,749	36	3
SOCAL	Jan 1982 Dec 2000	6,941	877	144
Mariana Islands	Jan 2000 Dec 2012	4,735	263	9

The first three datasets listed in the table at the top of this slide, for the Mediterranean Sea, contain only mass strandings (no singles). The last two (Southern California (SOCAL) and the Mariana Islands) are primarily single strandings, with only a few mass strandings in the SOCAL dataset. We note the following regarding our data sources :

- We obtained sonar use information and stranding observations for the three Mediterranean datasets from the open literature as described in [16].
- We compiled sonar use information for the SOCAL dataset from the Navy's Employment Schedule Database (EMPSKED, now called WebSked) [12], various fleet internet and SIPRNet sites, exercise after-action reports, and scheduling data provided by the Southern California Offshore Range (SCORE) operations center. We compiled SOCAL stranding data from hard copy stranding reports maintained by NOAA's West Coast Stranding Network office.
- We obtained sonar use data for the Mariana Islands dataset from [27], from various Navy exercise reports and scheduling documents, and from the Navy's SPORTS database. We obtained stranding data for the Mariana dataset from [27], supplemented with an open-source literature search.

The SOCAL dataset contained information on the decay state of the stranded mammal at time of discovery, noted as one of five categories: 1 for alive, 2 for fresh dead, 3 for moderate decomposition, 4 for advanced decomposition, and 5 for skeletal remains. Such information is useful to help mitigate the uncertainties associated with not knowing the true date of the stranding (recall that stranding databases give the date the stranded mammal was reported to be on the beach, which might not be the day it actually arrived there). The information on decay status will allow us to treat the true stranding date as a random variable with a probability distribution appropriate for each of the decay states given above.



Real-world datasets – *continued*

- Mediterranean
  - Sonar use: 1992–2004, location by basin
  - Strandings: Beaked whale mass, rough geographic location
- SOCAL
  - Sonar use: 1982–2002; geographic coordinates (SPORTS)
  - Strandings: Multi-species singles, geographic coordinates (NOAA data)
- Hawai'i-Mariana Islands
  - Strandings only, 100 years, singles, rough location information
- Mariana Islands ("Simonis Study")
  - Sonar use: 2007–2019; geographic coordinates (SPORTS)
  - Strandings: Beaked whale single; geographic coordinates
- NOAA National Stranding Database data
  - SOCAL, HI, MidLant
  - Last 5 years

41

(Continued …)
Each of these datasets (or some similar version of them) has been used in past studies:

- The three Mediterranean basin datasets are subsets of the data used in the first quantitative study of the correlation between military sonar and marine mammal strandings, described in [17]. That study was initiated following the well-publicized mass strandings of beaked whales in Greece immediately following the nearby testing of low- and mid-frequency acoustic sources by a NATO ASW research lab. Based on sparse data, that study concluded that the stranding rate during sonar use periods was significantly higher (at the 0.95 significance level) than the stranding rate when sonar is not present.

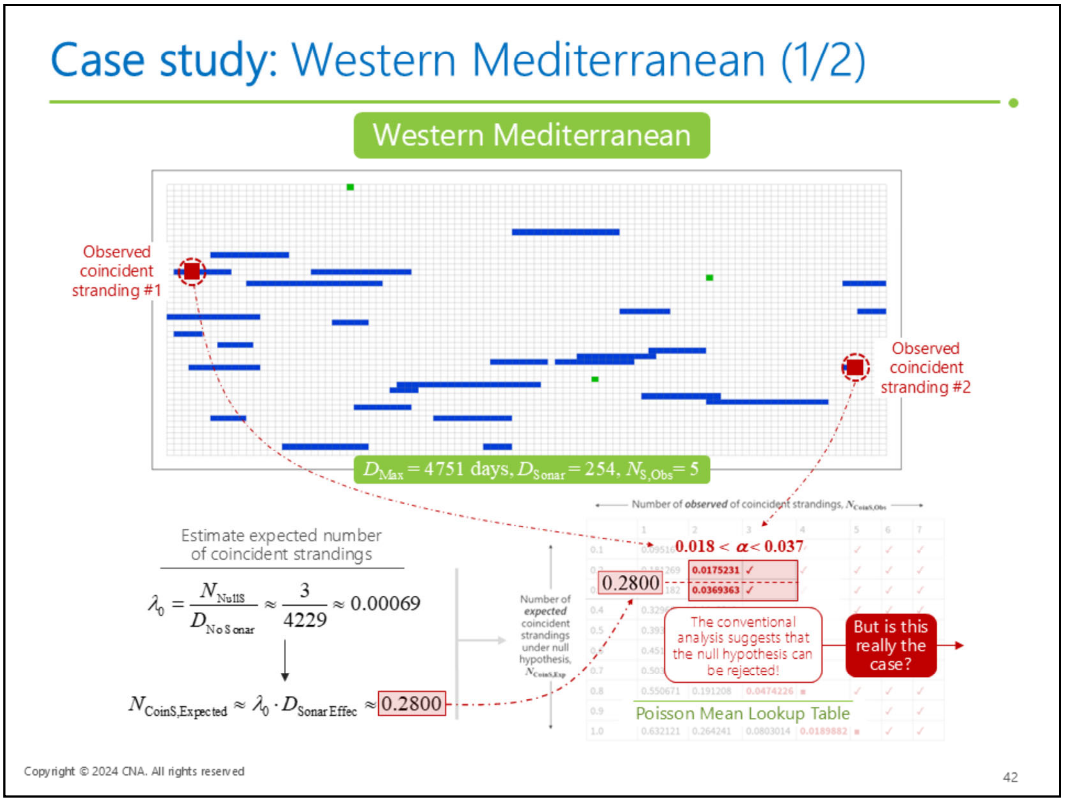
- The SOCAL dataset, which included the decay status of the observed animals, was used in the study described in [16]. That study found no significant difference in stranding rates between sonar and non-sonar periods.

- The Mariana dataset is an update to the data that was used in the study described in [27], which showed a significantly higher (at the 0.95 level) stranding during sonar periods compared to non-sonar periods. We updated the data from [27] with one additional stranding, and we used the Navy's SPORTS database to compile much more complete information on military sonar use.

We also located two additional datasets:

- Stranding data for Hawaii and the Mariana Islands, which contains single strandings over roughly 100 years, and includes approximate location information. This dataset was compiled by NOAA's West Coast Stranding Network office and provided to us by our study sponsor.

- We were also provided data from the NOAA National Stranding Database, consisting of single and mass stranding events for the Pacific.



Case study: Western Mediterranean (1/2)

Western Mediterranean

Observed coincident stranding #1
Observed coincident stranding #2
Number of observed of coincident strandings, $N_{\text{CoinS,Obs}}$
Number of expected coincident strandings under null hypothesis, $N_{\text{CoinS,Exp}}$
1
2
3
4
5
6
7
0.1
0.4
0.5
0.7
0.8
0.9
1.0
0.550671
0.191208
0.0474226
0.632121
0.264241
0.0803014
0.0189882
0.0175231 ✓
0.0369363 ✓
0.2800
$0.018 < \alpha < 0.037$
Poisson Mean Lookup Table

$D_{\text{Max}} = 4751$ days, $D_{\text{Sonar}} = 254$, $N_{\text{S,Obs}} = 5$

Estimate expected number of coincident strandings

$$\lambda_0 = \frac{N_{\text{NullS}}}{D_{\text{No Sonar}}} \approx \frac{3}{4229} \approx 0.00069$$

$$N_{\text{CoinS,Expected}} \approx \lambda_0 \cdot D_{\text{Sonar Effec}} \approx \boxed{0.2800}$$

The conventional analysis suggests that the null hypothesis can be rejected!

But is this really the case?

Copyright © 2024 CNA. All rights reserved

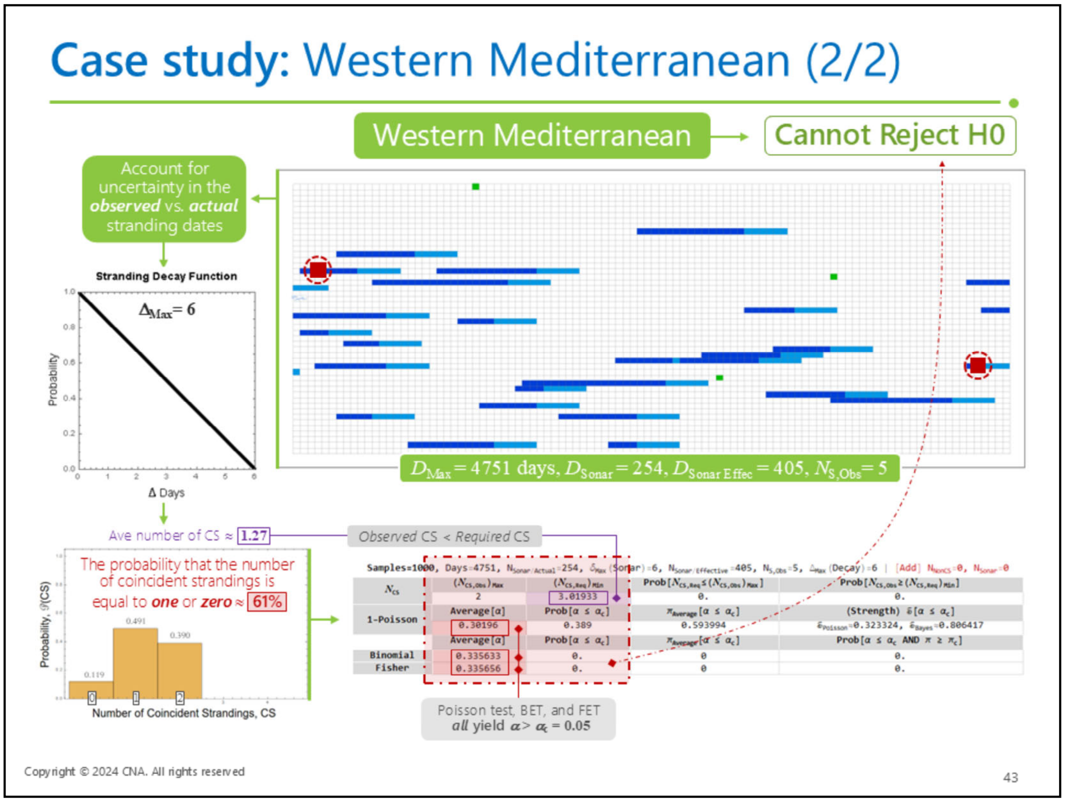
42

Slides 42-50 show examples of how the methodology introduced thus far can be applied to real-world datasets. (The practical steps required to run the Mathematica software developed for this study are summarized in **Appendix I**.)

The first case study uses data from the western Mediterranean to illustrate how existing methods are unable, on their own, to mitigate the inherent uncertainties in the data. The time-series displayed at the top of the slide shows a total of five strandings, two of which are ostensibly identified as coincident with sonar. Results of the conventional analysis are summarized at the bottom left of the slide, where the null stranding rate is calculated to be 0.00069 strandings per day, and the *expected* number of coincident strandings is estimated to be 0.28. Consulting the *Poisson Mean Lookup Table* (see slide 20), we find that the P-value lies somewhere between 0.018 and 0.037, leading us to conclude that the null hypothesis can be rejected (because even the upper value is less than α_c = 0.05).

But is this really the case?

The next slide shows that if basic underlying uncertainties in the data are taken into account (by applying the methods described earlier), rejection is no longer warranted.



Case study: Western Mediterranean (2/2)

Western Mediterranean → Cannot Reject H0

Account for uncertainty in the ***observed*** vs. ***actual*** stranding dates

$D_{Max} = 4751$ days, $D_{Sonar} = 254$, $D_{Sonar\,Effec} = 405$, $N_{S,Obs} = 5$

Ave number of CS ≈ 1.27

The probability that the number of coincident strandings is equal to ***one*** or ***zero*** ≈ 61%

Observed CS < Required CS

Samples=1000, Days=4751, $N_{Sonar/Actual}$=254, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=405, $N_{S,Obs}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob$[N_{CS,Req} \le (N_{CS,Obs})_{Max}]$	Prob$[N_{CS,Obs} \ge (N_{CS,Req})_{Min}]$
	2	3.01933	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{Average}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.30196	0.389	0.593994	$\delta_{Poisson}$=0.323324, δ_{Bayes}=0.806417
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{Average}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.335633	0.	0	0.
Fisher	0.335656	0.	0	0.

Poisson test, BET, and FET *all* yield $\alpha > \alpha_c = 0.05$

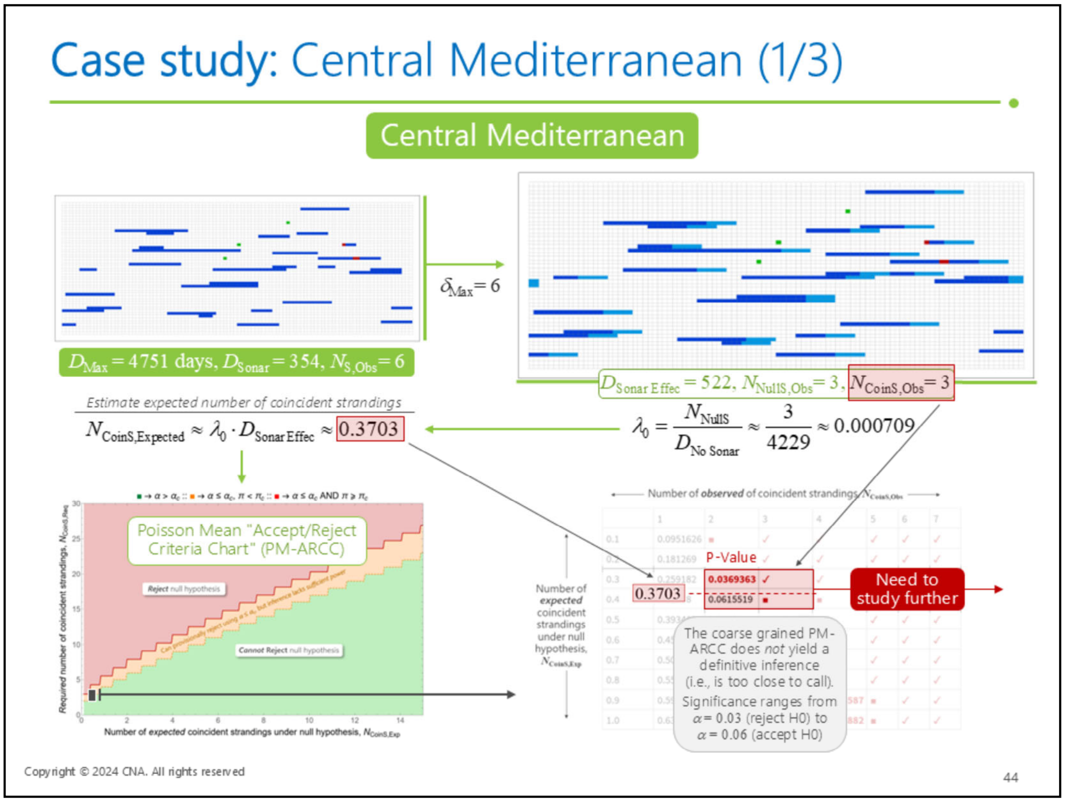
43

This slide extends the conventional analysis of the western Mediterranean dataset—which, as the previous slide shows, prima facie rejects the null hypothesis by taking into account the uncertainty in the observed versus actual stranding dates. The stranding decay function is defined as a simple linear ramp, with a maximum value of one at zero days (i.e., the highest probability is that strandings occur on the same day as they are reported) and a minimum value of zero at six days (i.e., $\delta_x = 6$). As mentioned earlier, the exact form of this function is unknown.[1] What matters is that it provides a plausible and realistic way to account for an inherent uncertainty regarding when strandings actually occurred. Even more saliently, the specific results summarized here do not appreciably change when using other functional forms (including using a uniform distribution rather than a linearly decreasing one), provided that $\delta_x > 1$; in other words, there is a nonzero probability that reported stranding dates are unequal to actual (albeit unknown) earlier dates.

The histogram on the lower left of the slide shows the statistical distribution of the number of coincident strandings as determined by running MCS 1 for 1,000 samples. The probability that the number of coincident strandings is equal to one or zero is about 61 percent; in other words, it is *not* unequivocally equal to *two*, as the conventional approach assumes. This finding significantly changes the inferential calculus because a smaller number of coincident strandings decreases the strength of the statistical evidence required to reject the null hypothesis. The area highlighted in **red** in the matrix summarizing the output statistics run bears quantitative witness to this: the single-mean Poisson test, FET, and BET all yield P-values much larger than $\alpha_c = 0.05$.

In other words, accounting for *just* the uncertainty in the observed versus actual stranding dates (remember, myriad other uncertainties are left unaccounted for by conventional methods; see slide 25), we conclude that rejecting the null hypothesis is *not warranted!*

[1] This dataset does not include necropsy data, otherwise the functional form may be tailored to an animal's decay-state; see **Appendix F**.



Case study: Central Mediterranean (1/3)

Central Mediterranean

$D_{Max} = 4751$ days, $D_{Sonar} = 354$, $N_{S,Obs} = 6$

$\delta_{Max} = 6$

$D_{Sonar\,Effec} = 522$, $N_{NullS,Obs} = 3$, $N_{CoinS,Obs} = 3$

Estimate expected number of coincident strandings

$$N_{CoinS,Expected} \approx \lambda_0 \cdot D_{Sonar\,Effec} \approx \boxed{0.3703}$$

$$\lambda_0 = \frac{N_{NullS}}{D_{No\,Sonar}} \approx \frac{3}{4229} \approx 0.000709$$

Poisson Mean "Accept/Reject Criteria Chart" (PM-ARCC)

P-Value

Need to study further

The coarse grained PM-ARCC does *not* yield a definitive inference (i.e., is too close to call). Significance ranges from $\alpha = 0.03$ (reject H0) to $\alpha = 0.06$ (accept H0)

🟩 → $\alpha > \alpha_c$ :: 🟧 → $\alpha \le \alpha_c$, $\pi < \pi_c$ :: 🟥 → $\alpha \le \alpha_c$ AND $\pi \ge \pi_c$
Reject null hypothesis
Can potentially reject using α ≤ αc but inference lacks sufficient power
Cannot Reject null hypothesis
Required number of coincident strandings, $N_{CoinS,Req}$
Number of expected coincident strandings under null hypothesis, $N_{CoinS,Exp}$
Number of observed of coincident strandings, $N_{CoinS,Obs}$
Number of expected coincident strandings under null hypothesis, $N_{CoinS,Exp}$
0.3703
0.0369363 ✓
0.0615519 ■

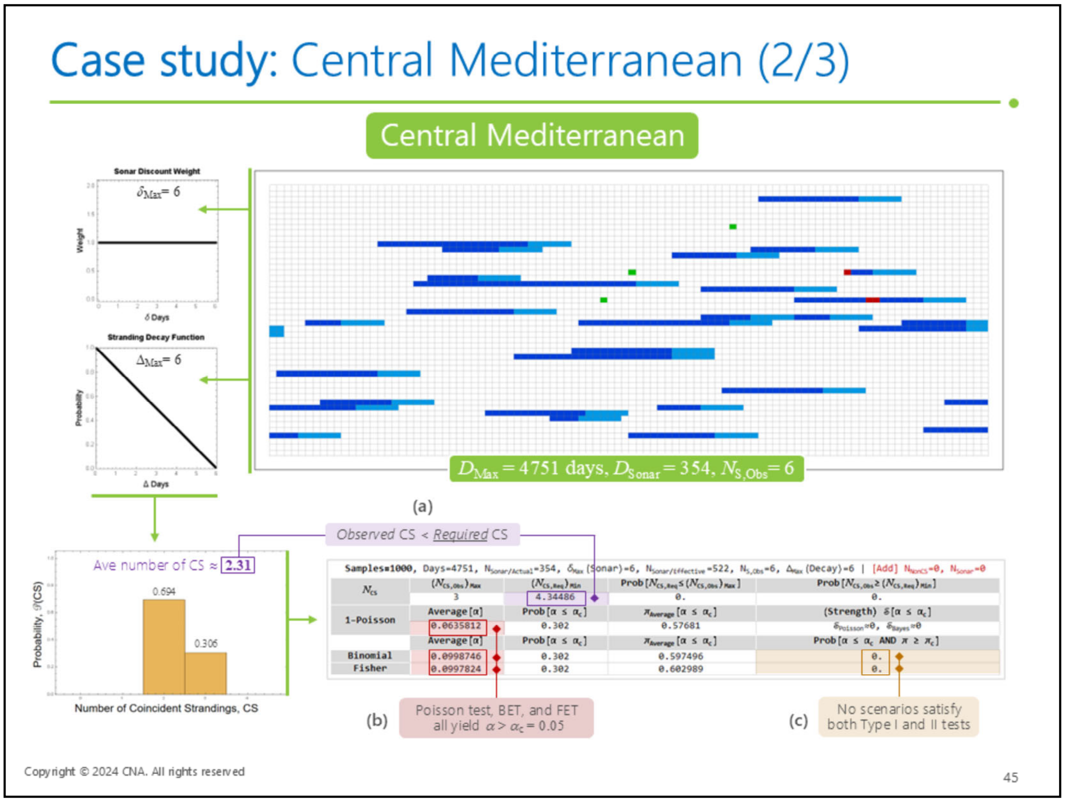
Copyright © 2024 CNA. All rights reserved

44

The second case study uses data from the central Mediterranean to show how exploring uncertainties in the data can strengthen (or lend more credence to) an otherwise ambiguous inference (drawn using existing methods). Here, the time-series is displayed at the top left of the slide. When augmented by a δ_x = 6 day window, the time-series nominally shows a total of six strandings, *three* of which are identified as coincident with sonar.

The conventional analysis is summarized as follows: the null stranding rate is estimated to be about 0.00071 strandings/day, the number of expected coincident strandings is ≈ 0.37, and (after consulting the Poisson mean *Accept/Reject Criteria Chart* on slide 24) the P-value lies somewhere between α = 0.03 (reject null hypothesis) and α = 0.06 (accept null hypothesis). So, at first cut, this result is ambiguous because the statistical evidence appears insufficient to either reject or accept the null hypothesis.

The next two slides show how this ambiguity may be partly ameliorated by additional analysis that explicitly accounts for uncertainties.



Case study: Central Mediterranean (2/3)

Central Mediterranean

$D_{\text{Max}} = 4751$ days, $D_{\text{Sonar}} = 354$, $N_{\text{S,Obs}} = 6$

(a)

Observed CS < Required CS

Samples=1000, Days=4751, $N_{\text{Sonar/Actual}}$=354, δ_{Max} (Sonar)=6, $N_{\text{Sonar/Effective}}$=522, $N_{\text{S,Obs}}$=6, Δ_{Max} (Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob$[N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}]$	Prob$[N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}]$
	3	4.34486	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.0635812	0.302	0.57681	$\delta_{\text{Poisson}} \approx 0$, $\delta_{\text{Bayes}} \approx 0$
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.0998746	0.302	0.597496	0.
Fisher	0.0997824	0.302	0.602989	0.

(b) Poisson test, BET, and FET all yield $\alpha > \alpha_c = 0.05$

(c) No scenarios satisfy both Type I and II tests

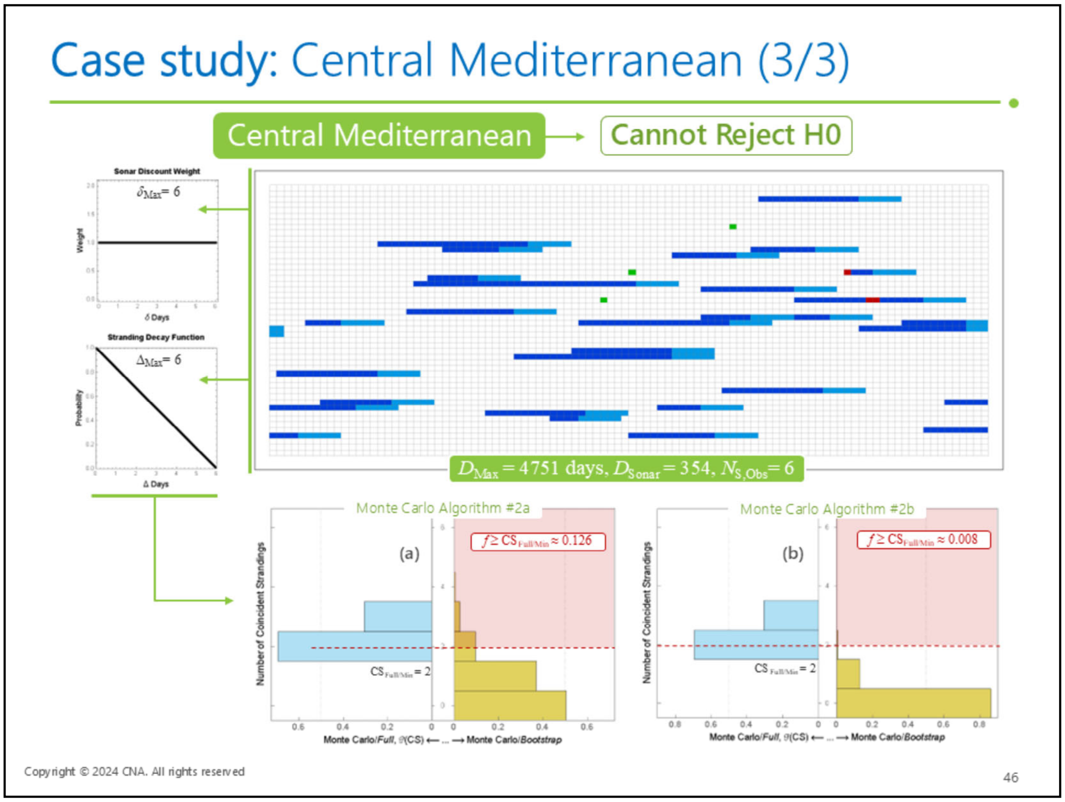
45

This slide extends the conventional analysis of the central Mediterranean dataset by taking into account (1) the uncertainty associated with assigning an actual stranding date to the observed stranding date and (2) the uncertainty associated with identifying a given stranding as coincident with sonar.

The *sonar discount* and *stranding decay* functions are defined as indicated at the top left of the slide, using $\delta_x = \Delta_x = 6$ days.

The result of running MCS 1 (using 1,000 samples) shows that the average number of coincident strandings is about 2.3, which is (as expected) less than the three coincident strandings that appear in the original dataset, but which does not account for the likelihood that some or all strandings actually occurred a few days prior to when they were reported.

This smaller-than-nominally-observed number of coincident strandings suggests that the supporting evidence to reject the null hypothesis is even less than what led to the already ambiguous result on the last slide. This is born out in two ways: (a) the average number of coincident strandings (accounting for uncertainty) is significantly less than the required minimum number of coincident strandings required to satisfy both the Type I and Type II tests (as highlighted in **purple**), and (b) all three statistical tests (single-mean Poisson test, BET, and FET) yield $\alpha > \alpha_c$.



Case study: Central Mediterranean (3/3)

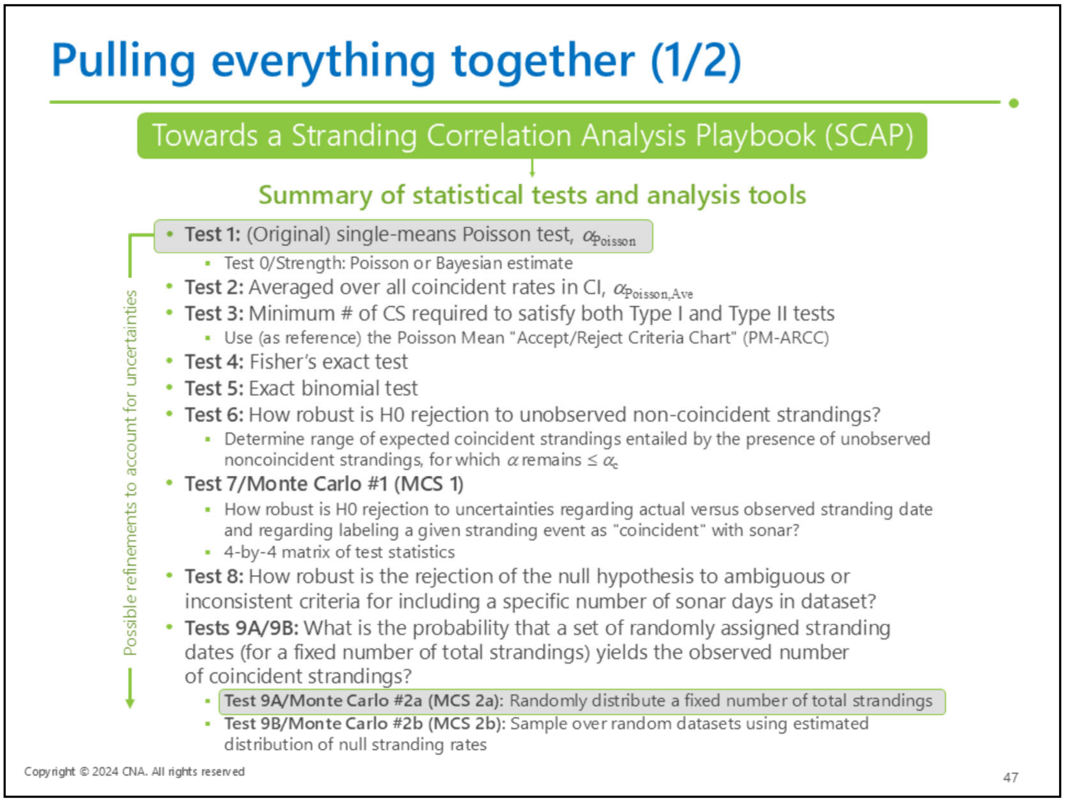
Central Mediterranean → Cannot Reject H0

46

This slide illustrates how some of the methods introduced earlier may be used to gather additional evidence to strengthen the inference to *not* reject the null hypothesis.

The two histograms at the bottom of the slide show the output of MCS 2a and MCS 2b, respectively (see slides 32 to 36). In each case, the probability of observing at least the number of coincident strandings as expected from the null stranding rate (i.e., two, as determined using the original dataset, but without accounting for any uncertainties) is only 12.6 percent for MCS 2a and 0.8 percent for MCS 2b.

Although this additional analysis does not substantively change the original inference (which was ambiguous at best, but which also provided insufficient evidence to reject the null hypothesis), the takeaway is that it strengthens the veracity of concluding that the statistics do not warrant rejection.



Pulling everything together (1/2)

Towards a Stranding Correlation Analysis Playbook (SCAP)

Summary of statistical tests and analysis tools

Possible refinements to account for uncertainties

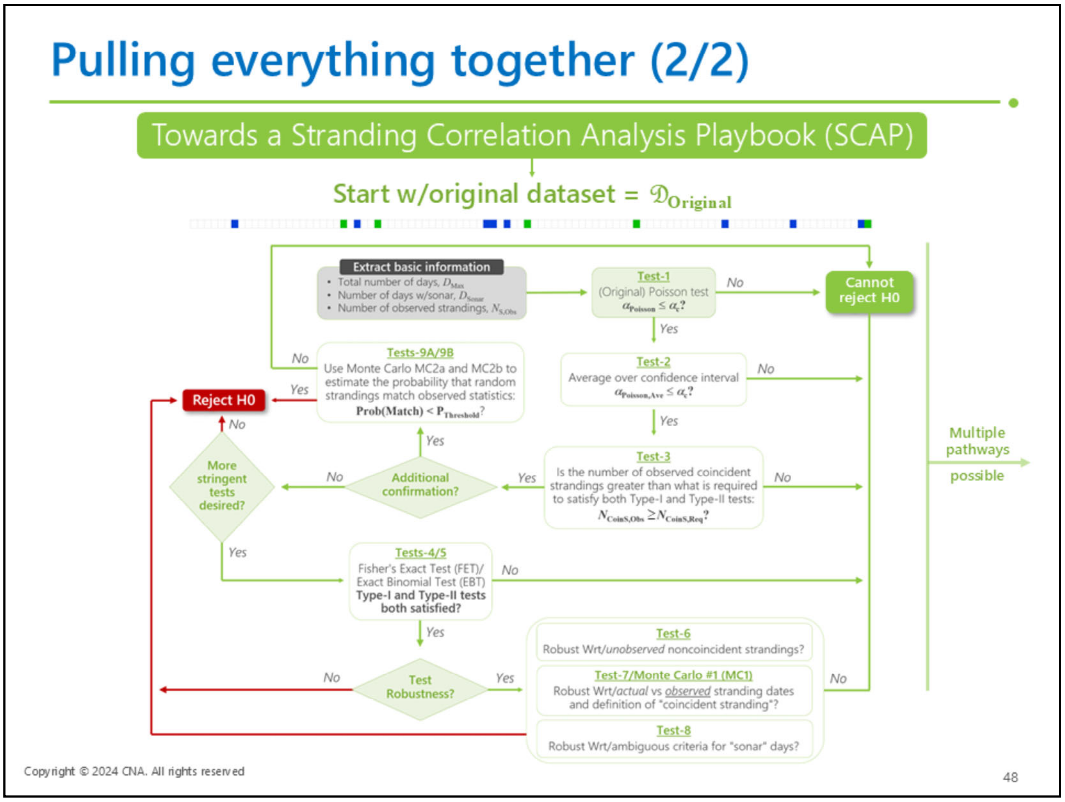
- **Test 1:** (Original) single-means Poisson test, α_{Poisson}
  - Test 0/Strength: Poisson or Bayesian estimate
- **Test 2:** Averaged over all coincident rates in CI, $\alpha_{\text{Poisson,Ave}}$
- **Test 3:** Minimum # of CS required to satisfy both Type I and Type II tests
  - Use (as reference) the Poisson Mean "Accept/Reject Criteria Chart" (PM-ARCC)
- **Test 4:** Fisher's exact test
- **Test 5:** Exact binomial test
- **Test 6:** How robust is H0 rejection to unobserved non-coincident strandings?
  - Determine range of expected coincident strandings entailed by the presence of unobserved noncoincident strandings, for which α remains ≤ α_c
- **Test 7/Monte Carlo #1 (MCS 1)**
  - How robust is H0 rejection to uncertainties regarding actual versus observed stranding date and regarding labeling a given stranding event as "coincident" with sonar?
  - 4-by-4 matrix of test statistics
- **Test 8:** How robust is the rejection of the null hypothesis to ambiguous or inconsistent criteria for including a specific number of sonar days in dataset?
- **Tests 9A/9B:** What is the probability that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?
  - **Test 9A/Monte Carlo #2a (MCS 2a):** Randomly distribute a fixed number of total strandings
  - **Test 9B/Monte Carlo #2b (MCS 2b):** Sample over random datasets using estimated distribution of null stranding rates

47

This slide summarizes the battery of statistical tests and analysis tools introduced thus far. Readers are encouraged to view this list as an assembly of parts making up the SCAP that immediately follows.

Apart from the sheer number of tests that appear on this list, the most salient point is that—as of this writing (February 2025)—**the vast majority of extant stranding studies use only some combination of the two tests that are highlighted in gray**. The other tests collectively refine this basic approach by providing tools that explicitly account for uncertainties.

Of course, it is not immediately clear which tests are the most suitable for a given scenario or whether a given test is necessary (or even applicable). The SCAP, introduced on the next slide, organizes these tests into a flowchart that analysts and other stakeholders can use to navigate the inferential process—that culminates in either a decision to reject the null hypothesis or the finding that the null hypothesis cannot be rejected.



Pulling everything together (2/2)

Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathfrak{D}_{\mathrm{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\mathrm{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\mathrm{Poisson}} \le \alpha_c$?

Cannot reject H0

Test-2
Average over confidence interval
$\alpha_{\mathrm{Poisson,Ave}} \le \alpha_c$?

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\mathrm{CoinS,Obs}} \ge N_{\mathrm{CoinS,Req}}$?

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\mathrm{Prob(Match)} < \mathrm{P}_{\mathrm{Threshold}}$?

Reject H0

Additional confirmation?

More stringent tests desired?

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Test Robustness?

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

Multiple pathways possible

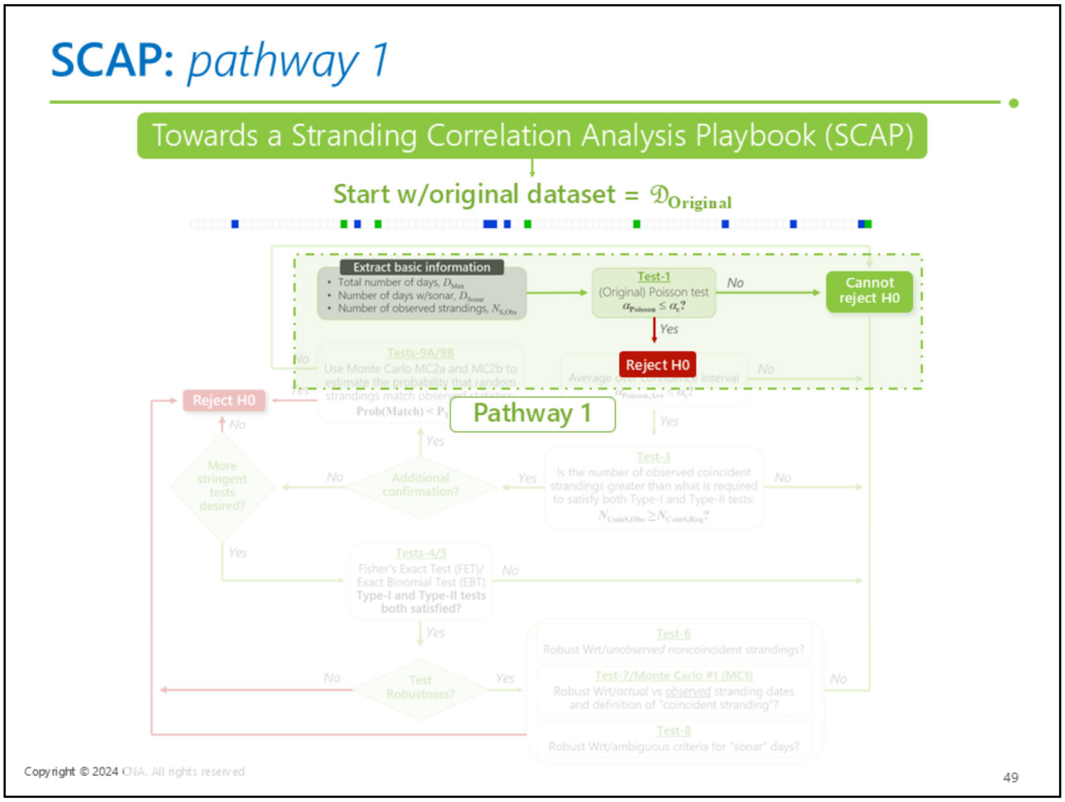
This slide presents a draft version of the SCAP. The design goal is twofold: (1) to weave together what otherwise would be a disorganized list of stand-alone statistical tests and simulation tools and (2) to provide multiple inferential pathways that stakeholders can choose to take, depending on individual preferences and requirements.

Because each of the next five slides isolates and describes the specifics of potential pathways, we limit our discussion of the complete SCAP, as it appears here, to its essential elements.

The presumption is that an analyst (or other stakeholder) will start with a dataset that contains sonar and strandings data. Note that the SCAP does not address any uncertainties that may be introduced during the *preparation* of this dataset (see **Appendix E**).

All pathways start by compiling a list of basic statistics that must be extracted from this dataset (as illustrated in the gray box at the top left of the slide). Of course, other statistics may need to be extracted later, depending on which pathways are followed.

The following slides outline five increasingly refined inferential pathways that may be followed while navigating the SCAP.



SCAP: *pathway 1*

Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathfrak{D}_{\mathbf{Original}}$

Extract basic information
- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\text{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\text{Poisson}} \le \alpha_c$?

Cannot reject H0

Reject H0

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
Prob(Match) < P_T

Average over confidence interval
$\alpha_{\text{Poisson,Ave}} \le \alpha_c$?

Pathway 1

Reject H0

More stringent tests desired?

Additional confirmation?

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\text{CoinS,Obs}} \ge N_{\text{CoinS,Req}}$?

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Test Robustness?

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

49

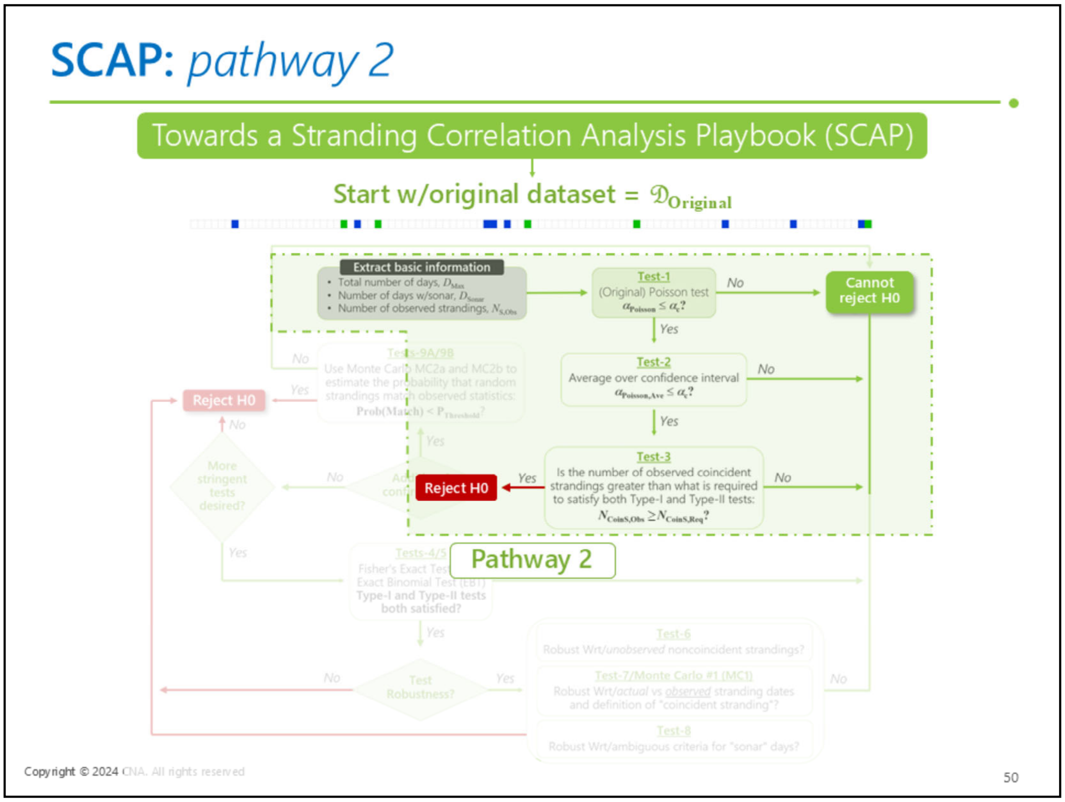
This slide highlights the first of five possible SCAP pathways that stakeholders may follow.

Test 1, which appears immediately to the left of the gray box, constitutes SCAP's de facto first step and refers to the single-mean Poisson test that lies at the heart of almost all stranding studies. Recall that the single-mean Poisson test adjudicates only significance (i.e., it tests only whether the P-value is less than some critical value, α_c) and not power. If the test *fails* (meaning if the answer to the question, "Is the P-value less than or equal to α_c?" is no), we infer that the null hypothesis cannot be credibly ignored given the existing statistical evidence. In particular, no additional tests or criteria are required, and the remaining parts of the SCAP can effectively be ignored.

The SCAP starts offering benefits if Test 1 is *passed,* or if the answer to the above question is yes. If Stakeholder A wishes to merely adhere to roughly the same level of rigor that characterizes most existing methods (which this first inferential pathway nominally entails), the null hypothesis may be rejected at this point, and the analysis will end.

However, Stakeholder B may wish to more rigorously interrogate and analyze the data, choosing not to automatically reject the null hypothesis solely on the basis of passing a statistical significance test whose power is unaccounted for. How confident are we that the rejection is robust with regard to uncertainties in the data? Does a lack of achieving a threshold confidence justify reversing our original inference (based on significance alone)? Would our decision change if additional tests are applied? Might other criteria be used to support making (or changing) an interim inference?

The following slides outline four increasingly interconnected pathways designed to provide answers to these questions.



SCAP: *pathway 2*

Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathfrak{D}_{\text{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\text{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\text{Poisson}} \le \alpha_c$?

Test-2
Average over confidence interval
$\alpha_{\text{Poisson,Ave}} \le \alpha_c$?

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\text{CoinS,Obs}} \ge N_{\text{CoinS,Req}}$?

Cannot reject H0

Reject H0

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
Prob(Match) < $P_{\text{Threshold}}$?

More stringent tests desired?

Tests-4/5
Fisher's Exact Test
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Pathway 2

Test Robustness?

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs observed stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

50

The second pathway adds two additional tests that may be applied to the data to strengthen the veracity of whatever final inference is drawn.

If the answer to the single-mean Poisson test (i.e., "is the P-value $\alpha_{\text{Poisson}} \le \alpha_c$?") is yes, then Test 2 asks the same question but uses a P-value that is estimated by averaging over all null coincident rates that fall within a given confidence interval of the observed null stranding rate (see **slides 16 to 17**).

If the answer to this more robust version of the original question is no (i.e., it is determined that $\alpha_{\text{Poisson,Ave}} > \alpha_c$), then it may be inferred that statistical evidence is lacking to reject the null hypothesis, and the analysis will end.

On the other hand, is the answer is yes (i.e., that $\alpha_{\text{Poisson,Ave}} \le \alpha_c$), then an additional test, Test 3, may be applied. In this case, the question is whether the *observed* number of coincident strandings, $N_{\text{CoinS,Obs}}$, is at least as large as the minimal number that is required to satisfy both significance (or Type I errors) and power (or Type II errors), $N_{\text{CoinS,Req}}$, as described on **slides 18 to 21**.

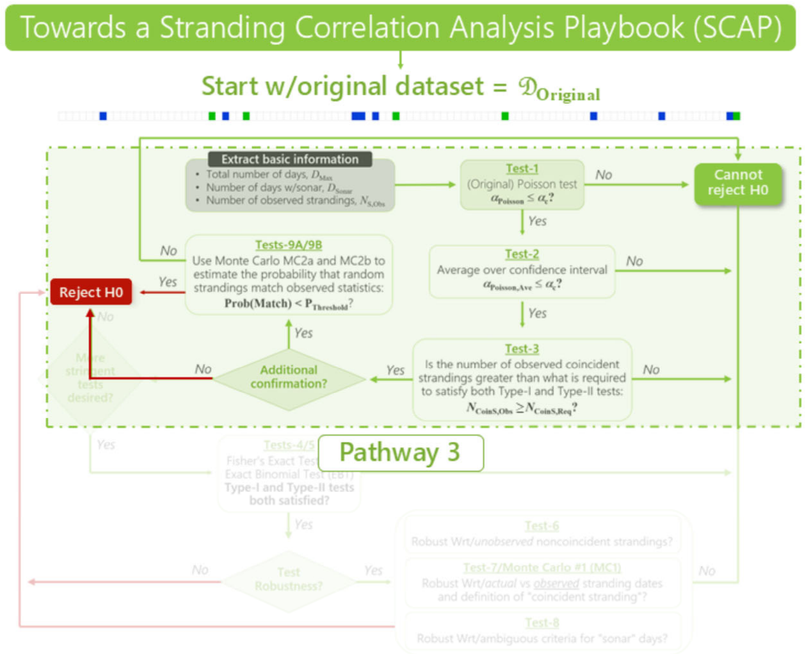
If $N_{\text{CoinS,Obs}} < N_{\text{CoinS,Req}}$, then *the null hypothesis cannot be rejected*, despite the fact that $\alpha_{\text{Poisson}} \le \alpha_c$ (as per the original affirmative answer to the question posed by Test 1).

The null hypothesis may be rejected by following the inferential flows in Pathway #2 if and only if it is determined that $N_{\text{CoinS,Obs}} < N_{\text{CoinS,Req}}$.

Pathway #2 is arguably the simplest (and certainly the most straightforward and intuitively sensible) way to refine and strengthen the veracity of existing tests.

If a stakeholder wishes for additional confirmation or to administer yet more stringent tests that better account for uncertainties in the data, an immediate option is to follow Pathway #3, described on the next slide.

SCAP: *pathway 3*



Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathfrak{D}_{\text{Original}}$

Extract basic information
- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\text{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\text{Poisson}} \le \alpha_c$?

Cannot reject H0

Test-2
Average over confidence interval
$\alpha_{\text{Poisson,Ave}} \le \alpha_c$?

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\text{CoinS,Obs}} \ge N_{\text{CoinS,Req}}$?

Additional confirmation?

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\text{Prob(Match)} < \text{P}_{\text{Threshold}}$?

Reject H0

More stringent tests desired?

Tests-4/5
Fisher's Exact Test
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Test Robustness?

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

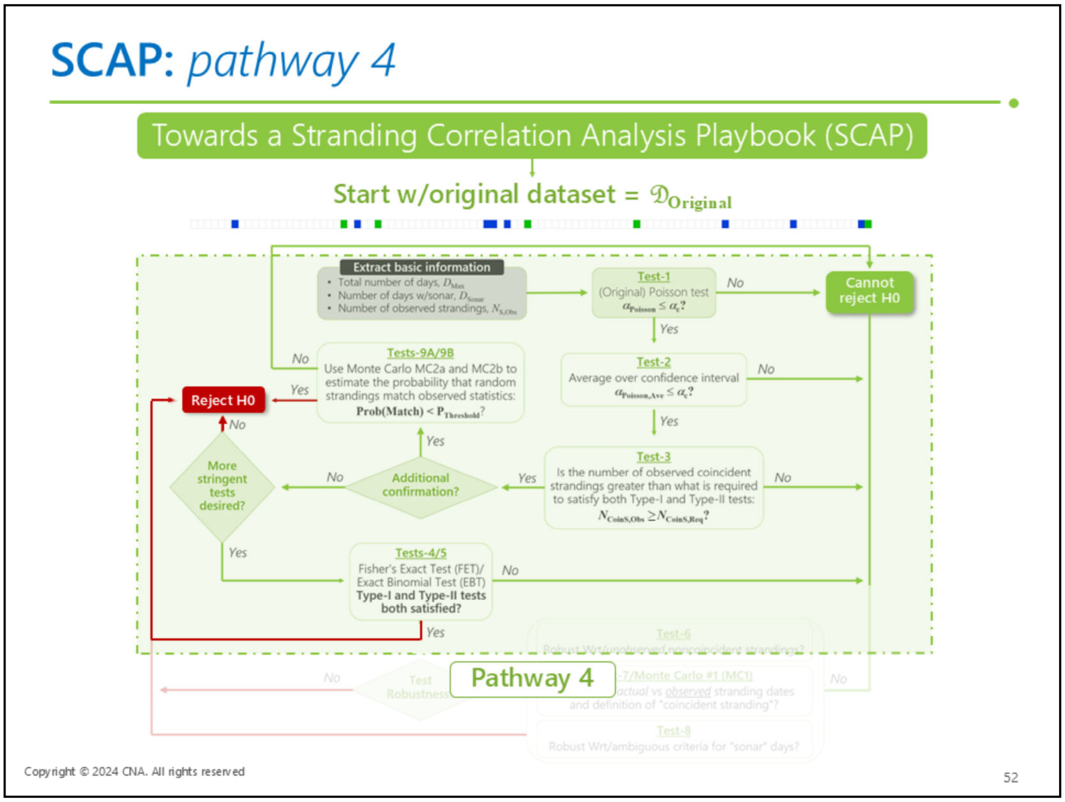
Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

Pathway 3

51

The option to follow (or keep following) the third pathway opens up if the answer to Test 3 (as described on the preceding slide) is yes and if the stakeholder wishes to obtain additional confirmation that the statistical evidence warrants rejecting the null hypothesis. We remind the reader that a given stakeholder must decide which inferential pathway to follow. On this slide, we have effectively entered Pathway #3 with the presumption that the outcomes of all prior tests—as encountered on Pathway #1 (Test 1) and Pathway #2 (Tests 2 and 3)—have been yes.

If no additional confirmation is needed, Pathway #3 terminates by rejecting the null hypothesis. If additional confirmation is desired, the option is to run a MCS (either MCS 2a, MCS 2b, or both; see **slides 35 to 39**) to determine the likelihood that a set of random bootstrapped strandings yields the same number of coincident strandings as are actually observed. If the probability that these two are equal is less than some minimum threshold, $P_{\text{Threshold}}$ (the value of which is left to the stakeholder's discretion), then the statistical evidence has already accrued—and by itself was already deemed sufficient to reject in Pathway #2—and is only strengthened. Thus, a positive outcome of Tests 9A and 9B warrants rejecting the null hypothesis and terminating the analysis.



SCAP: *pathway 4*

Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathfrak{D}_{\mathbf{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\mathrm{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\mathrm{Poisson}} \le \alpha_c$?

Cannot reject H0

Test-2
Average over confidence interval
$\alpha_{\mathrm{Poisson,Ave}} \le \alpha_c$?

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\mathrm{CoinS,Obs}} \ge N_{\mathrm{CoinS,Req}}$?

Additional confirmation?

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\mathrm{Prob(Match)} < \mathrm{P}_{\mathrm{Threshold}}$?

Reject H0

More stringent tests desired?

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Test Robustness

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

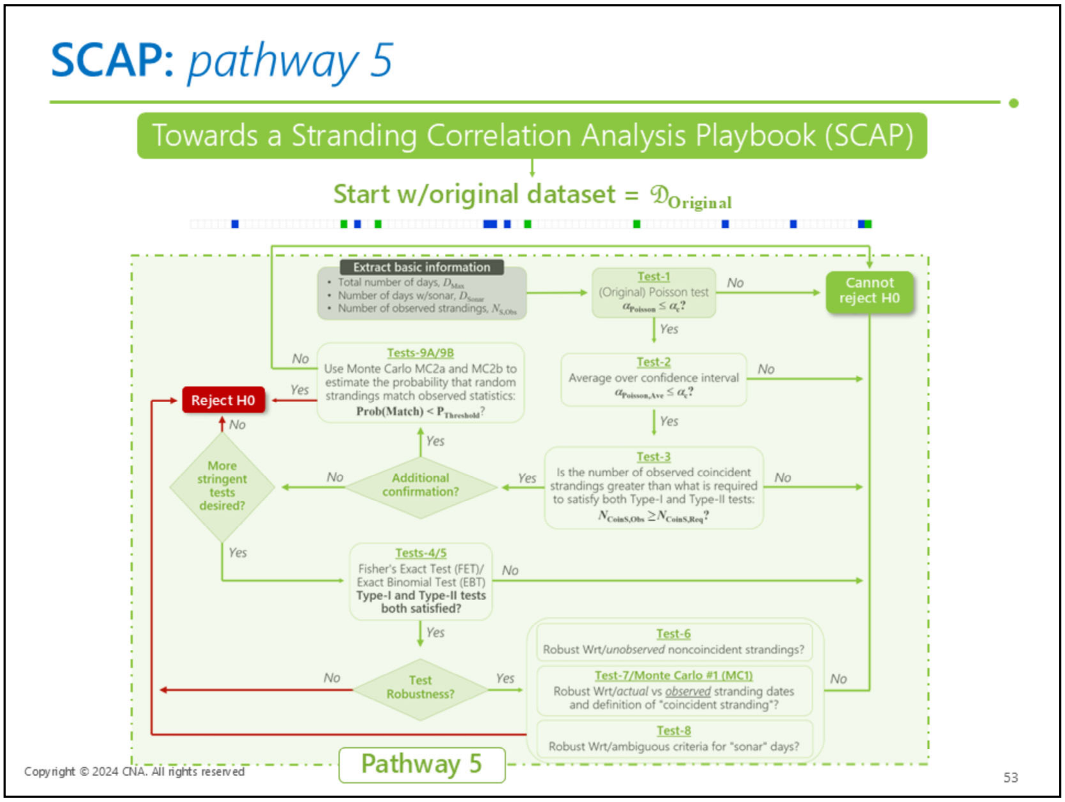
-7/Monte Carlo #1 (MC1)
actual vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

Pathway 4

52

Pathway #4 is similar to Pathway #3, but it is predicated on a stakeholder wishing to apply more stringent statistical tests. If the stakeholder does not wish to do so, the pathway leads to the same conclusion as at the end of Pathway #3 (i.e., reject the null hypothesis). If the stakeholder does wish to do so, the option is to apply both FET and the EBT (see **slide 26** and **Appendix A**).



SCAP: *pathway 5*

Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathfrak{D}_{\mathbf{Original}}$

Extract basic information
- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\text{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\text{Poisson}} \le \alpha_c$?

Cannot reject H0

Test-2
Average over confidence interval
$\alpha_{\text{Poisson,Ave}} \le \alpha_c$?

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\text{CoinS,Obs}} \ge N_{\text{CoinS,Req}}$?

Additional confirmation?

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\mathbf{Prob(Match)} < \mathbf{P}_{\mathbf{Threshold}}$?

Reject H0

More stringent tests desired?

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Test Robustness?

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

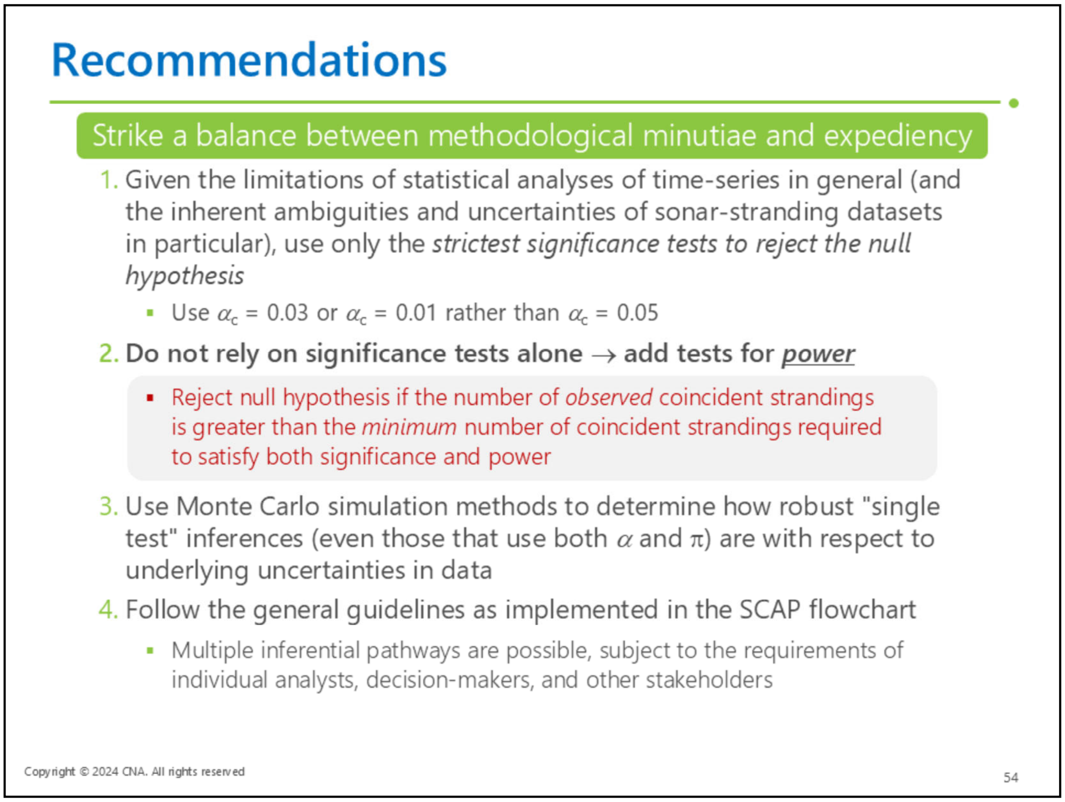
Pathway 5

53

Pathway #5 includes all elements of the complete SCAP. It builds on Pathway #4 by adding an optional three battery of tests of robustness:

1. Test-6 may be used to determine whether the statistical evidence to reject the null hypothesis is robust with respect to unobserved non-coincident strandings (see slide).
2. Test-7 may be used to determine whether the evidence is robust with respect to the uncertainty between actual and observed stranding dates and the ambiguity of how "coincident strandings" are defined (see slide).
3. Test-8 may be used to determine whether the evidence is robust with respect to the ambiguous criteria used to define the set of sonar days (see slide).

Passing some or all of these tests (which tests to apply is, as always, at the stakeholder's discretion) warrants a final rejection of the null hypothesis. If all tests fail, the null hypothesis cannot be rejected on the grounds of excessive uncertainties inherent in the original dataset.



Recommendations

Strike a balance between methodological minutiae and expediency

1. Given the limitations of statistical analyses of time-series in general (and the inherent ambiguities and uncertainties of sonar-stranding datasets in particular), use only the *strictest significance tests to reject the null hypothesis*
   - Use $\alpha_c = 0.03$ or $\alpha_c = 0.01$ rather than $\alpha_c = 0.05$
2. **Do not rely on significance tests alone → add tests for *power***
   - Reject null hypothesis if the number of *observed* coincident strandings is greater than the *minimum* number of coincident strandings required to satisfy both significance and power
3. Use Monte Carlo simulation methods to determine how robust "single test" inferences (even those that use both α and π) are with respect to underlying uncertainties in data
4. Follow the general guidelines as implemented in the SCAP flowchart
   - Multiple inferential pathways are possible, subject to the requirements of individual analysts, decision-makers, and other stakeholders

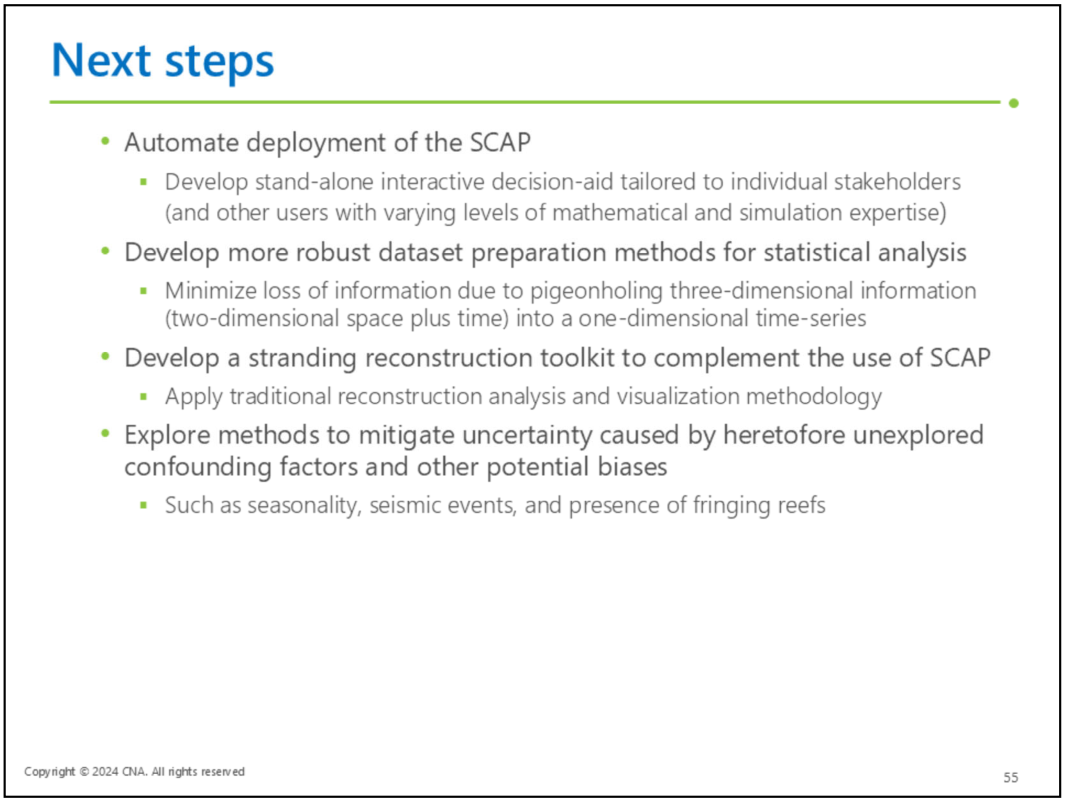
54

This study concludes with four general recommendations. The first is the easiest to implement but also the least far-reaching because it requires only that a more stringent threshold value of significance be applied to existing methodology (which is assumed to otherwise remain the same). This top-level recommendation is made in view of the myriad potential sources of ambiguities and uncertainties inherent in the analysis of sonar-stranding correlations.

The second recommendation is key because it both refines existing methodology (by adding an estimate of power to significance) and modifies the way in which statistical inferences are drawn. Specifically, rather than comparing P-values to some arbitrary threshold, the ***observed* number of coincident strandings is compared to the number of coincident stranding required to satisfy both significance and power**. As long as what constitutes a coincident stranding is unambiguously defined and properly derived from the data, this method is both more intuitive and unequivocal because it respects, and simultaneously minimizes, both Type I (false positive) and Type II (false negative) errors.

The third recommendation is to use any (or all) of the MCS introduced throughout the discussion to determine the degree to which significance and power tests alone are robust to underlying uncertainties in the data.

The final recommendation is overarching. It is to follow the general guidelines in the SCAP. Recall, that Pathway 1 effectively reproduces the existing methodology, whereas Pathways 2 to 5 include an increasingly refined battery of tests and simulations.



Next steps

- Automate deployment of the SCAP
  - Develop stand-alone interactive decision-aid tailored to individual stakeholders (and other users with varying levels of mathematical and simulation expertise)
- Develop more robust dataset preparation methods for statistical analysis
  - Minimize loss of information due to pigeonholing three-dimensional information (two-dimensional space plus time) into a one-dimensional time-series
- Develop a stranding reconstruction toolkit to complement the use of SCAP
  - Apply traditional reconstruction analysis and visualization methodology
- Explore methods to mitigate uncertainty caused by heretofore unexplored confounding factors and other potential biases
  - Such as seasonality, seismic events, and presence of fringing reefs

55

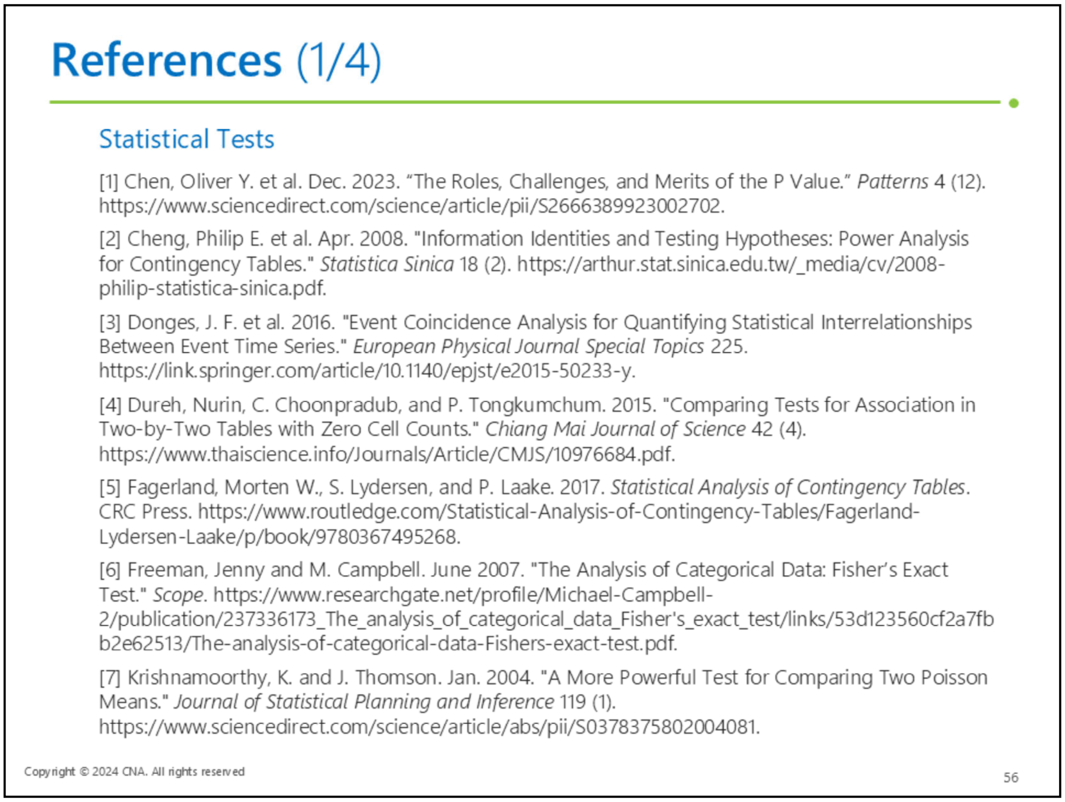
Short discussions of open issues and potential avenues for future exploration and development have been sprinkled throughout the main narrative.

One obvious follow-on effort is to automate the deployment of the SCAP. In its current form (as described on slides 48 to 53), SCAP is a simple at-a-glance flowchart designed to help stakeholders understand and navigate (an often technically cumbersome) inferential process. Although stakeholders may choose to follow different pathways, as determined by their own requirements and individual predilections, the actual analysis (which runs the gamut from calculating the statistics of 2-by-2 contingency tables to setting up, running, and interpreting the output of multiple Monte Carlo simulations) is implied but otherwise not embedded within the SCAP itself. Because all of the elements of SCAP have already been developed for this study (albeit, many in draft form for illustrative purposes only), the text-based flowchart may easily be transformed into a fully interactive stand-alone decision-aid.

Appendix E describes a framework for estimating null stranding rates not from time-series data (as is traditionally done but is riddled with potential bias-generating ambiguities) but by using the combined space plus time data describing the full scenario. Although doing so (beyond outlining one such approach that may be taken) is beyond the scope of this study, it is a natural follow-on research effort that may significantly push the envelope for future sonar-stranding analysis.

Stranding analyses may also be enhanced by developing a stand-alone reconstruction toolkit and embedding it within SCAP. By *reconstruction*, we mean the analysis that is currently done piecemeal, depending on the availability and quality of data, to credibly and confidently distinguish between null and coincident strandings.

Another follow-on effort is to explore methods to mitigate uncertainties resulting from various confounding factors. Although such methods are also intrinsically statistical and well known, they have rarely been applied to sonar-stranding analyses.



References (1/4)

Statistical Tests

[1] Chen, Oliver Y. et al. Dec. 2023. “The Roles, Challenges, and Merits of the P Value.” *Patterns* 4 (12). https://www.sciencedirect.com/science/article/pii/S2666389923002702.

[2] Cheng, Philip E. et al. Apr. 2008. "Information Identities and Testing Hypotheses: Power Analysis for Contingency Tables." *Statistica Sinica* 18 (2). https://arthur.stat.sinica.edu.tw/_media/cv/2008-philip-statistica-sinica.pdf.

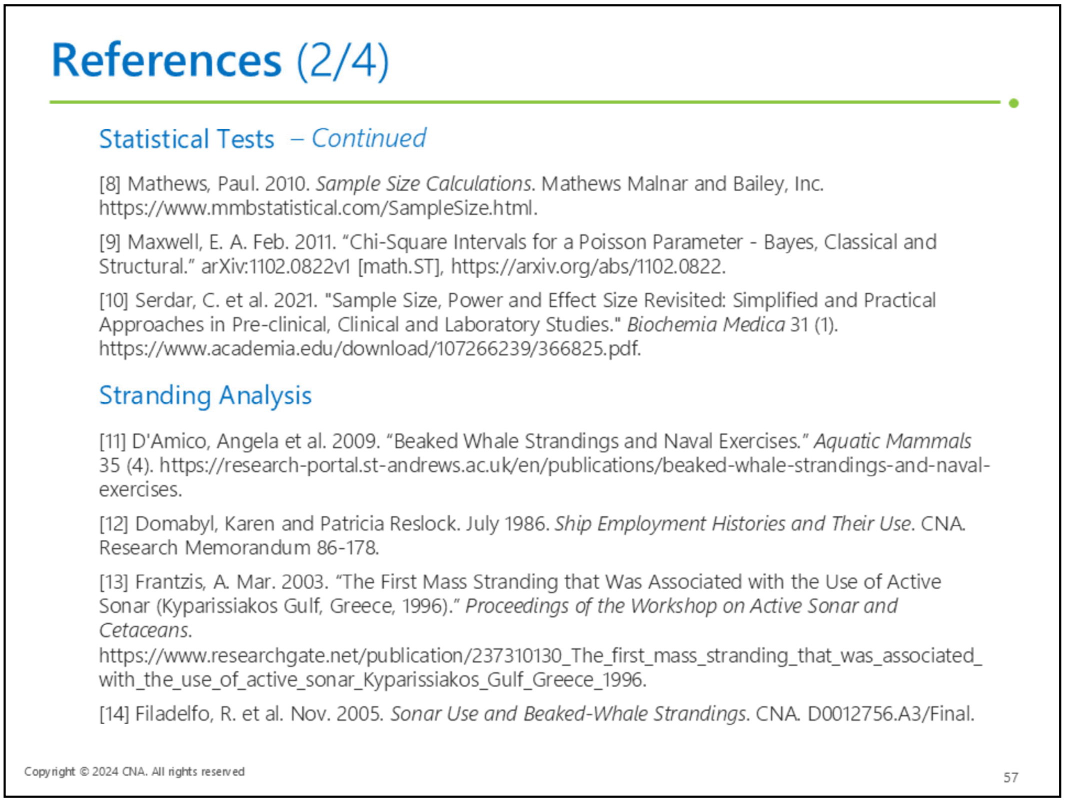
[3] Donges, J. F. et al. 2016. "Event Coincidence Analysis for Quantifying Statistical Interrelationships Between Event Time Series." *European Physical Journal Special Topics* 225. https://link.springer.com/article/10.1140/epjst/e2015-50233-y.

[4] Dureh, Nurin, C. Choonpradub, and P. Tongkumchum. 2015. "Comparing Tests for Association in Two-by-Two Tables with Zero Cell Counts." *Chiang Mai Journal of Science* 42 (4). https://www.thaiscience.info/Journals/Article/CMJS/10976684.pdf.

[5] Fagerland, Morten W., S. Lydersen, and P. Laake. 2017. *Statistical Analysis of Contingency Tables*. CRC Press. https://www.routledge.com/Statistical-Analysis-of-Contingency-Tables/Fagerland-Lydersen-Laake/p/book/9780367495268.

[6] Freeman, Jenny and M. Campbell. June 2007. "The Analysis of Categorical Data: Fisher’s Exact Test." *Scope*. https://www.researchgate.net/profile/Michael-Campbell-2/publication/237336173_The_analysis_of_categorical_data_Fisher's_exact_test/links/53d123560cf2a7fbb2e62513/The-analysis-of-categorical-data-Fishers-exact-test.pdf.

[7] Krishnamoorthy, K. and J. Thomson. Jan. 2004. "A More Powerful Test for Comparing Two Poisson Means." *Journal of Statistical Planning and Inference* 119 (1). https://www.sciencedirect.com/science/article/abs/pii/S0378375802004081.



References (2/4)

Statistical Tests *– Continued*

[8] Mathews, Paul. 2010. *Sample Size Calculations*. Mathews Malnar and Bailey, Inc. https://www.mmbstatistical.com/SampleSize.html.

[9] Maxwell, E. A. Feb. 2011. “Chi-Square Intervals for a Poisson Parameter - Bayes, Classical and Structural.” arXiv:1102.0822v1 [math.ST], https://arxiv.org/abs/1102.0822.

[10] Serdar, C. et al. 2021. "Sample Size, Power and Effect Size Revisited: Simplified and Practical Approaches in Pre-clinical, Clinical and Laboratory Studies." *Biochemia Medica* 31 (1). https://www.academia.edu/download/107266239/366825.pdf.

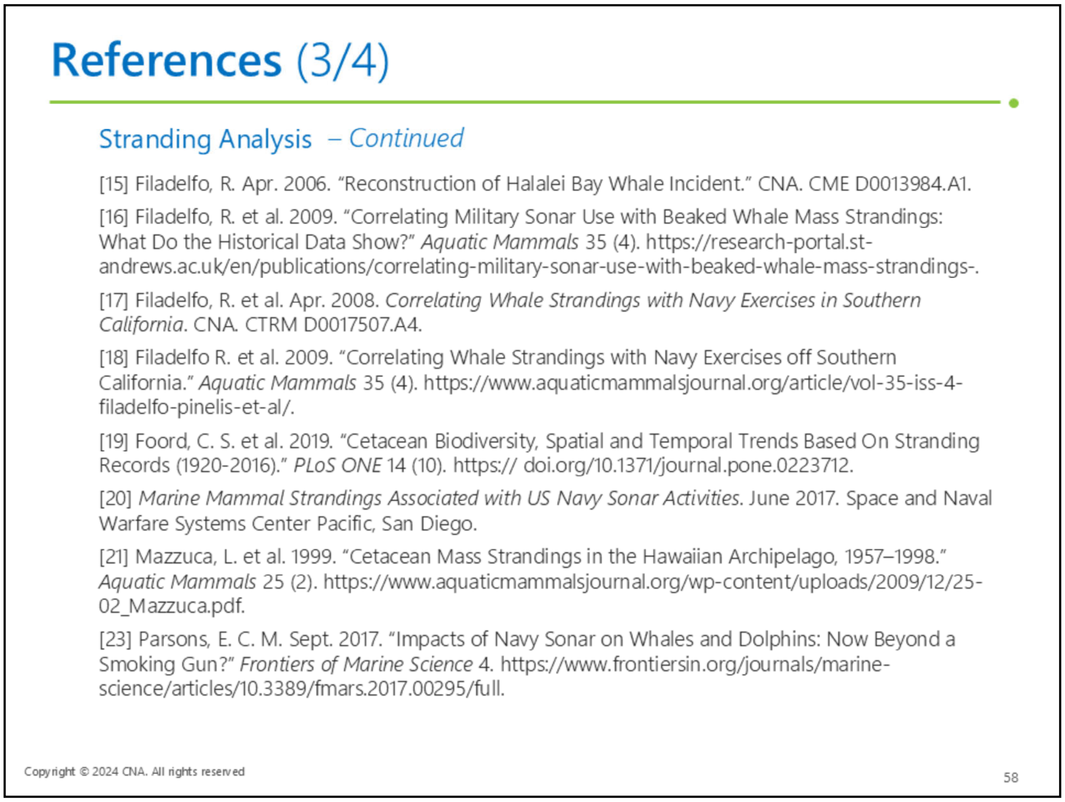
Stranding Analysis

[11] D'Amico, Angela et al. 2009. “Beaked Whale Strandings and Naval Exercises.” *Aquatic Mammals* 35 (4). https://research-portal.st-andrews.ac.uk/en/publications/beaked-whale-strandings-and-naval-exercises.

[12] Domabyl, Karen and Patricia Reslock. July 1986. *Ship Employment Histories and Their Use*. CNA. Research Memorandum 86-178.

[13] Frantzis, A. Mar. 2003. “The First Mass Stranding that Was Associated with the Use of Active Sonar (Kyparissiakos Gulf, Greece, 1996).” *Proceedings of the Workshop on Active Sonar and Cetaceans*. https://www.researchgate.net/publication/237310130_The_first_mass_stranding_that_was_associated_with_the_use_of_active_sonar_Kyparissiakos_Gulf_Greece_1996.

[14] Filadelfo, R. et al. Nov. 2005. *Sonar Use and Beaked-Whale Strandings*. CNA. D0012756.A3/Final.



References (3/4)

Stranding Analysis – *Continued*

[15] Filadelfo, R. Apr. 2006. "Reconstruction of Halalei Bay Whale Incident." CNA. CME D0013984.A1.

[16] Filadelfo, R. et al. 2009. "Correlating Military Sonar Use with Beaked Whale Mass Strandings: What Do the Historical Data Show?" *Aquatic Mammals* 35 (4). https://research-portal.st-andrews.ac.uk/en/publications/correlating-military-sonar-use-with-beaked-whale-mass-strandings-.

[17] Filadelfo, R. et al. Apr. 2008. *Correlating Whale Strandings with Navy Exercises in Southern California*. CNA. CTRM D0017507.A4.

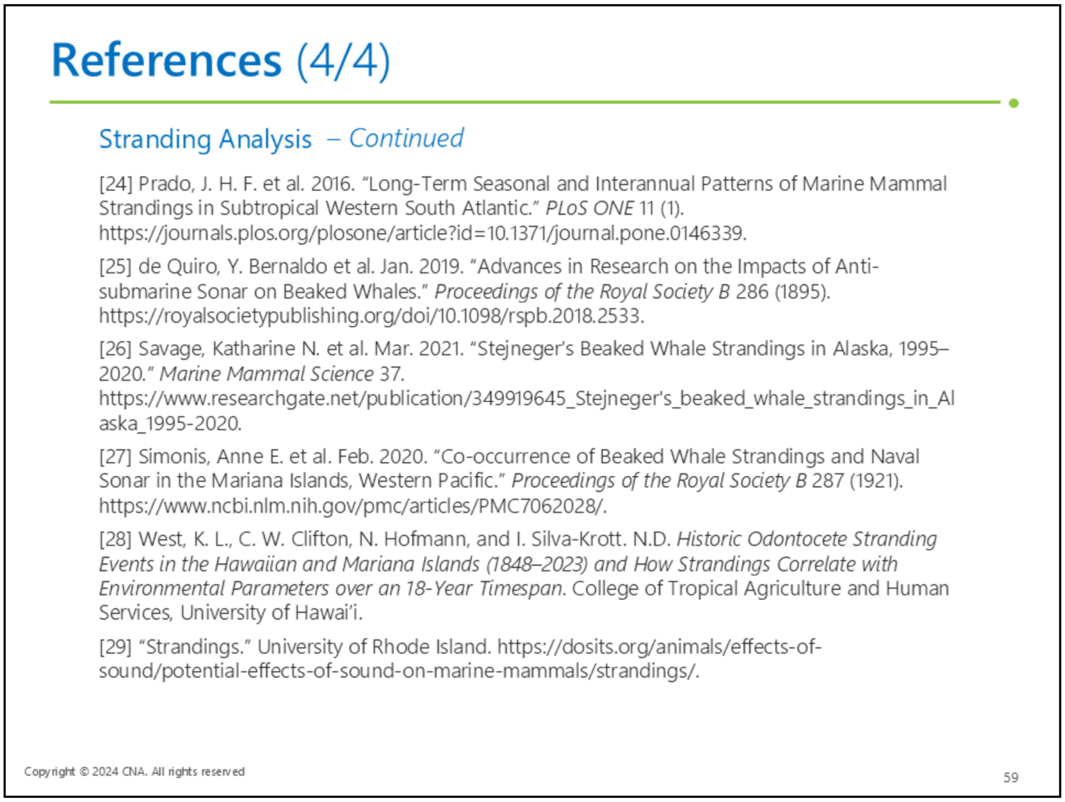
[18] Filadelfo R. et al. 2009. "Correlating Whale Strandings with Navy Exercises off Southern California." *Aquatic Mammals* 35 (4). https://www.aquaticmammalsjournal.org/article/vol-35-iss-4-filadelfo-pinelis-et-al/.

[19] Foord, C. S. et al. 2019. "Cetacean Biodiversity, Spatial and Temporal Trends Based On Stranding Records (1920-2016)." *PLoS ONE* 14 (10). https:// doi.org/10.1371/journal.pone.0223712.

[20] *Marine Mammal Strandings Associated with US Navy Sonar Activities*. June 2017. Space and Naval Warfare Systems Center Pacific, San Diego.

[21] Mazzuca, L. et al. 1999. "Cetacean Mass Strandings in the Hawaiian Archipelago, 1957–1998." *Aquatic Mammals* 25 (2). https://www.aquaticmammalsjournal.org/wp-content/uploads/2009/12/25-02_Mazzuca.pdf.

[23] Parsons, E. C. M. Sept. 2017. "Impacts of Navy Sonar on Whales and Dolphins: Now Beyond a Smoking Gun?" *Frontiers of Marine Science* 4. https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2017.00295/full.



References (4/4)

Stranding Analysis – *Continued*

[24] Prado, J. H. F. et al. 2016. "Long-Term Seasonal and Interannual Patterns of Marine Mammal Strandings in Subtropical Western South Atlantic." *PLoS ONE* 11 (1). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146339.

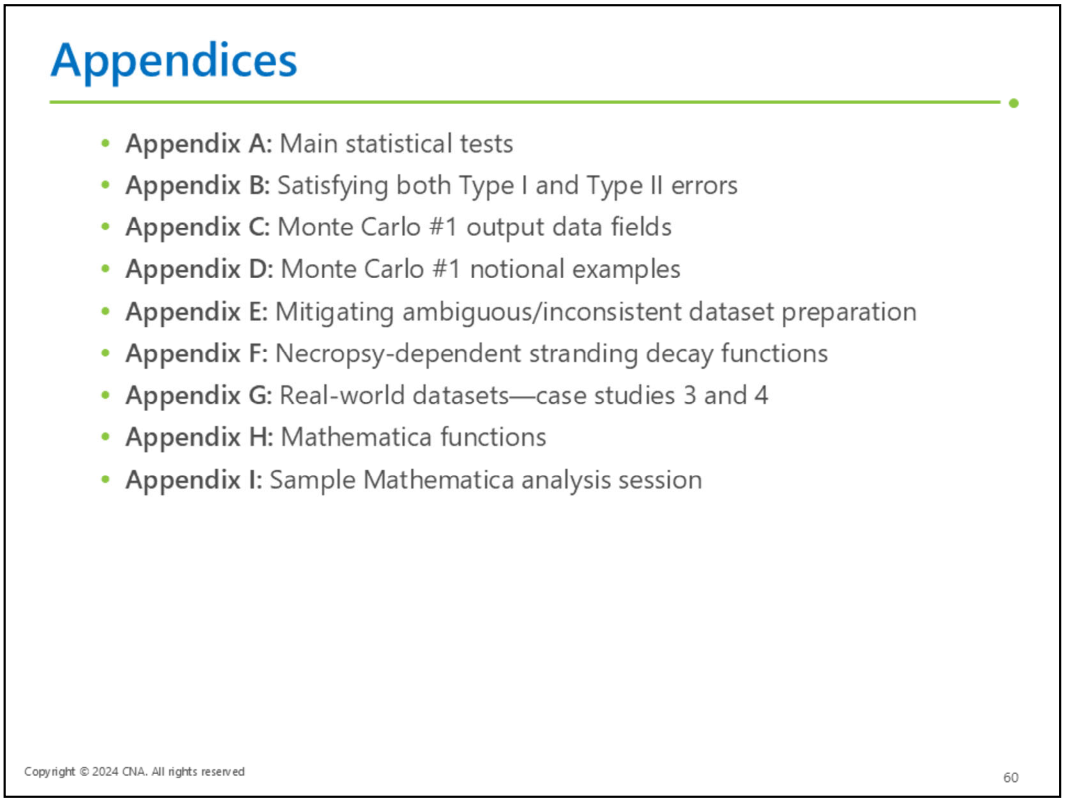
[25] de Quiro, Y. Bernaldo et al. Jan. 2019. "Advances in Research on the Impacts of Anti-submarine Sonar on Beaked Whales." *Proceedings of the Royal Society B* 286 (1895). https://royalsocietypublishing.org/doi/10.1098/rspb.2018.2533.

[26] Savage, Katharine N. et al. Mar. 2021. "Stejneger's Beaked Whale Strandings in Alaska, 1995–2020." *Marine Mammal Science* 37. https://www.researchgate.net/publication/349919645_Stejneger's_beaked_whale_strandings_in_Alaska_1995-2020.

[27] Simonis, Anne E. et al. Feb. 2020. "Co-occurrence of Beaked Whale Strandings and Naval Sonar in the Mariana Islands, Western Pacific." *Proceedings of the Royal Society B* 287 (1921). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7062028/.

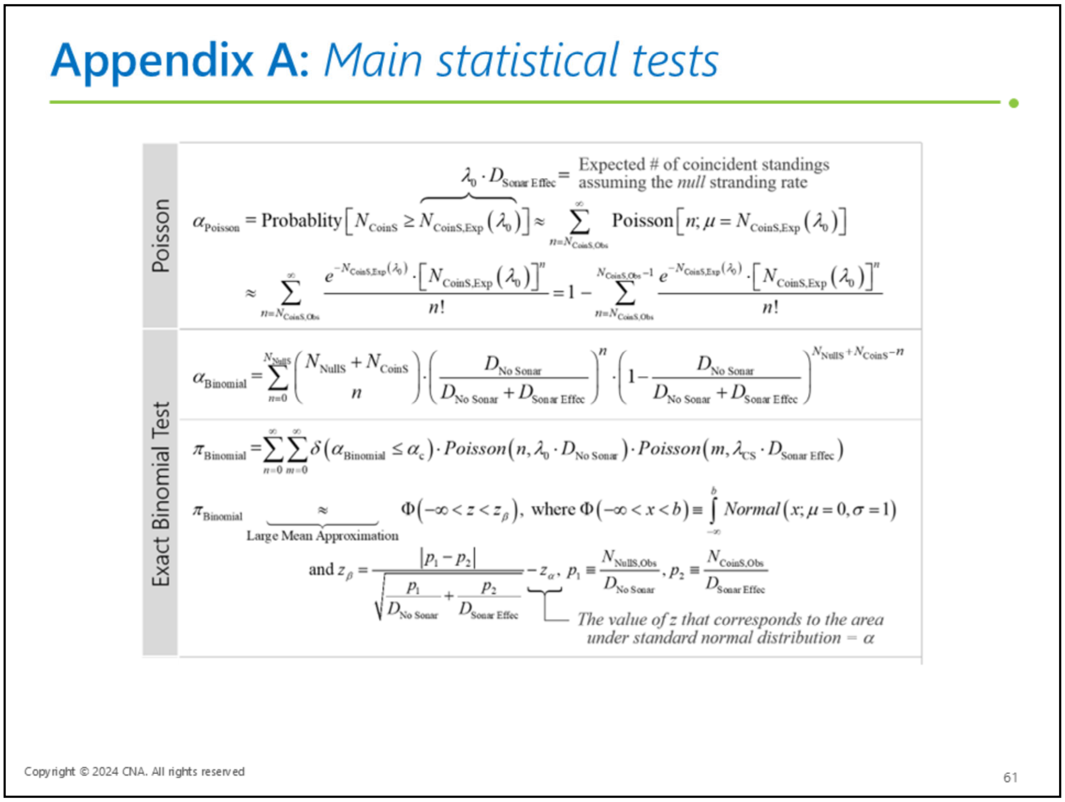
[28] West, K. L., C. W. Clifton, N. Hofmann, and I. Silva-Krott. N.D. *Historic Odontocete Stranding Events in the Hawaiian and Mariana Islands (1848–2023) and How Strandings Correlate with Environmental Parameters over an 18-Year Timespan*. College of Tropical Agriculture and Human Services, University of Hawai'i.

[29] "Strandings." University of Rhode Island. https://dosits.org/animals/effects-of-sound/potential-effects-of-sound-on-marine-mammals/strandings/.



Appendices

- **Appendix A:** Main statistical tests
- **Appendix B:** Satisfying both Type I and Type II errors
- **Appendix C:** Monte Carlo #1 output data fields
- **Appendix D:** Monte Carlo #1 notional examples
- **Appendix E:** Mitigating ambiguous/inconsistent dataset preparation
- **Appendix F:** Necropsy-dependent stranding decay functions
- **Appendix G:** Real-world datasets—case studies 3 and 4
- **Appendix H:** Mathematica functions
- **Appendix I:** Sample Mathematica analysis session



Appendix A: *Main statistical tests*

Poisson

$$\alpha_{\text{Poisson}} = \text{Probablity}\left[N_{\text{CoinS}} \geq \overbrace{N_{\text{CoinS,Exp}}\left(\lambda_0\right)}^{\lambda_0 \cdot D_{\text{Sonar Effec}} = \substack{\text{Expected \# of coincident standings} \\ \text{assuming the } \textit{null} \text{ stranding rate}}}\right] \approx \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right]$$

$$\approx \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \frac{e^{-N_{\text{CoinS,Exp}}(\lambda_0)} \cdot \left[N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right]^n}{n!} = 1 - \sum_{n=N_{\text{CoinS,Obs}}}^{N_{\text{CoinS,Obs}}-1} \frac{e^{-N_{\text{CoinS,Exp}}(\lambda_0)} \cdot \left[N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right]^n}{n!}$$

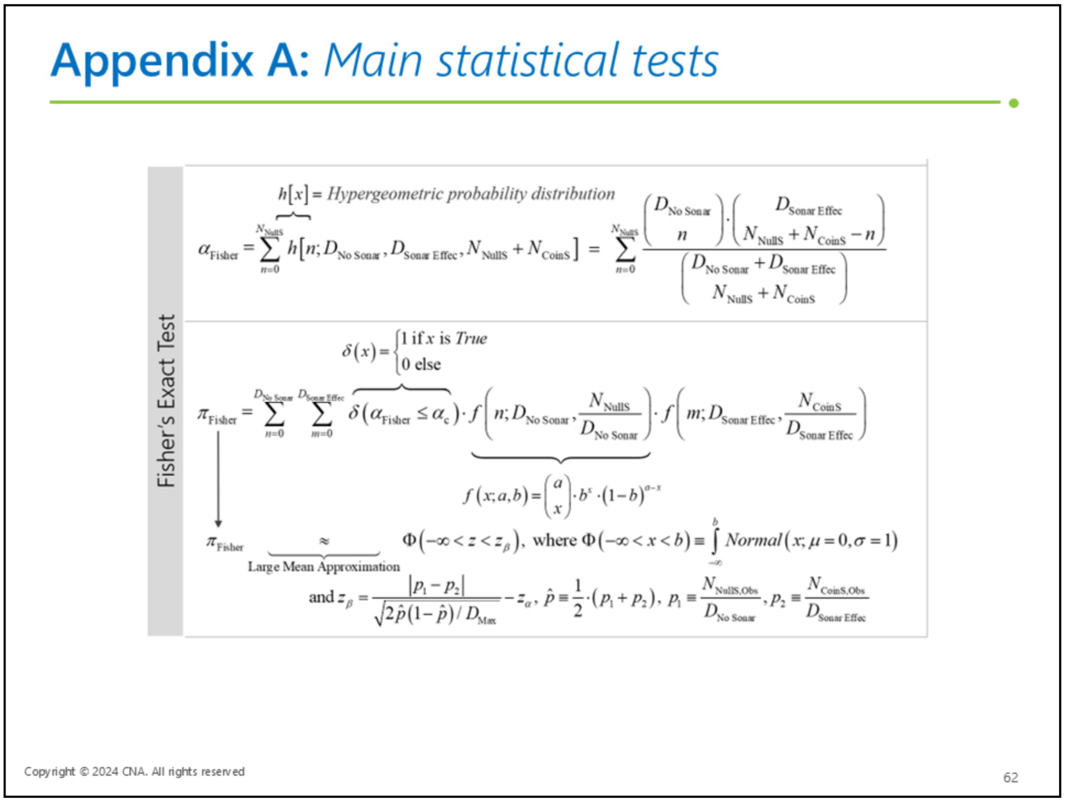
Exact Binomial Test

$$\alpha_{\text{Binomial}} = \sum_{n=0}^{N_{\text{NullS}}} \binom{N_{\text{NullS}} + N_{\text{CoinS}}}{n} \cdot \left(\frac{D_{\text{No Sonar}}}{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}}\right)^n \cdot \left(1 - \frac{D_{\text{No Sonar}}}{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}}\right)^{N_{\text{NullS}} + N_{\text{CoinS}} - n}$$

$$\pi_{\text{Binomial}} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \delta\left(\alpha_{\text{Binomial}} \leq \alpha_{\text{c}}\right) \cdot \mathit{Poisson}\left(n, \lambda_0 \cdot D_{\text{No Sonar}}\right) \cdot \mathit{Poisson}\left(m, \lambda_{\text{CS}} \cdot D_{\text{Sonar Effec}}\right)$$

$$\pi_{\text{Binomial}} \underbrace{\approx}_{\text{Large Mean Approximation}} \Phi\left(-\infty < z < z_\beta\right), \text{ where } \Phi\left(-\infty < x < b\right) \equiv \int_{-\infty}^{b} \mathit{Normal}\left(x; \mu = 0, \sigma = 1\right)$$

$$\text{and } z_\beta = \frac{\left|p_1 - p_2\right|}{\sqrt{\frac{p_1}{D_{\text{No Sonar}}} + \frac{p_2}{D_{\text{Sonar Effec}}}}} \underbrace{- z_\alpha}_{\substack{\textit{The value of z that corresponds to the area} \\ \textit{under standard normal distribution} = \alpha}}, \; p_1 \equiv \frac{N_{\text{NullS,Obs}}}{D_{\text{No Sonar}}}, \; p_2 \equiv \frac{N_{\text{CoinS,Obs}}}{D_{\text{Sonar Effec}}}$$



Appendix A: *Main statistical tests*

Fisher's Exact Test

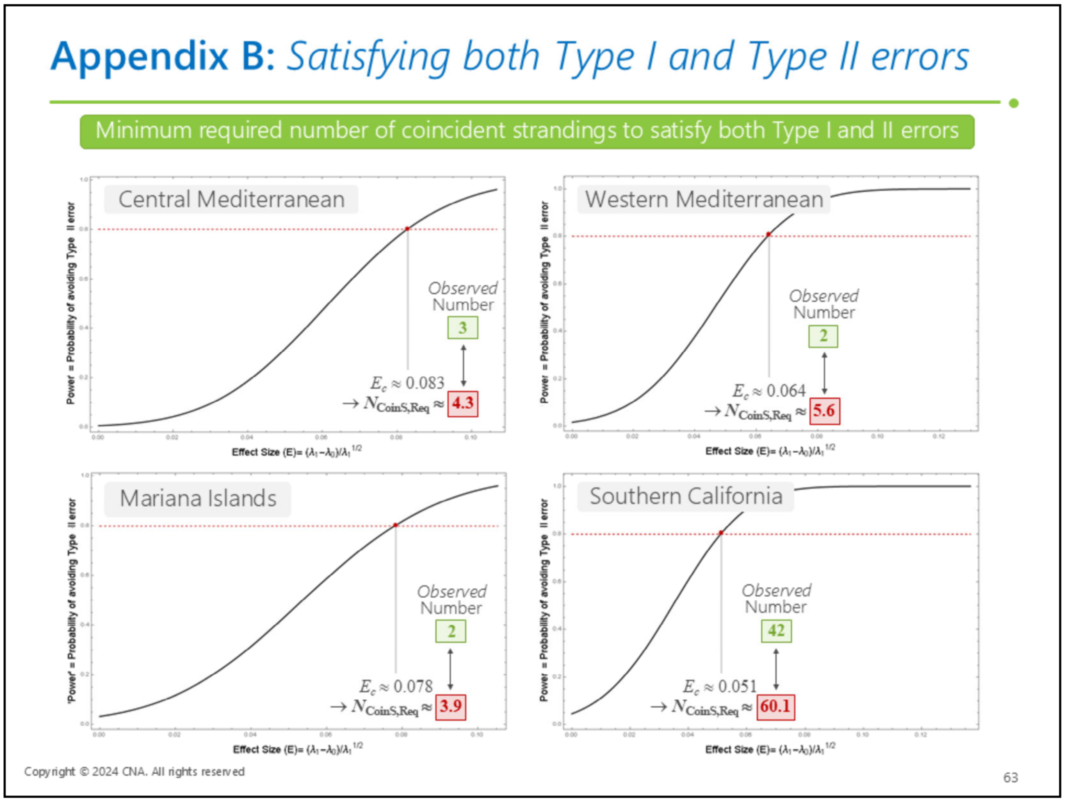
$$\alpha_{\text{Fisher}} = \sum_{n=0}^{N_{\text{NullS}}} \overbrace{h}^{h[x] = \textit{Hypergeometric probability distribution}}\left[n; D_{\text{No Sonar}}, D_{\text{Sonar Effec}}, N_{\text{NullS}} + N_{\text{CoinS}}\right] = \sum_{n=0}^{N_{\text{NullS}}} \frac{\binom{D_{\text{No Sonar}}}{n}\cdot\binom{D_{\text{Sonar Effec}}}{N_{\text{NullS}} + N_{\text{CoinS}} - n}}{\binom{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}}{N_{\text{NullS}} + N_{\text{CoinS}}}}$$

$$\pi_{\text{Fisher}} = \sum_{n=0}^{D_{\text{No Sonar}}} \sum_{m=0}^{D_{\text{Sonar Effec}}} \overbrace{\delta\left(\alpha_{\text{Fisher}} \le \alpha_{\text{c}}\right)}^{\delta(x) = \begin{cases} 1 \text{ if } x \text{ is } \textit{True} \\ 0 \text{ else} \end{cases}} \cdot \underbrace{f\left(n; D_{\text{No Sonar}}, \frac{N_{\text{NullS}}}{D_{\text{No Sonar}}}\right)}_{f(x;a,b) = \binom{a}{x}\cdot b^{x}\cdot(1-b)^{a-x}} \cdot f\left(m; D_{\text{Sonar Effec}}, \frac{N_{\text{CoinS}}}{D_{\text{Sonar Effec}}}\right)$$

$$\Big\downarrow$$

$$\pi_{\text{Fisher}} \underbrace{\approx}_{\text{Large Mean Approximation}} \Phi\left(-\infty < z < z_{\beta}\right), \text{ where } \Phi\left(-\infty < x < b\right) \equiv \int_{-\infty}^{b} \textit{Normal}\left(x; \mu = 0, \sigma = 1\right)$$

$$\text{and } z_{\beta} = \frac{\left|p_1 - p_2\right|}{\sqrt{2\hat{p}\left(1 - \hat{p}\right) / D_{\text{Max}}}} - z_{\alpha},\ \hat{p} \equiv \frac{1}{2}\cdot\left(p_1 + p_2\right),\ p_1 \equiv \frac{N_{\text{NullS,Obs}}}{D_{\text{No Sonar}}}, p_2 \equiv \frac{N_{\text{CoinS,Obs}}}{D_{\text{Sonar Effec}}}$$



Appendix B: *Satisfying both Type I and Type II errors*

Minimum required number of coincident strandings to satisfy both Type I and II errors

Copyright © 2024 CNA. All rights reserved

63

This slide shows estimates of the minimum number of observed coincident strandings required to simultaneously satisfy both Type I (significance) and Type II tests (power); see slides 19 to 22.

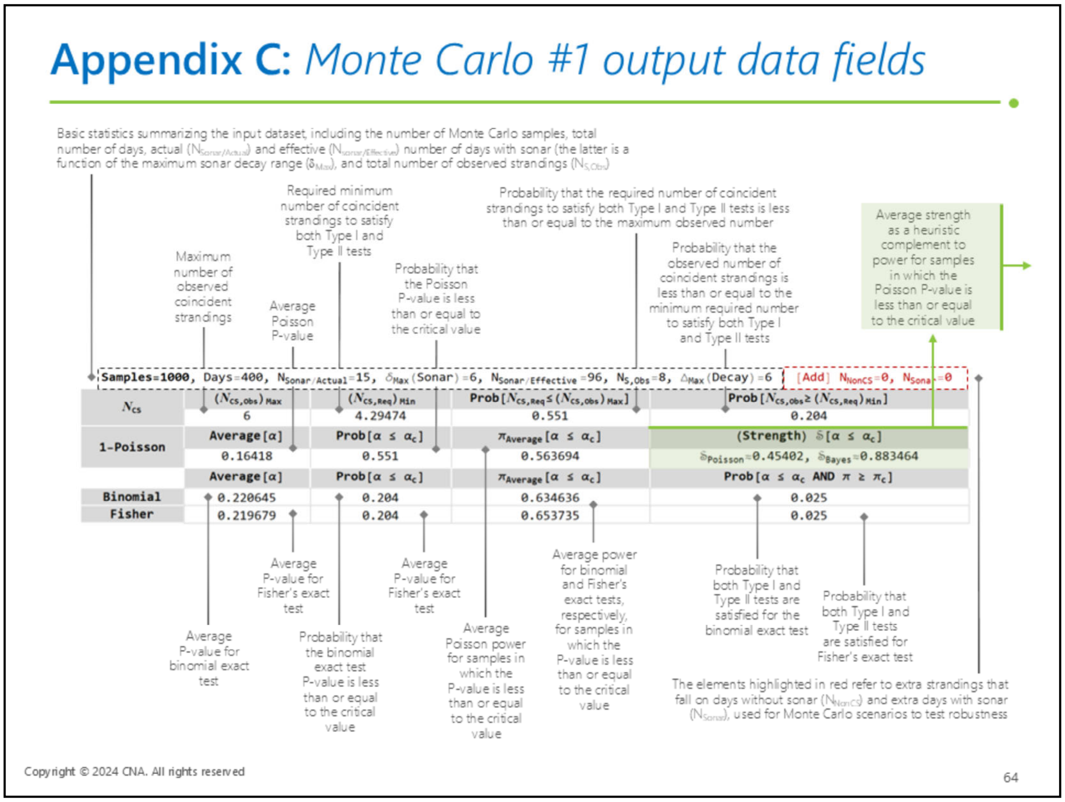
The algorithm consists of two loops. The first loop is to find the minimum null stranding rate for which the significance $\alpha_{\text{Min}} \leq \alpha_{\text{c}}$. Note that whatever this null stranding rate is would not (necessarily) be the one that is observed (i.e., the number of observed strandings on non-sonar days divided by the total number of non-sonar days); it is merely an interim value that must be computed first as part of this combined Type I–Type II test.

Once α_{Min} is estimated, the second loop is to find the minimum number of coincident strandings (technically, this is to find the minimum effect size, as defined on slide 19, which represents the minimum statistically discernable difference between expected and observed coincident strandings) for which the Type II (false negative) test is satisfied.

We produced estimates for four real-world datasets (see slide 40): Central Mediterranean, Western Mediterranean, Mariana Islands, and Southern California.

The **red** dashed lines depict the desired power threshold, π_c=0.8. The value of $N_{\text{CoinS,Req}}$ (highlighted in the **red** box) is the required minimum number of coincident strandings to satisfy both Type I and Type II tests. The observed number is highlighted in the **green** box.

We see that in each case, the number of observed coincident strandings is less than $N_{\text{CoinS,Req}}$. Because none of these datasets simultaneously satisfies both Type I and Type II tests, we conclude that the null hypothesis cannot be rejected for any of them.



Appendix C: *Monte Carlo #1 output data fields*

Basic statistics summarizing the input dataset, including the number of Monte Carlo samples, total number of days, actual ($N_{Sonar/Actual}$) and effective ($N_{sonar/Effective}$) number of days with sonar (the latter is a function of the maximum sonar decay range (δ_{Max}), and total number of observed strandings ($N_{S,Obs}$)

Maximum number of observed coincident strandings

Average Poisson P-value

Required minimum number of coincident strandings to satisfy both Type I and Type II tests

Probability that the Poisson P-value is less than or equal to the critical value

Probability that the required number of coincident strandings to satisfy both Type I and Type II tests is less than or equal to the maximum observed number

Probability that the observed number of coincident strandings is less than or equal to the minimum required number to satisfy both Type I and Type II tests

Average strength as a heuristic complement to power for samples in which the Poisson P-value is less than or equal to the critical value

Samples=1000, Days=400, $N_{Sonar/Actual}$=15, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=96, $N_{S,Obs}$=8, Δ_{Max}(Decay)=6 [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob[$N_{CS,Req} \le (N_{CS,Obs})_{Max}$]	Prob[$N_{CS,Obs} \ge (N_{CS,Req})_{Min}$]
	6	4.29474	0.551	0.204
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{Average}[\alpha \le \alpha_c]$	(Strength) $\mathbb{S}[\alpha \le \alpha_c]$
	0.16418	0.551	0.563694	$\mathbb{S}_{Poisson}$=0.45402, $\mathbb{S}_{Bayes}$=0.883464
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{Average}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.220645	0.204	0.634636	0.025
Fisher	0.219679	0.204	0.653735	0.025

Average P-value for binomial exact test

Average P-value for Fisher's exact test

Probability that the binomial exact test P-value is less than or equal to the critical value

Average P-value for Fisher's exact test

Average Poisson power for samples in which the P-value is less than or equal to the critical value

Average power for binomial and Fisher's exact tests, respectively, for samples in which the P-value is less than or equal to the critical value

Probability that both Type I and Type II tests are satisfied for the binomial exact test

Probability that both Type I and Type II tests are satisfied for Fisher's exact test

The elements highlighted in red refer to extra strandings that fall on days without sonar (N_{NonCS}) and extra days with sonar (N_{Sonar}), used for Monte Carlo scenarios to test robustness

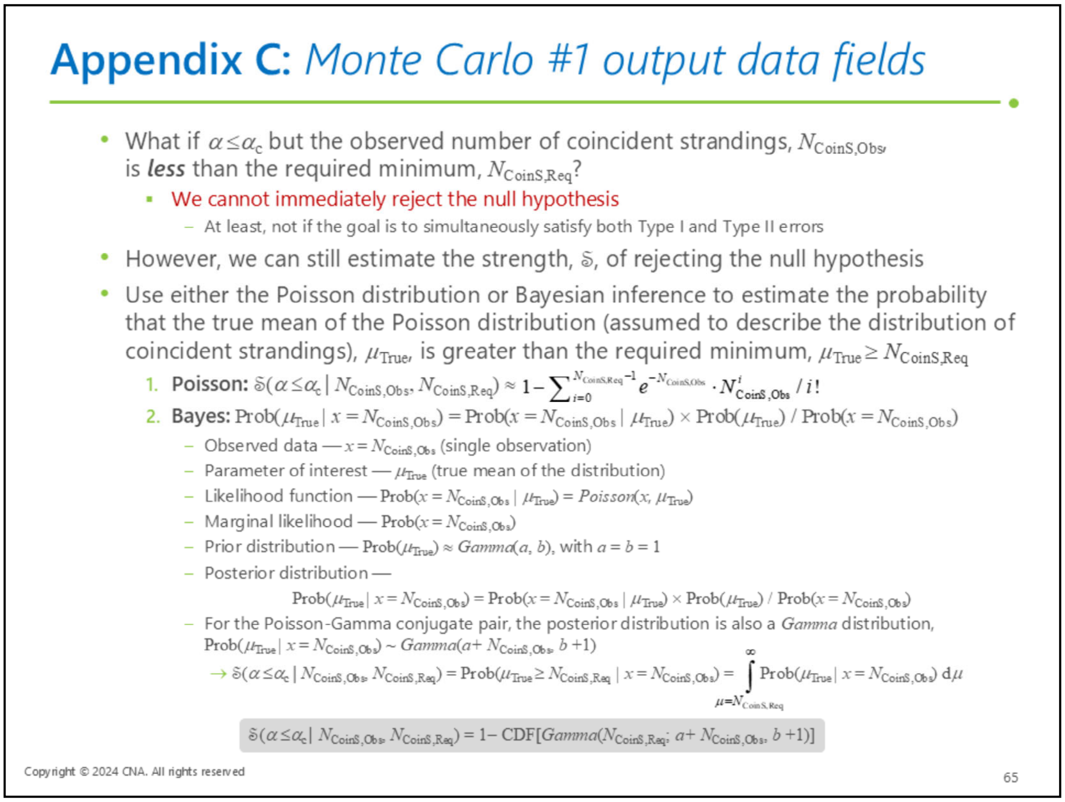
64

This appendix summarizes each of the information fields that appear as part of the output of running MCS #1 (see Slides 30 and 31 and **Appendices H** and **I**).

The top row contains basic statistics summarizing the input data, including the number of Monte Carlo samples, total number of days, actual ($N_{Sonar/Actual}$) and effective ($N_{sonar/Effective}$) number of days with sonar (the latter is a function of the maximum sonar decay range (d_{Max})), and total number of observed strandings ($N_{S,Obs}$).

The remaining rows summarize four types of statistical tests: N_{CS} refers to comparing the observed number of coincident strandings with the minimum number required to satisfy both Type I and Type II tests (see slides 19 and 20), the *1-Poisson* refers to the single-mean Poisson test (see slides 7 and 15); *Binomial* refers to the EBT (see slide 24 and **Appendix A**), and *Fisher* refers to the FET (see slide 26 and **Appendix A**).

The entries highlighted in green on the right of the slide refer to a measure called strength that (loosely speaking) is intended to serve as a complement to statistical power. Although additional details are provided on the next slide, please note that strength is *not* a standard measure, so it is best viewed as an experimental heuristic that we found useful during the early parts of our analysis (although it does not appear in the main narrative). It is included in this appendix for completeness.



Appendix C: *Monte Carlo #1 output data fields*

- What if $\alpha \le \alpha_c$ but the observed number of coincident strandings, $N_{\text{CoinS,Obs}}$, is ***less*** than the required minimum, $N_{\text{CoinS,Req}}$?
  - We cannot immediately reject the null hypothesis
    - At least, not if the goal is to simultaneously satisfy both Type I and Type II errors
- However, we can still estimate the strength, $\mathbb{S}$, of rejecting the null hypothesis
- Use either the Poisson distribution or Bayesian inference to estimate the probability that the true mean of the Poisson distribution (assumed to describe the distribution of coincident strandings), μ_{True}, is greater than the required minimum, $\mu_{\text{True}} \ge N_{\text{CoinS,Req}}$
  1. **Poisson:** $\mathbb{S}(\alpha \le \alpha_c \mid N_{\text{CoinS,Obs}}, N_{\text{CoinS,Req}}) \approx 1 - \sum_{i=0}^{N_{\text{CoinS,Req}}-1} e^{-N_{\text{CoinS,Obs}}} \cdot N_{\text{CoinS,Obs}}^{i} / i!$
  2. **Bayes:** $\text{Prob}(\mu_{\text{True}} \mid x = N_{\text{CoinS,Obs}}) = \text{Prob}(x = N_{\text{CoinS,Obs}} \mid \mu_{\text{True}}) \times \text{Prob}(\mu_{\text{True}}) / \text{Prob}(x = N_{\text{CoinS,Obs}})$
     - Observed data — $x = N_{\text{CoinS,Obs}}$ (single observation)
     - Parameter of interest — μ_{True} (true mean of the distribution)
     - Likelihood function — $\text{Prob}(x = N_{\text{CoinS,Obs}} \mid \mu_{\text{True}}) = \mathit{Poisson}(x, \mu_{\text{True}})$
     - Marginal likelihood — $\text{Prob}(x = N_{\text{CoinS,Obs}})$
     - Prior distribution — $\text{Prob}(\mu_{\text{True}}) \approx \mathit{Gamma}(a, b)$, with $a = b = 1$
     - Posterior distribution —

       $$\text{Prob}(\mu_{\text{True}} \mid x = N_{\text{CoinS,Obs}}) = \text{Prob}(x = N_{\text{CoinS,Obs}} \mid \mu_{\text{True}}) \times \text{Prob}(\mu_{\text{True}}) / \text{Prob}(x = N_{\text{CoinS,Obs}})$$
     - For the Poisson-Gamma conjugate pair, the posterior distribution is also a *Gamma* distribution, $\text{Prob}(\mu_{\text{True}} \mid x = N_{\text{CoinS,Obs}}) \sim \mathit{Gamma}(a + N_{\text{CoinS,Obs}}, b + 1)$

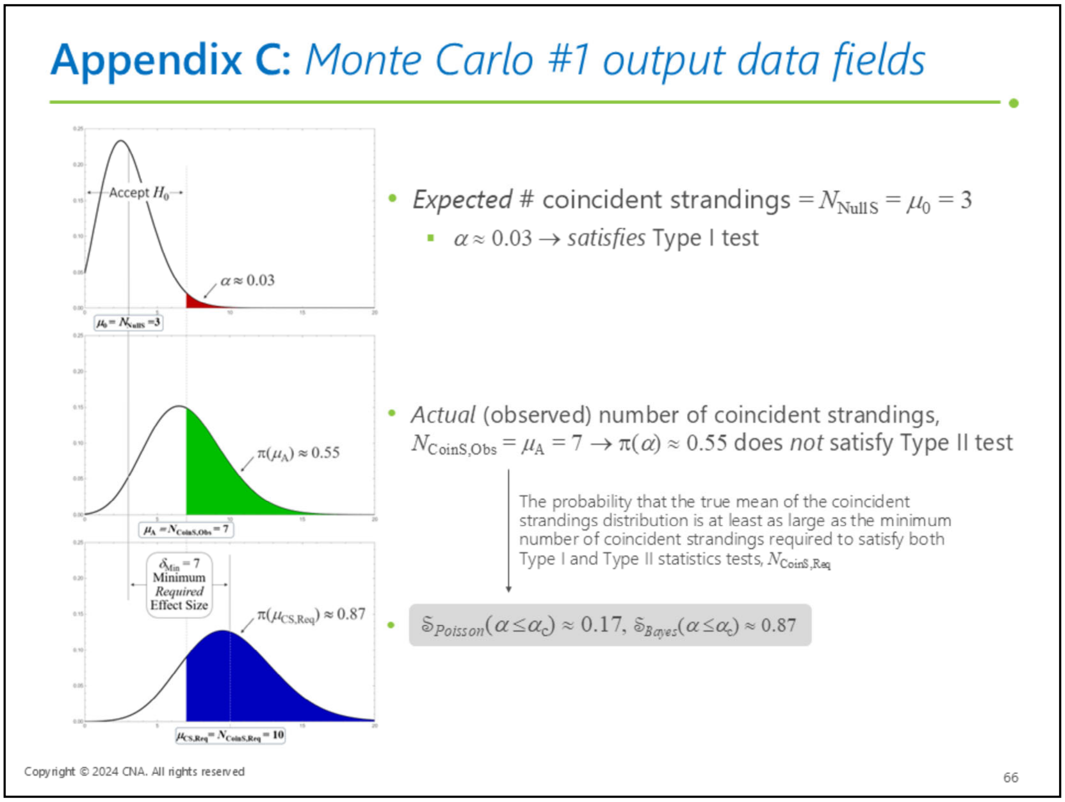
       $\rightarrow \mathbb{S}(\alpha \le \alpha_c \mid N_{\text{CoinS,Obs}}, N_{\text{CoinS,Req}}) = \text{Prob}(\mu_{\text{True}} \ge N_{\text{CoinS,Req}} \mid x = N_{\text{CoinS,Obs}}) = \int_{\mu = N_{\text{CoinS,Req}}}^{\infty} \text{Prob}(\mu_{\text{True}} \mid x = N_{\text{CoinS,Obs}})\, d\mu$

$$\mathbb{S}(\alpha \le \alpha_c \mid N_{\text{CoinS,Obs}}, N_{\text{CoinS,Req}}) = 1 - \text{CDF}[\mathit{Gamma}(N_{\text{CoinS,Req}}; a + N_{\text{CoinS,Obs}}, b + 1)]$$

65

As discussed on slide 18 of the main narrative, one of the limitations of single-mean Poisson test is that the statistical power of rejecting the null hypothesis never exceeds ~0.63, which is far short of typically used thresholds (e.g., 0.80 or 0.85). This limitation is an artifact of the Poisson distribution and results from the fact that the best estimate to use for the alternative hypothesis is the observed number of coincident strandings. As a result, the power to reject the null hypothesis can never pass the Type II test (for thresholds over 0.63), even when significance is far below the required limit (say, 0.05). Now, this does not mean the Type I and Type II tests cannot be simultaneously satisfied; rather, it means that the power of *an already administered Type I test* (that results in an $\alpha \le \alpha_c$) will always be less than is typically required. This issue is mitigated by following the recommendation made in the main narrative, as appears in Pathways two to five in the SCAP (see slides 47 to 51). Specifically, the recommendation is to estimate the minimum number of observed coincident strandings required to simultaneously satisfy both Type I (significance) and Type II tests (power). By doing so, power is not computed at a *given* value of α (based on the null stranding rate) but rather on the basis of whatever null stranding rate (possibly less than that which is actually observed) yields the desired threshold, α_c.

If stakeholders wish to continue to base their decisions to reject on significance (rather than comparing the observed and required number of coincident strandings), one way is to use strength as a heuristic complement of power. Strength estimates the probability that the true Poisson mean, μ_{True}, that describes the distribution of coincident strandings is greater than the *required* minimum. The greater the strength, the greater the likelihood that the true mean of the Poisson distribution entailed by the observed number of coincident strandings will be at least as great as required to satisfy both Type I and Type II tests.

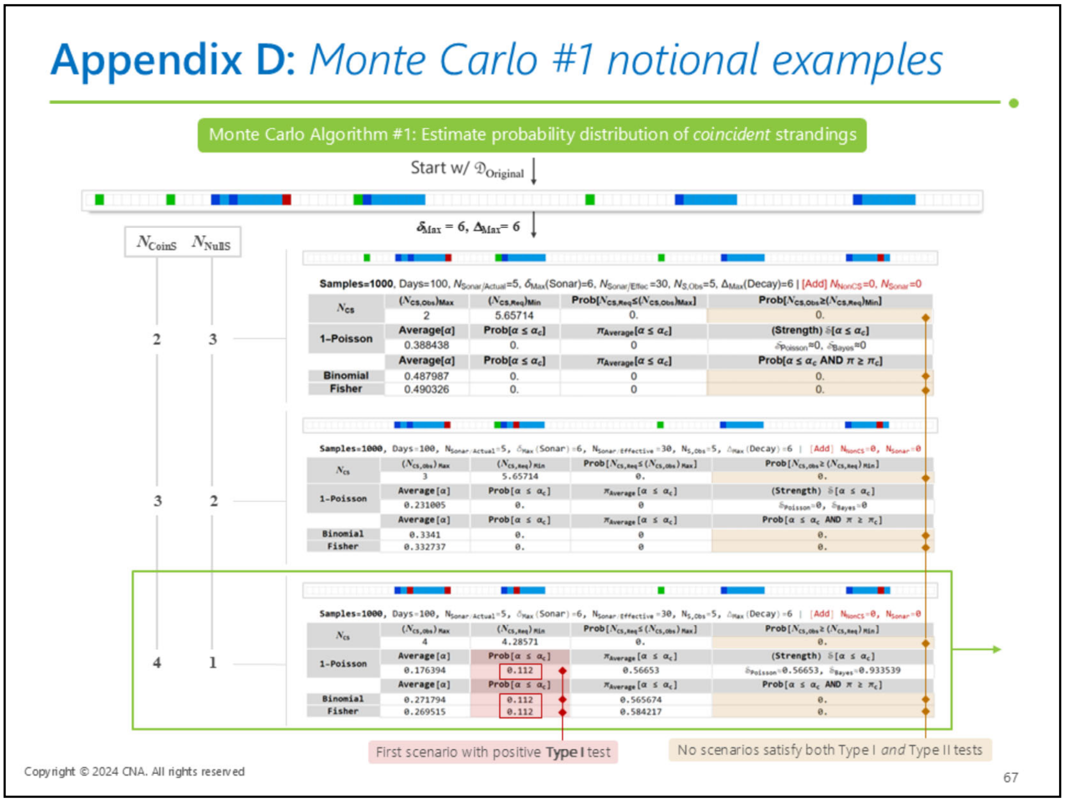


Appendix C: *Monte Carlo #1 output data fields*

- *Expected* # coincident strandings $= N_{\text{Null S}} = \mu_0 = 3$
  - $\alpha \approx 0.03 \rightarrow$ *satisfies* Type I test
- *Actual* (observed) number of coincident strandings, $N_{\text{CoinS,Obs}} = \mu_A = 7 \rightarrow \pi(\alpha) \approx 0.55$ does *not* satisfy Type II test

The probability that the true mean of the coincident strandings distribution is at least as large as the minimum number of coincident strandings required to satisfy both Type I and Type II statistics tests, $N_{\text{CoinS,Req}}$

- $\mathbb{S}_{Poisson}(\alpha \le \alpha_c) \approx 0.17$, $\mathbb{S}_{Bayes}(\alpha \le \alpha_c) \approx 0.87$



Appendix D: *Monte Carlo #1 notional examples*

Monte Carlo Algorithm #1: Estimate probability distribution of *coincident* strandings

Start w/ $\mathfrak{D}_{\text{Original}}$

$\delta_{\text{Max}} = 6, \Delta_{\text{Max}} = 6$

N_{CoinS}	N_{NullS}
2	3
3	2
4	1

Samples=1000, Days=100, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$=30, $N_{S,\text{Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}$]
	2	5.65714	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.388438	0.	0	δ_{Poisson}≈0, δ_{Bayes}≈0
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.487987	0.	0	0.
Fisher	0.490326	0.	0	0.

Samples=1000, Days=100, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=30, $N_{S,\text{Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

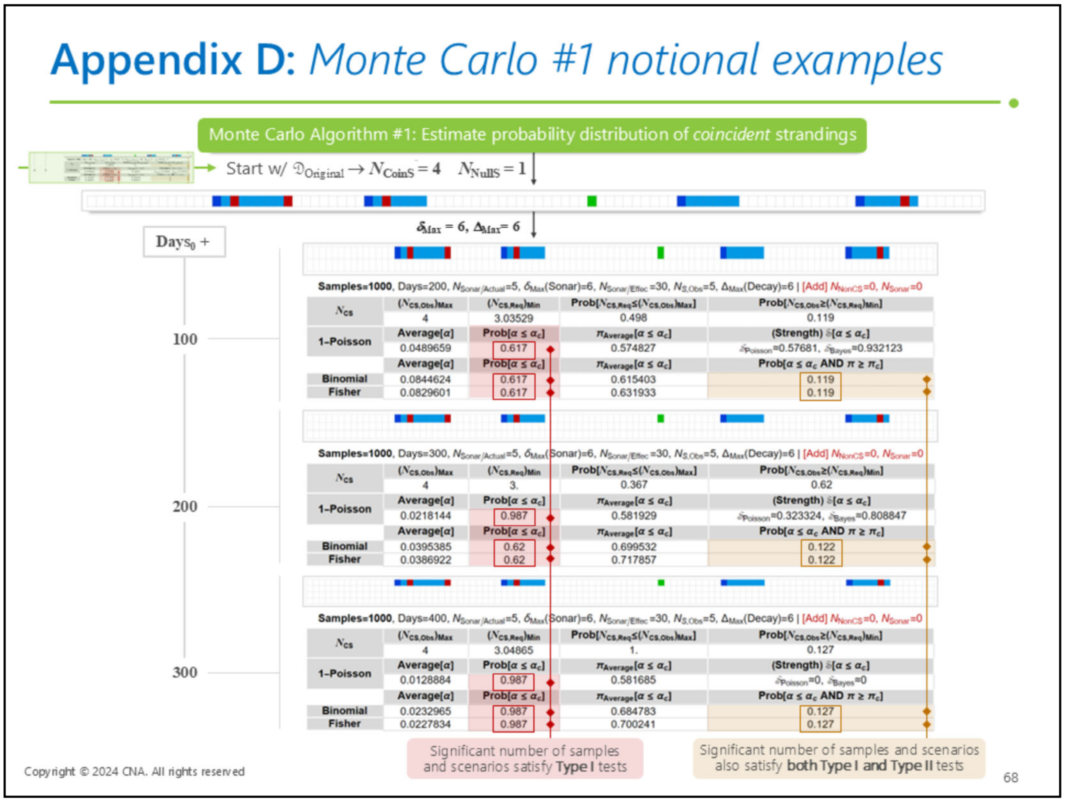
N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}$]
	3	5.65714	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.231005	0.	0	δ_{Poisson}=0, δ_{Bayes}=0
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.3341	0.	0	0.
Fisher	0.332737	0.	0	0.

Samples=1000, Days=100, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=30, $N_{S,\text{Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}$]
	4	4.28571	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.176394	0.112	0.56653	δ_{Poisson}=0.56653, δ_{Bayes}=0.933539
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.271794	0.112	0.565674	0.
Fisher	0.269515	0.112	0.584217	0.

First scenario with positive **Type I** test

No scenarios satisfy both Type I *and* Type II tests



Appendix D: *Monte Carlo #1 notional examples*

Monte Carlo Algorithm #1: Estimate probability distribution of *coincident* strandings

Start w/ $\mathfrak{D}_{\mathrm{Original}} \rightarrow N_{\mathrm{CoinS}} = 4 \quad N_{\mathrm{NullS}} = 1$

$\delta_{\mathrm{Max}} = 6,\ \Delta_{\mathrm{Max}} = 6$

Samples=1000, Days=200, $N_{\mathrm{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\mathrm{Sonar/Effec}}$=30, $N_{S,\mathrm{Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\mathrm{CS,Obs}})_{\mathrm{Max}}$	$(N_{\mathrm{CS,Req}})_{\mathrm{Min}}$	Prob[$N_{\mathrm{CS,Req}} \le (N_{\mathrm{CS,Obs}})_{\mathrm{Max}}$]	Prob[$N_{\mathrm{CS,Obs}} \ge (N_{\mathrm{CS,Req}})_{\mathrm{Min}}$]
	4	3.03529	0.498	0.119
1–Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\mathrm{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.0489659	0.617	0.574827	$\delta_{\mathrm{Poisson}}$≈0.57681, δ_{Bayes}≈0.932123
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\mathrm{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.0844624	0.617	0.615403	0.119
Fisher	0.0829601	0.617	0.631933	0.119

Samples=1000, Days=300, $N_{\mathrm{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\mathrm{Sonar/Effec}}$=30, $N_{S,\mathrm{Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

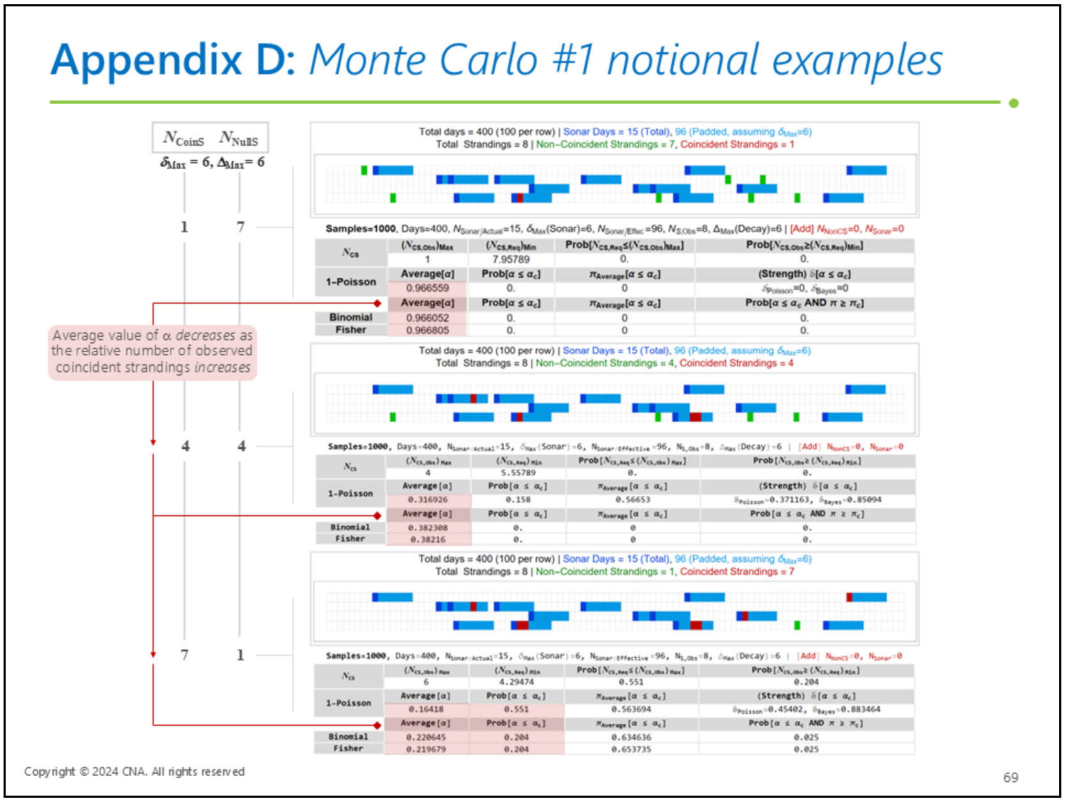
N_{CS}	$(N_{\mathrm{CS,Obs}})_{\mathrm{Max}}$	$(N_{\mathrm{CS,Req}})_{\mathrm{Min}}$	Prob[$N_{\mathrm{CS,Req}} \le (N_{\mathrm{CS,Obs}})_{\mathrm{Max}}$]	Prob[$N_{\mathrm{CS,Obs}} \ge (N_{\mathrm{CS,Req}})_{\mathrm{Min}}$]
	4	3.	0.367	0.62
1–Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\mathrm{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.0218144	0.987	0.581929	$\delta_{\mathrm{Poisson}}$≈0.323324, δ_{Bayes}≈0.808847
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\mathrm{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.0395385	0.62	0.699532	0.122
Fisher	0.0386922	0.62	0.717857	0.122

Samples=1000, Days=400, $N_{\mathrm{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\mathrm{Sonar/Effec}}$=30, $N_{S,\mathrm{Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\mathrm{CS,Obs}})_{\mathrm{Max}}$	$(N_{\mathrm{CS,Req}})_{\mathrm{Min}}$	Prob[$N_{\mathrm{CS,Req}} \le (N_{\mathrm{CS,Obs}})_{\mathrm{Max}}$]	Prob[$N_{\mathrm{CS,Obs}} \ge (N_{\mathrm{CS,Req}})_{\mathrm{Min}}$]
	4	3.04865	1.	0.127
1–Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\mathrm{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.0128884	0.987	0.581685	$\delta_{\mathrm{Poisson}}$≈0, δ_{Bayes}≈0
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\mathrm{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.0232965	0.987	0.684783	0.127
Fisher	0.0227834	0.987	0.700241	0.127

Significant number of samples and scenarios satisfy **Type I** tests

Significant number of samples and scenarios also satisfy **both Type I and Type II** tests



Appendix D: *Monte Carlo #1 notional examples*

N_{CoinS}	N_{NullS}
1	7
4	4
7	1

$\delta_{\text{Max}} = 6$, $\Delta_{\text{Max}} = 6$

Average value of α *decreases* as the relative number of observed coincident strandings *increases*

Total days = 400 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non-Coincident Strandings = 7, Coincident Strandings = 1

Samples=1000, Days=400, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$=96, $N_{\text{S,Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}$]
	1	7.95789	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.966559	0.	0	δ_{Poisson}=0, δ_{Bayes}=0
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.966052	0.	0	0.
Fisher	0.966805	0.	0	0.

Total days = 400 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non-Coincident Strandings = 4, Coincident Strandings = 4

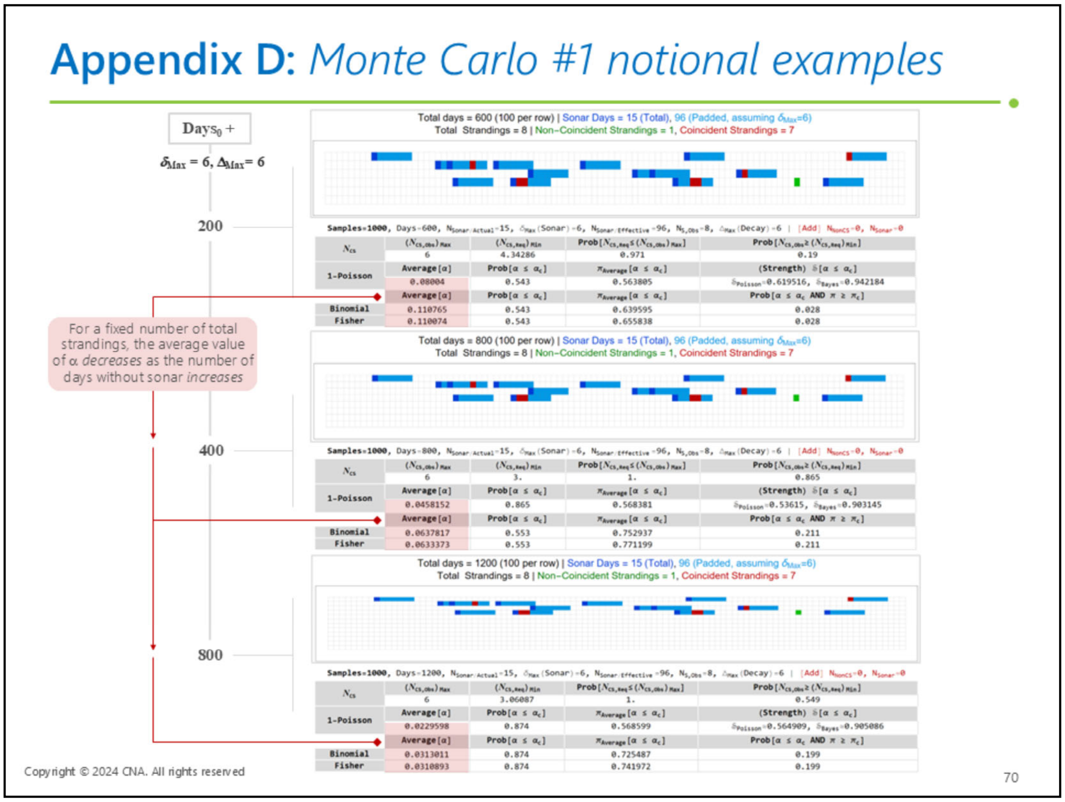
Samples=1000, Days=400, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=96, $N_{\text{S,Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}$]
	4	5.55789	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.316926	0.158	0.56653	δ_{Poisson}=0.371163, δ_{Bayes}=0.85094
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.382308	0.	0	0.
Fisher	0.38216	0.	0	0.

Total days = 400 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non-Coincident Strandings = 1, Coincident Strandings = 7

Samples=1000, Days=400, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=96, $N_{\text{S,Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}$]
	6	4.29474	0.551	0.204
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.16418	0.551	0.563694	δ_{Poisson}=0.45402, δ_{Bayes}=0.883464
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.220645	0.204	0.634636	0.025
Fisher	0.219679	0.204	0.653735	0.025



Appendix D: *Monte Carlo #1 notional examples*

For a fixed number of total strandings, the average value of α *decreases* as the number of days without sonar *increases*

Total days = 600 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non-Coincident Strandings = 1, Coincident Strandings = 7

Samples=1000, Days=600, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=96, $N_{S,\text{Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob$[N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}]$	Prob$[N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}]$
	6	4.34286	0.971	0.19
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.08004	0.543	0.563805	δ_{Poisson}=0.619516, δ_{Bayes}=0.942184
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.110765	0.543	0.639595	0.028
Fisher	0.110074	0.543	0.655838	0.028

Total days = 800 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non-Coincident Strandings = 1, Coincident Strandings = 7

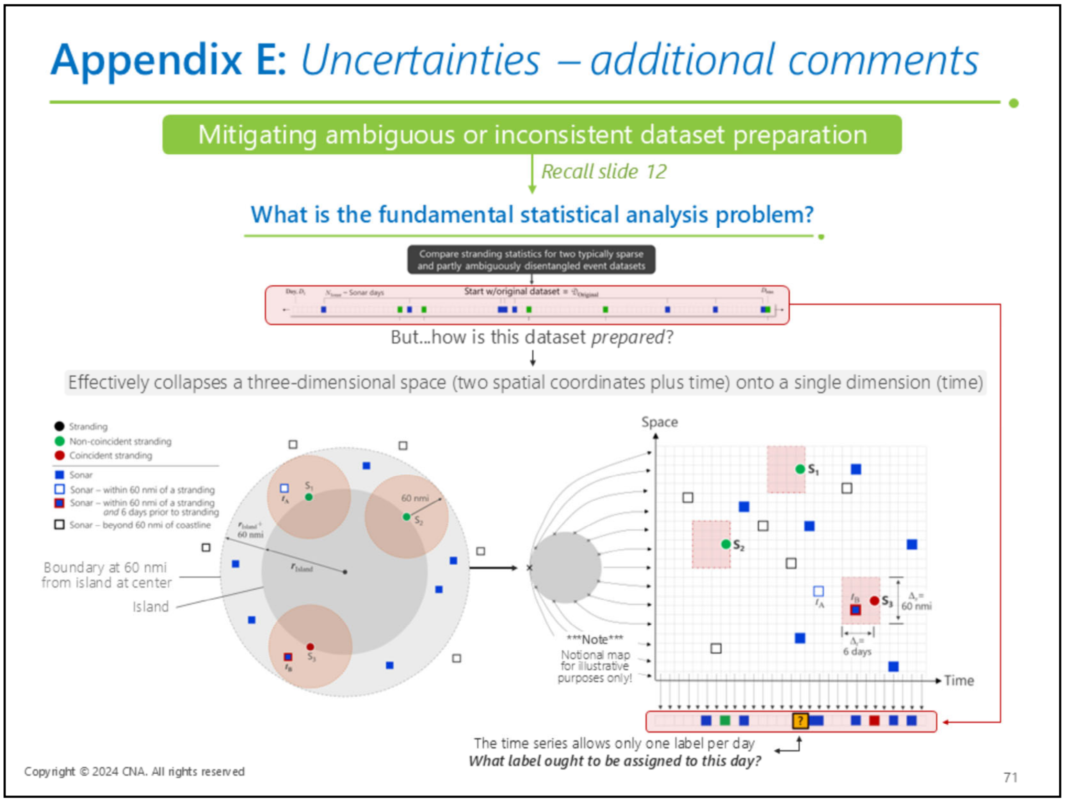
Samples=1000, Days=800, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=96, $N_{S,\text{Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob$[N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}]$	Prob$[N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}]$
	6	3.	1.	0.865
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.0458152	0.865	0.568381	δ_{Poisson}=0.53615, δ_{Bayes}=0.903145
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.0637817	0.553	0.752937	0.211
Fisher	0.0633373	0.553	0.771199	0.211

Total days = 1200 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non-Coincident Strandings = 1, Coincident Strandings = 7

Samples=1000, Days=1200, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=96, $N_{S,\text{Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob$[N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}]$	Prob$[N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}]$
	6	3.06087	1.	0.549
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.0229598	0.874	0.568599	δ_{Poisson}=0.564909, δ_{Bayes}=0.905086
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.0313011	0.874	0.725487	0.199
Fisher	0.0310893	0.874	0.741972	0.199



Appendix E: *Uncertainties – additional comments*

Mitigating ambiguous or inconsistent dataset preparation

Recall slide 12

What is the fundamental statistical analysis problem?

Compare stranding statistics for two typically sparse and partly ambiguously disentangled event datasets

Day D_1
N_{Sonar} – Sonar days
Start w/original dataset = $\mathfrak{D}_{\text{Original}}$
D_{last}

But...how is this dataset *prepared*?

Effectively collapses a three-dimensional space (two spatial coordinates plus time) onto a single dimension (time)

● Stranding
● Non-coincident stranding
● Coincident stranding
■ Sonar
□ Sonar – within 60 nmi of a stranding
■ Sonar – within 60 nmi of a stranding *and* 6 days prior to stranding
□ Sonar – beyond 60 nmi of coastline
Boundary at 60 nmi from island at center
Island
r_{Island} + 60 nmi
r_{Island}
60 nmi
S_1
S_2
S_3
t_A
t_B
Space
Time
Δ_s = 60 nmi
Δ_t = 6 days
?

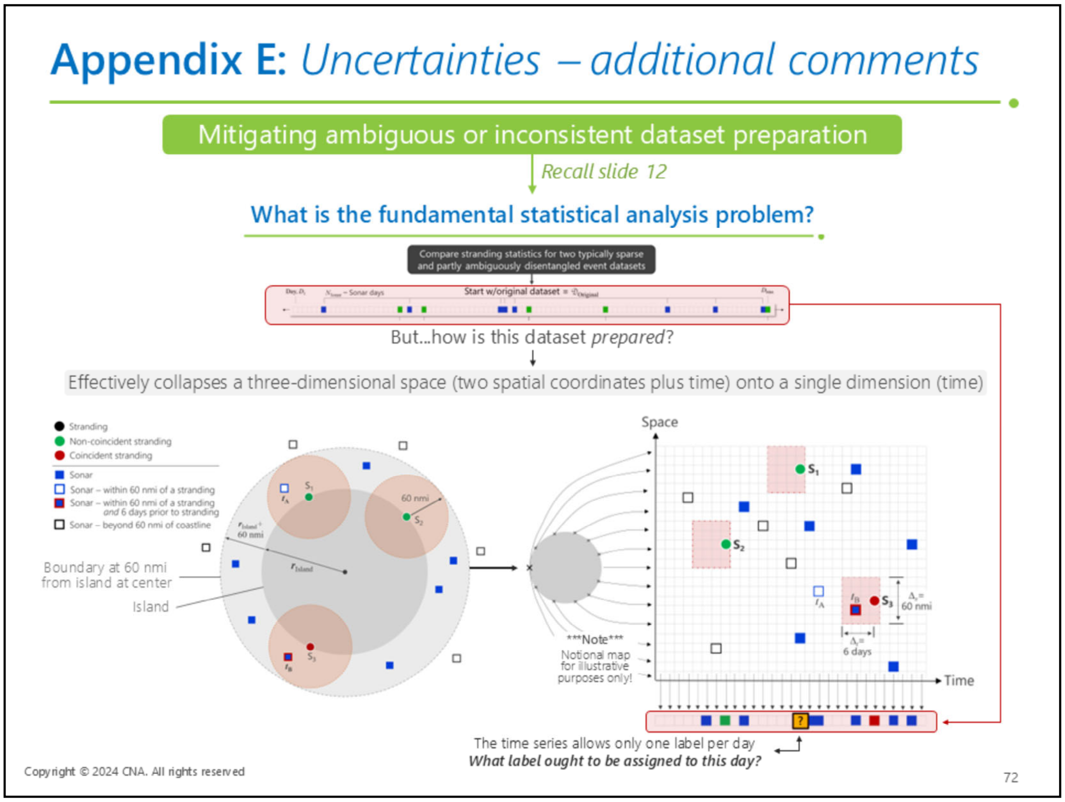
Note
Notional map for illustrative purposes only!

The time series allows only one label per day
What label ought to be assigned to this day?

Slide #27 ("*Plethora of Uncertainties*") summarizes several key uncertainties associated with sonar-stranding correlation analysis. Slide 34 expands on the second of these uncertainties—ambiguous or inconsistent criteria for including sonar in datasets—by illustrating how MCS can be used to test the degree to which a given statistical test is robust to uncertainties. For example, an analyst might ask whether a nominal decision to reject the null hypothesis (using the original dataset, $\mathfrak{D}_{\text{Original}}$) remains the statistically most credible inference to draw if a certain number of sonar days are added to $\mathfrak{D}_{\text{original}}$. These added sonar days can be used as simple proxies for any ambiguous criteria that may have been used to select which days are associated with sonar. Assuming a greater number of sonar days (than are actually observed in $\mathfrak{D}_{\text{original}}$) results in a larger number of expected coincident strandings, which in turn increases the effective P-value. If, despite this, the P-value remains below the critical threshold (say, below 0.05), the nominal rejection is said to be robust with respect to this kind of ambiguity.

However, as this appendix shows (albeit, mostly heuristically), the often ill-defined way in which active sonar is codified in datasets is a harbinger of a deeper methodological problem—specifically, that information must inevitably be *lost* during the preparation of $\mathfrak{D}_{\text{original}}$ because three-dimensional real-world data (two space dimensions plus one time dimension) are effectively pigeonholed into a one-dimensional time-series.

Observe that all stranding analyses, to date, consist of applying one or more statistical tests to a one-dimensional dataset. Each element (that is, each day) in this dataset codifies a single bit of three types of information: the element is empty, it is highlighted in **blue** (to denote the presence of sonar that was active for some part of that day), or it contains a stranding (that may be highlighted in **green** to indicate a null stranding or in **red** to highlight a coincident stranding).



Appendix E: *Uncertainties – additional comments*

Mitigating ambiguous or inconsistent dataset preparation

Recall slide 12

What is the fundamental statistical analysis problem?

Compare stranding statistics for two typically sparse and partly ambiguously disentangled event datasets

But...how is this dataset *prepared*?

Effectively collapses a three-dimensional space (two spatial coordinates plus time) onto a single dimension (time)

Note Notional map for illustrative purposes only!

The time series allows only one label per day

What label ought to be assigned to this day?

Copyright © 2024 CNA. All rights reserved

72

Day, D_1
N_{Sonar} – Sonar days
Start w/original dataset = $\mathcal{D}_{\text{Original}}$
D_{max}
● Stranding
● Non-coincident stranding
● Coincident stranding
■ Sonar
□ Sonar – within 60 nmi of a stranding
■ Sonar – within 60 nmi of a stranding *and* 6 days prior to stranding
□ Sonar – beyond 60 nmi of coastline
Boundary at 60 nmi from island at center
Island
r_{Island} + 60 nmi
r_{Island}
60 nmi
S_1
S_2
S_3
t_A
t_B
Space
Time
Δ_x = 60 nmi
Δ_t = 6 days
?

(Continued …)

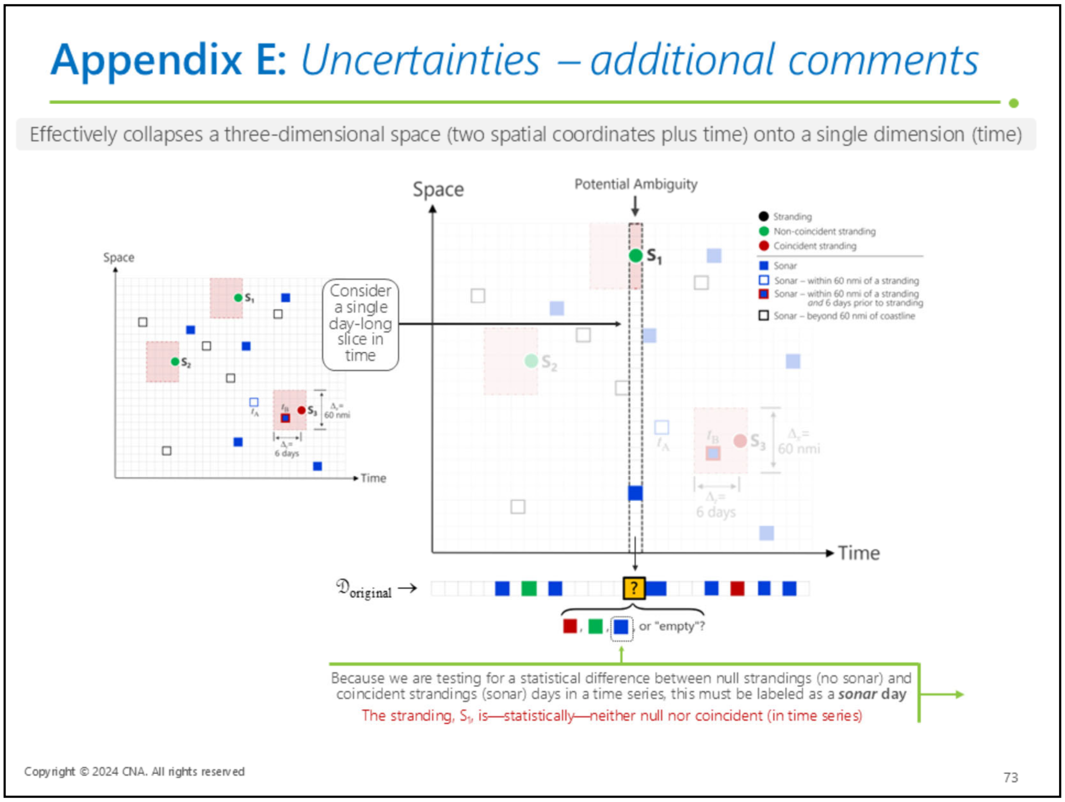
The graphic at the bottom left of the slide identifies the salient real-world elements that eventually find their way into the dataset. It shows a top-level view of a notional two-dimensional scenario centered on a circular island (highlighted in **dark gray**) with a radius = r_{Island}. *Note that the all-important time dimension is implicit in this graphic* (as discussed below). Three stranding events are depicted at points S_1, S_2, and S_3. The circles of radius = $60\ \text{nmi}$ centered on each of these strandings depict the respective areas within which a stranding may be determined to be coincident with sonar if sonar was active there during a δ_x-day window of time (typically equal to six days) preceding a given stranding. The circle centered on the island and highlighted in **light gray** denotes the maximum area such that if sonar is active outside of it, we would assume that it cannot possibly be correlated with any strandings (subject to the 60 nmi and δ_x-day window constraints).

Now, somehow the information contained in this "two space dimensions plus one time dimension" scenario must be pigeonholed into a one-dimensional time-series. We use the term *pigeonholed* deliberately. The pigeonhole principle asserts that if N objects are distributed throughout $N + x$ boxes, where $x > 0$, then there will be at least one box that contains at least two objects.[1]

The graphic at the bottom right of the slide illustrates schematically one way that three dimensions can be pigeonholed into a one-dimensional dataset—specifically, by "unwrapping" or unrolling the island's boundary onto a single spatial dimension (along the y-axis) and using the x-axis to depict the temporal dimension.

Most of the elements can be resolved without ambiguity. But what label ought to be assigned to the day marked with a "**?**" and highlighted in **orange**?

[1] B. Rittaud and A. Heefer, "The pigeonhole principle, two centuries before Dirichlet," in *The Best Writing on Mathematics 2015*, edited by Mircea Pitici, Princeton University Press, 2016



Appendix E: *Uncertainties – additional comments*

Effectively collapses a three-dimensional space (two spatial coordinates plus time) onto a single dimension (time)

Space
Time
S₁
S₂
S₃
t_A
t_B
Δ_s= 60 nmi
Δ_t= 6 days
Consider a single day-long slice in time
Space
Potential Ambiguity
S₁
S₂
S₃
t_A
t_B
Δ_s= 60 nmi
Δ_t= 6 days
Time
● Stranding
● Non-coincident stranding
● Coincident stranding
■ Sonar
□ Sonar – within 60 nmi of a stranding
■ Sonar – within 60 nmi of a stranding *and* 6 days prior to stranding
□ Sonar – beyond 60 nmi of coastline
$\mathfrak{D}_{\text{original}}$ →
?
■, ■, ■, or "empty"?

Because we are testing for a statistical difference between null strandings (no sonar) and coincident strandings (sonar) days in a time series, this must be labeled as a ***sonar* day**

The stranding, S₁, is—statistically—neither null nor coincident (in time series)

The graphic on the right side of this slide zooms in on the space-time plot to the left to highlight the column that harbors a potential (pigeonholed) ambiguity.

Each day in the dataset, $\mathfrak{D}_{\text{Original}}$, can accommodate/highlight a single piece of information, but two dynamic elements are at play: (1) the stranding, S_1 (highlighted in **green**), and (2) sonar, highlighted in **blue**).

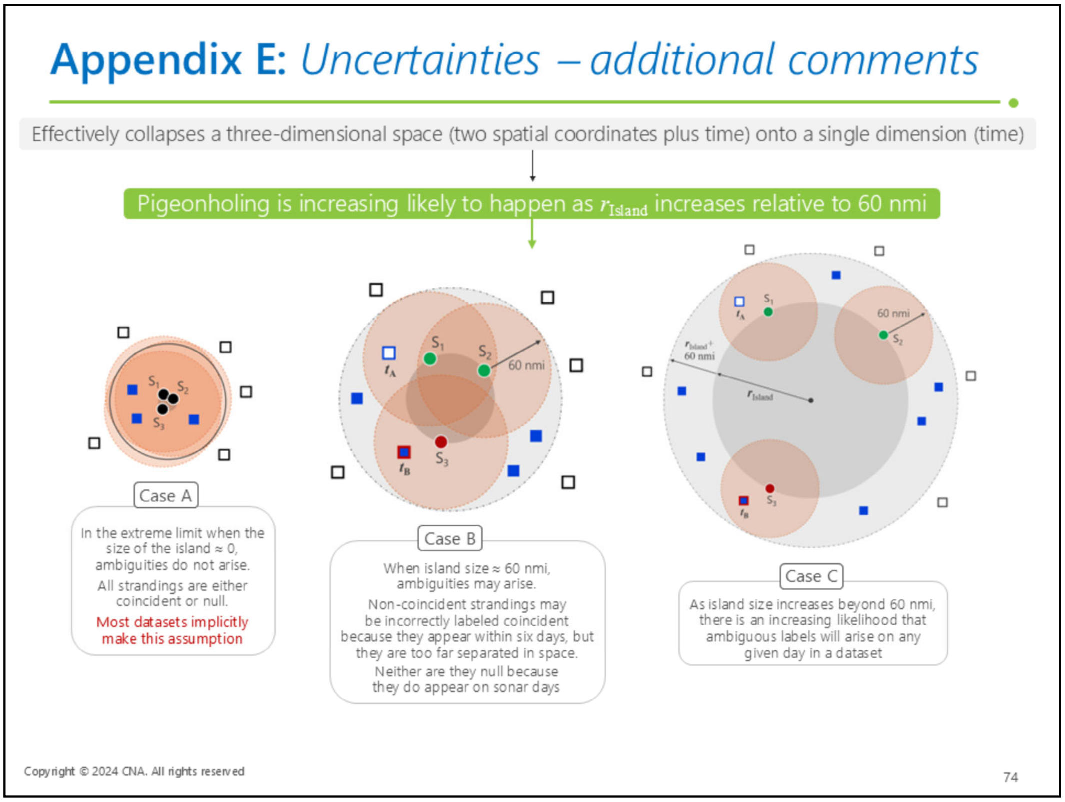
Which of these two elements are in $\mathfrak{D}_{\text{Original}}$?

This question can be asked from two points of view. The first point of view is from the perspective of an existing dataset. That is, suppose we are given $\mathfrak{D}_{\text{Original}}$, with the day corresponding to "**?**" labeled either as "stranding" (i.e., green box) or "sonar" (i.e., blue box)—it does not matter which. The point is that it must be a single label, not two. Absent the full space-time scenario (as depicted on the bottom left of the previous slide) and given only $\mathfrak{D}_{\text{Original}}$, either label necessarily hides the existence of the other, which in turn biases the associated statistical analysis.

The second point of view is from the perspective of *creating* the dataset. As we have just argued, there is room enough for a single label. So that, just as before, assigning a "stranding" effectively hides "sonar," and vice versa.

Even when the data describing the full (two-dimensional space plus time) scenario is available and (at least in principle) allows for proper distinctions to be drawn between null and coincident strandings, most stranding analysis nonetheless consists of applying a regimen of statistical tests to time-series data, which unavoidably biases the statistics because it pigeonholes relevant information.

The situation is actually even more complicated than this. For example, an additional pigeonhole-like bias may be introduced by weighing the contribution of days during which there is a single active sonar equally to days during which there were multiple active sonars. We leave the analysis of such hypothetical scenarios to a future study.



Appendix E: *Uncertainties – additional comments*

Effectively collapses a three-dimensional space (two spatial coordinates plus time) onto a single dimension (time)

Pigeonholing is increasing likely to happen as r_{Island} increases relative to 60 nmi

Case A

In the extreme limit when the size of the island ≈ 0, ambiguities do not arise.

All strandings are either coincident or null.

Most datasets implicitly make this assumption

Case B

When island size ≈ 60 nmi, ambiguities may arise.

Non-coincident strandings may be incorrectly labeled coincident because they appear within six days, but they are too far separated in space.

Neither are they null because they do appear on sonar days

Case C

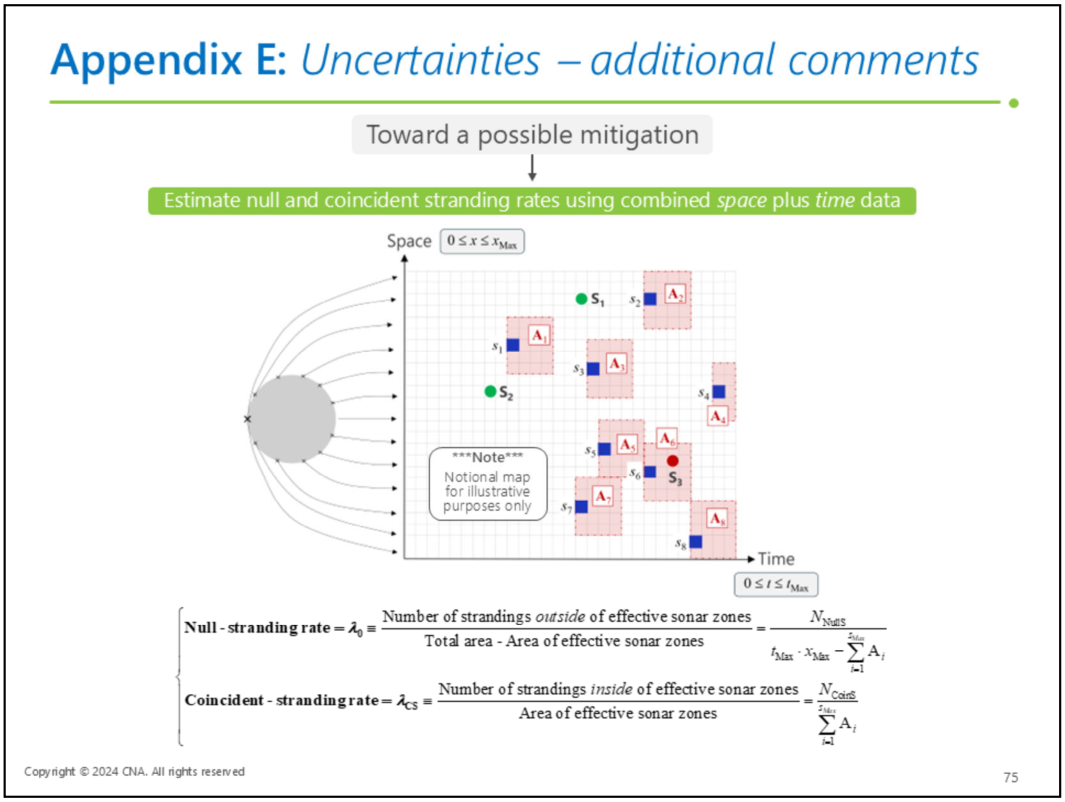
As island size increases beyond 60 nmi, there is an increasing likelihood that ambiguous labels will arise on any given day in a dataset

74

Pigeonholing is increasingly likely to happen as the size of the island, r_{Island}, increases relative to the maximum distance a sonar is allowed to be from a stranding for the stranding to be considered coincident. Nominally, this distance is set to 60 nmi.

Most stranding analysis implicitly assumes that $r_{\text{Island}} \approx 0$ because only in this case are strandings unambiguous: either *a* was coincident with sonar that was active within δ_x = six days and 60 nmi of the stranding, or it was not (i.e., it was a null stranding). See Case A.

As r_{Island} increases to 60 nmi (Case B) and larger (Case C), the probability also increases that some sonar will have been active within δ_x = 6 days of a stranding X but within the 60 nmi perimeter of another stranding Y, or vice versa. In this way, labeling ambiguities are increasingly likely to arise on any given day in the dataset.



Appendix E: *Uncertainties – additional comments*

Toward a possible mitigation

Estimate null and coincident stranding rates using combined *space* plus *time* data

Note
Notional map for illustrative purposes only

$$\begin{cases} \textbf{Null-stranding rate} = \lambda_0 \equiv \dfrac{\text{Number of strandings } \textit{outside} \text{ of effective sonar zones}}{\text{Total area - Area of effective sonar zones}} = \dfrac{N_{\text{NullS}}}{t_{\text{Max}} \cdot x_{\text{Max}} - \sum_{i=1}^{s_{\text{Max}}} \mathrm{A}_i} \\ \textbf{Coincident-stranding rate} = \lambda_{CS} \equiv \dfrac{\text{Number of strandings } \textit{inside} \text{ of effective sonar zones}}{\text{Area of effective sonar zones}} = \dfrac{N_{\text{CoinS}}}{\sum_{i=1}^{s_{\text{Max}}} \mathrm{A}_i} \end{cases}$$

75

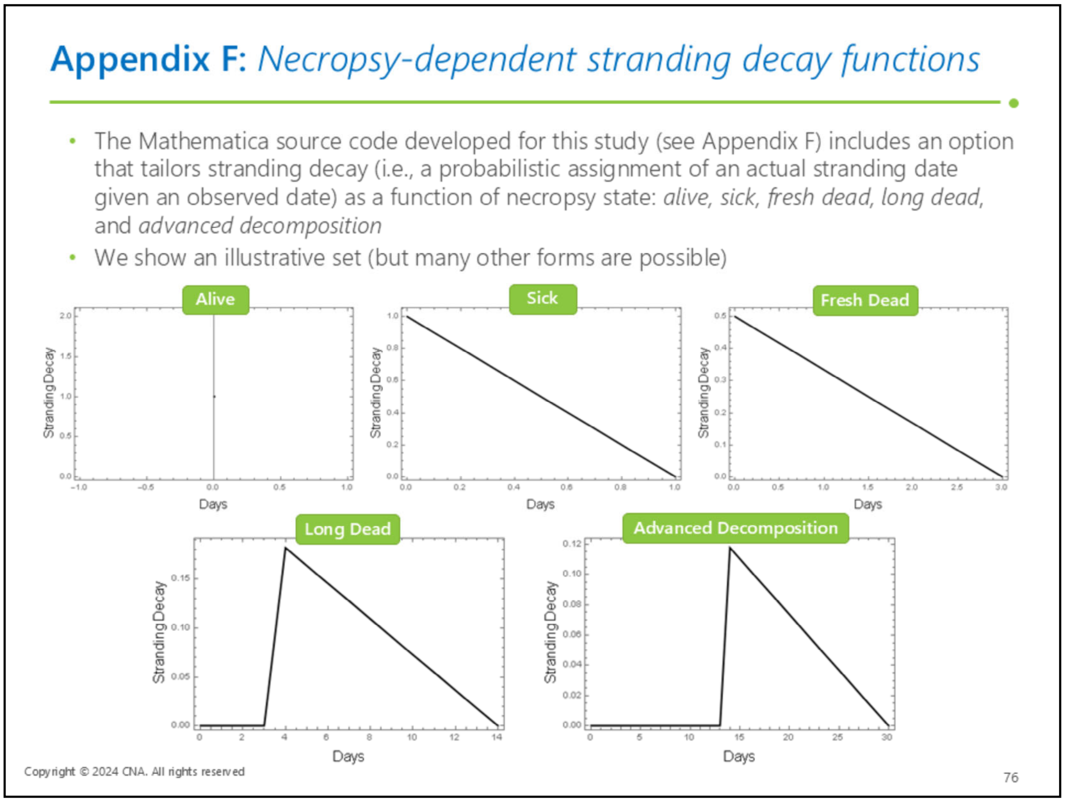
The final slide in this appendix illustrates one way in which null stranding rates may be estimated not from time-series data (as is traditionally done, and, as has just been argued, is rife with potential bias-generating ambiguities) but by using the combined space plus time data describing the full scenario.

The idea is to generalize how the null stranding rate, λ_0, is estimated. Conventionally, λ_0 is estimated by counting the number of strandings that occur on days without sonar and dividing by the number of such null sonar days (see slide 11). In like fashion, we can use the two-dimensional space-time representation of the same scenario to do the same thing in three steps: (1) cordon off all areas within which strandings are, by definition, coincident with sonar (these appear as dotted red areas in the figure on the slide); (2) count the number of strandings that are not in any of those areas (these are the "null strandings"); and (3) estimate λ_0 as the number of null strandings divided by the total space-time area minus the (red highlighted) sonar zones.

Similarly, whereas the *expected* number of coincident strandings is traditionally estimated by multiplying λ_0 by the number of effective sonar days, we multiply the generalized λ_0 by the total effective sonar area (i.e., by summing over the component areas).

By preserving the information and geometry relevant to a given scenario, this approach eliminates the ambiguities that might otherwise be latent in a traditional time-series.

[1] The reader is cautioned that this approach vastly oversimplifies the requisite analysis, which involves mapping the real-world geometry that describes an operating area of interest. Also left out of the discussion is how to calibrate the "size" of the space-time blocks used to define areas. Details await a future study.



Appendix F: *Necropsy-dependent stranding decay functions*

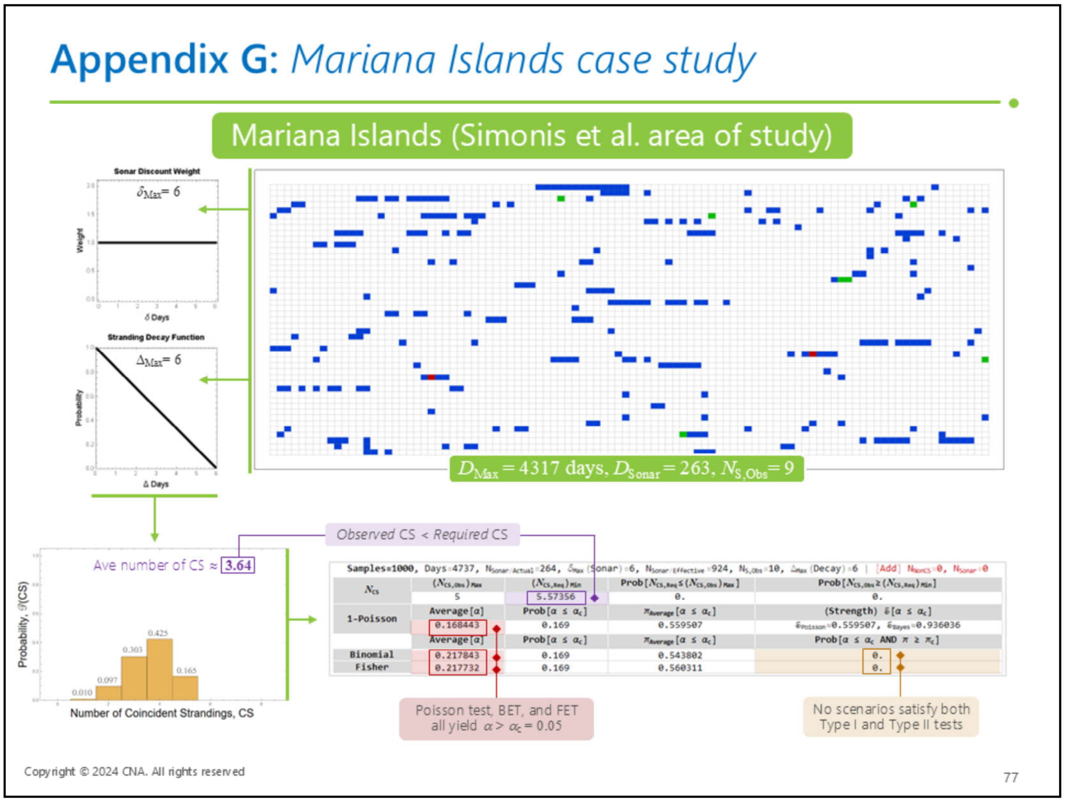
- The Mathematica source code developed for this study (see Appendix F) includes an option that tailors stranding decay (i.e., a probabilistic assignment of an actual stranding date given an observed date) as a function of necropsy state: *alive, sick, fresh dead, long dead,* and *advanced decomposition*
- We show an illustrative set (but many other forms are possible)

76

The Mathematica source code developed for this study includes an option to apply user-defined decay-state-dependent stranding decay functions when necropsy data are (typically only rarely) available. Case study four, which summarizes the analysis for the SOCAL dataset (described in [16] and summarized on slide 38 of the main narrative), is given in Appendix G, which immediately follows.

We caution readers that even when necropsy data are available, the classification and implication of a given decay state are highly subjective. About the only definitive conclusion that may be drawn is that the actual stranding dates for animals that have been long dead by the time an observation is made are likely much earlier, relatively speaking, than if an animal is obviously still alive and uninjured.

Beyond making such a broadly sweeping intuitive observation, little else is certain. The functional forms shown on this slide are emphatically not panacea depictions of what the true decay-stranding function looks like (analysts can easily modify the appearance of these functions); indeed, *there is no definitive function*. The best we can do is to gauge how strongly any uncertainty in actual versus observed stranding dates would influence stranding analysis based on specific dates. The weight of statistical evidence to reject the null hypothesis would be strengthened only if it remains robust with respect to this or any other kind of uncertainty (however crudely these uncertainties are taken into account).



Appendix G: *Mariana Islands case study*

Mariana Islands (Simonis et al. area of study)

Observed CS < Required CS

Samples=1000, Days=4737, $N_{Sonar/Actual}$=264, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=924, $N_{S,Obs}$=10, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob[$N_{CS,Req} \le (N_{CS,Obs})_{Max}$]	Prob[$N_{CS,Obs} \ge (N_{CS,Req})_{Min}$]
	5	5.57356	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{Average}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.168443	0.169	0.559507	$\delta_{Poisson}$=0.559507, δ_{Bayes}=0.936036
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{Average}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.217843	0.169	0.543802	0.
Fisher	0.217732	0.169	0.560311	0.

Poisson test, BET, and FET all yield $\alpha > \alpha_c = 0.05$

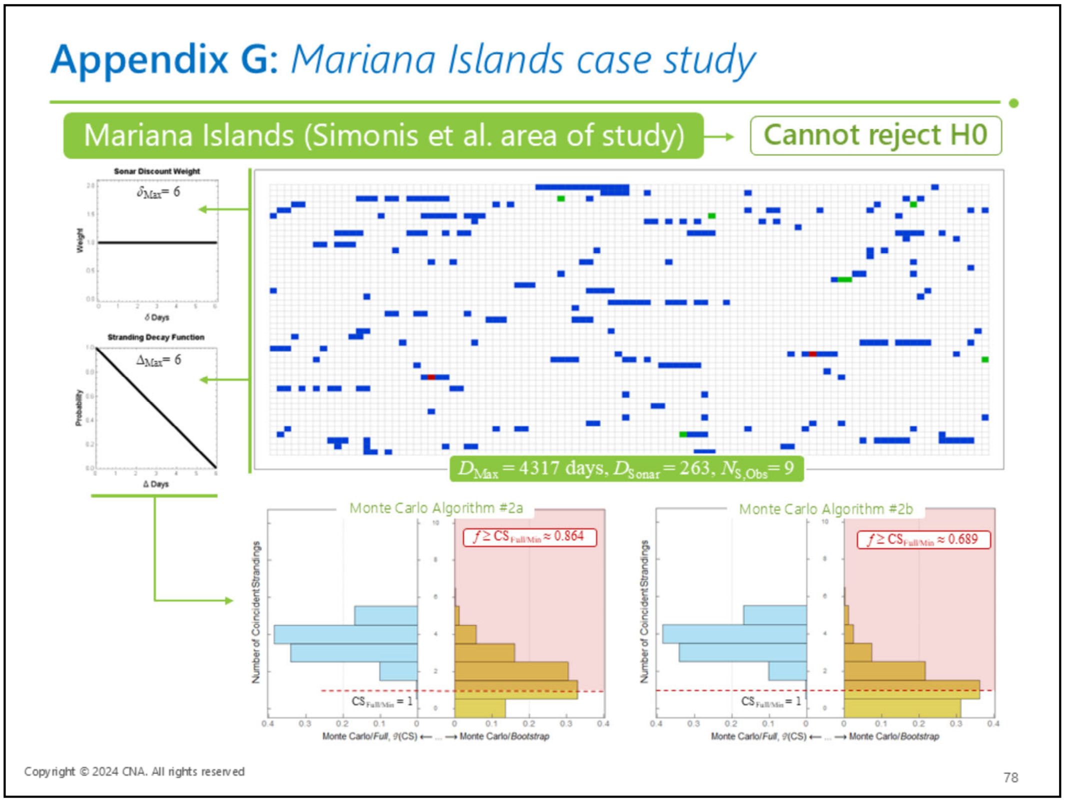
No scenarios satisfy both Type I and Type II tests

77

Appendix G contains two additional case studies using the datasets described on slide 40.

The Mariana dataset is an update to the data that was used in the study described in [27], which showed a significantly higher (at the 0.95 level) stranding rate during sonar periods compared to non-sonar periods. We updated the data from [27] with one additional stranding, and we used the Navy's SPORTS database to compile much more complete information on military sonar use.

The output includes the same items that were previously used to summarize the analysis for case studies one and two (slides 42 to 46); also see **Appendix C**.

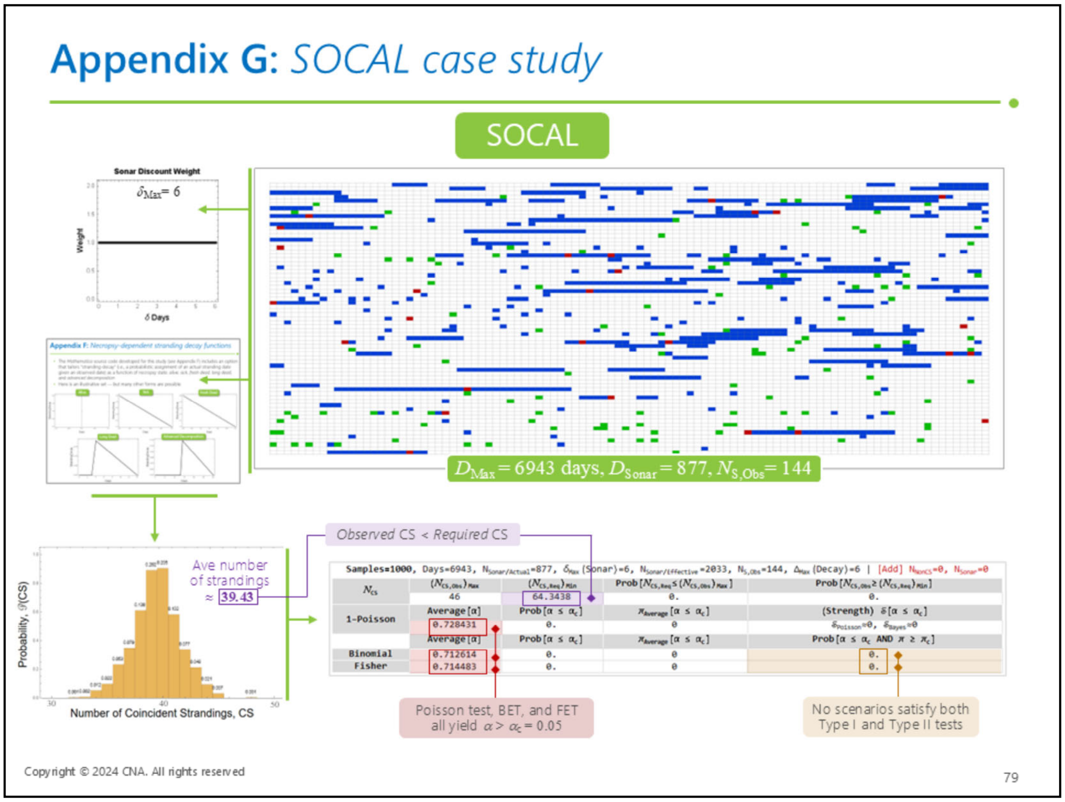


Appendix G: *Mariana Islands case study*

Mariana Islands (Simonis et al. area of study) → Cannot reject H0

$D_{Max} = 4317$ days, $D_{Sonar} = 263$, $N_{S,Obs} = 9$

78



Appendix G: *SOCAL case study*

SOCAL

Sonar Discount Weight
$\delta_{Max} = 6$
Weight
δ Days
Appendix F: Necropsy-dependent stranding decay functions
Probability, $\mathcal{P}$(CS)
Number of Coincident Strandings, CS
Ave number of strandings ≈ 39.43

$D_{Max} = 6943$ days, $D_{Sonar} = 877$, $N_{S,Obs} = 144$

Observed CS < Required CS

Samples=1000, Days=6943, $N_{Sonar/Actual}$=877, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=2033, $N_{S,Obs}$=144, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

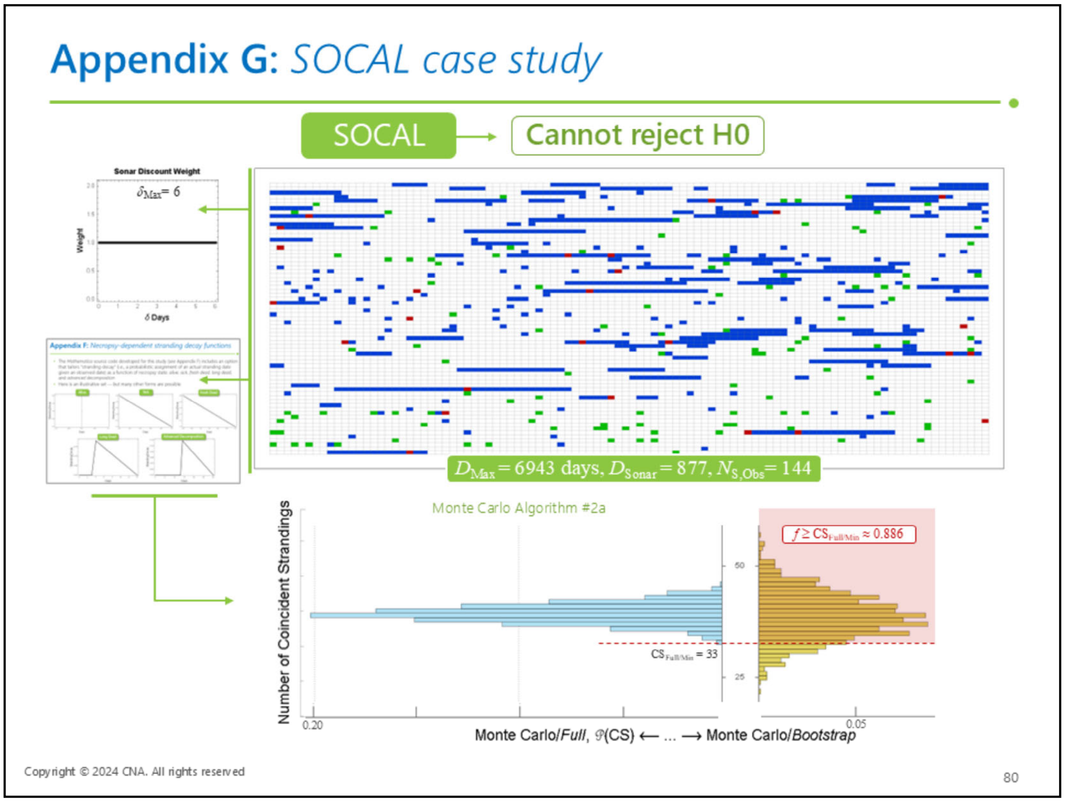
N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob$[N_{CS,Req} \le (N_{CS,Obs})_{Max}]$	Prob$[N_{CS,Obs} \ge (N_{CS,Req})_{Min}]$
	46	64.3438	0.	0.
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	(Strength) δ[α ≤ α_c]
	0.728431	0.	0	$\delta_{Poisson}$=0, δ_{Bayes}=0
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.712614	0.	0	0.
Fisher	0.714483	0.	0	0.

Poisson test, BET, and FET all yield $\alpha > \alpha_c = 0.05$

No scenarios satisfy both Type I and Type II tests

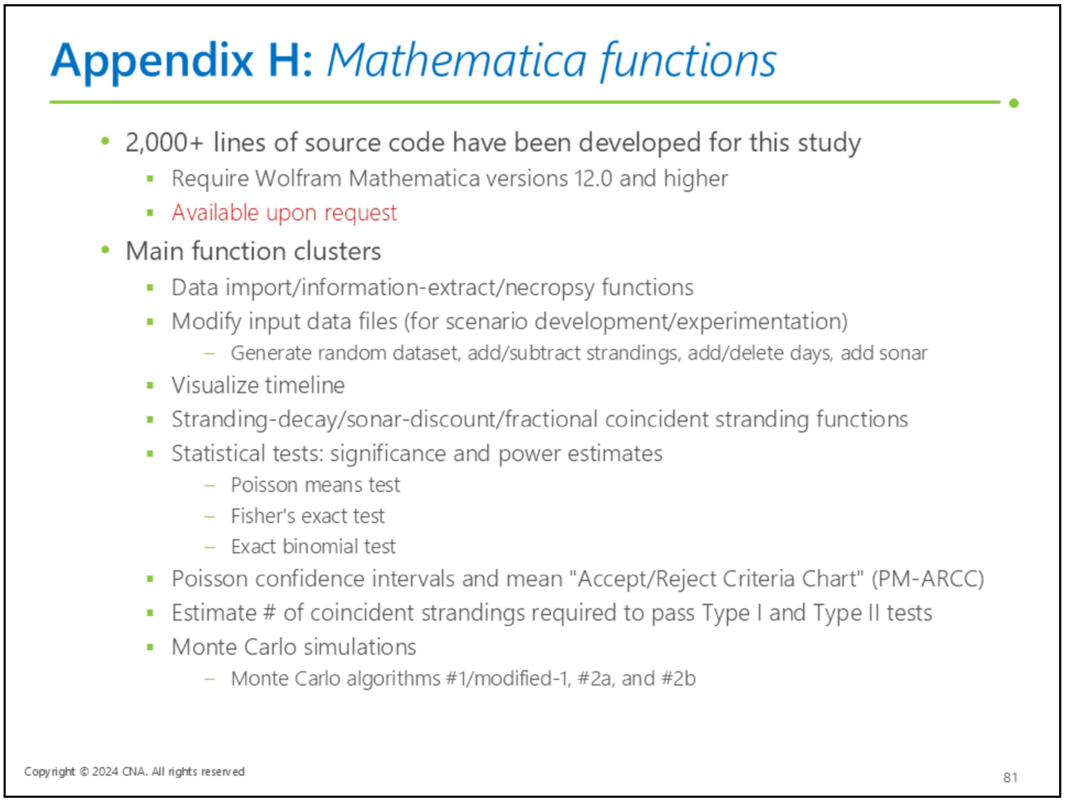
The SOCAL dataset was used in the study described in [16] and summarized on slide 38 of the main narrative. This particular dataset includes necropsy information. Specifically, each stranding is accompanying by one of four labels: (1) alive or sick/injured, (2) fresh dead, (3) long dead, or (4) advanced decomposition.

The stranding correlation results for this last case study are based on the stranding decay functions that appear in **Appendix F**.



Appendix G: *SOCAL case study*

SOCAL → Cannot reject H0



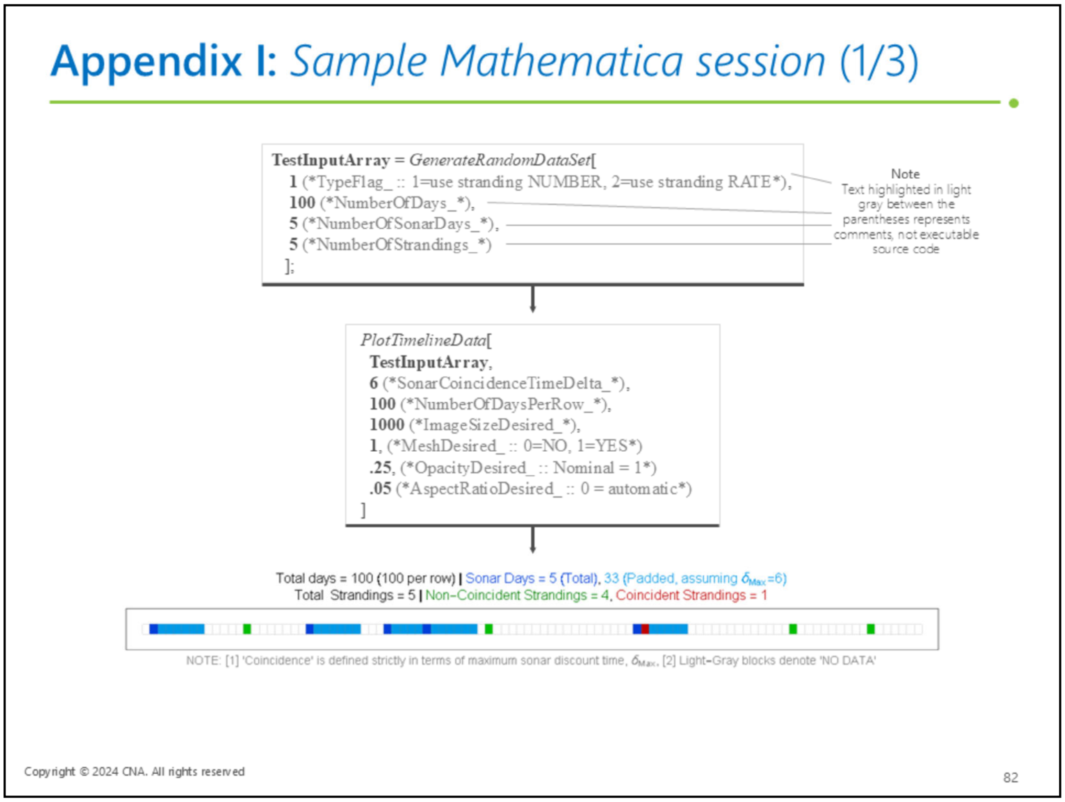
Appendix H: *Mathematica functions*

- 2,000+ lines of source code have been developed for this study
  - Require Wolfram Mathematica versions 12.0 and higher
  - Available upon request
- Main function clusters
  - Data import/information-extract/necropsy functions
  - Modify input data files (for scenario development/experimentation)
    - Generate random dataset, add/subtract strandings, add/delete days, add sonar
  - Visualize timeline
  - Stranding-decay/sonar-discount/fractional coincident stranding functions
  - Statistical tests: significance and power estimates
    - Poisson means test
    - Fisher's exact test
    - Exact binomial test
  - Poisson confidence intervals and mean "Accept/Reject Criteria Chart" (PM-ARCC)
  - Estimate # of coincident strandings required to pass Type I and Type II tests
  - Monte Carlo simulations
    - Monte Carlo algorithms #1/modified-1, #2a, and #2b

More than 2,000 lines of Mathematica source code[1] have been developed for this study. Individual functions include basic data import and timeline visualization, input data modification for experimentation and scenario development, the stranding decay and sonar discount functions, various statistical tests (including significance and power), and stand-alone MCSs.

The complete source code is available upon request.

[1] *Mathematica* homepage: https://www.wolfram.com/mathematica/.



Appendix I: *Sample Mathematica session* (1/3)

```
TestInputArray = GenerateRandomDataSet[
  1 (*TypeFlag_ :: 1=use stranding NUMBER, 2=use stranding RATE*),
  100 (*NumberOfDays_*),
  5 (*NumberOfSonarDays_*),
  5 (*NumberOfStrandings_*)
 ];
```

Note
Text highlighted in light gray between the parentheses represents comments, not executable source code

```
PlotTimelineData[
  TestInputArray,
  6 (*SonarCoincidenceTimeDelta_*),
  100 (*NumberOfDaysPerRow_*),
  1000 (*ImageSizeDesired_*),
  1, (*MeshDesired_ :: 0=NO, 1=YES*)
  .25, (*OpacityDesired_ :: Nominal = 1*)
  .05 (*AspectRatioDesired_ :: 0 = automatic*)
]
```

Total days = 100 (100 per row) | Sonar Days = 5 (Total), 33 (Padded, assuming δ_{Max}=6)
Total Strandings = 5 | Non-Coincident Strandings = 4, Coincident Strandings = 1

NOTE: [1] 'Coincidence' is defined strictly in terms of maximum sonar discount time, δ_{Max}, [2] Light-Gray blocks denote 'NO DATA'

Appendix I contains illustrative (albeit small and heavily truncated) input/output samples of a few of the 50+ Mathematica functions developed for this study (as described in **Appendix H**).

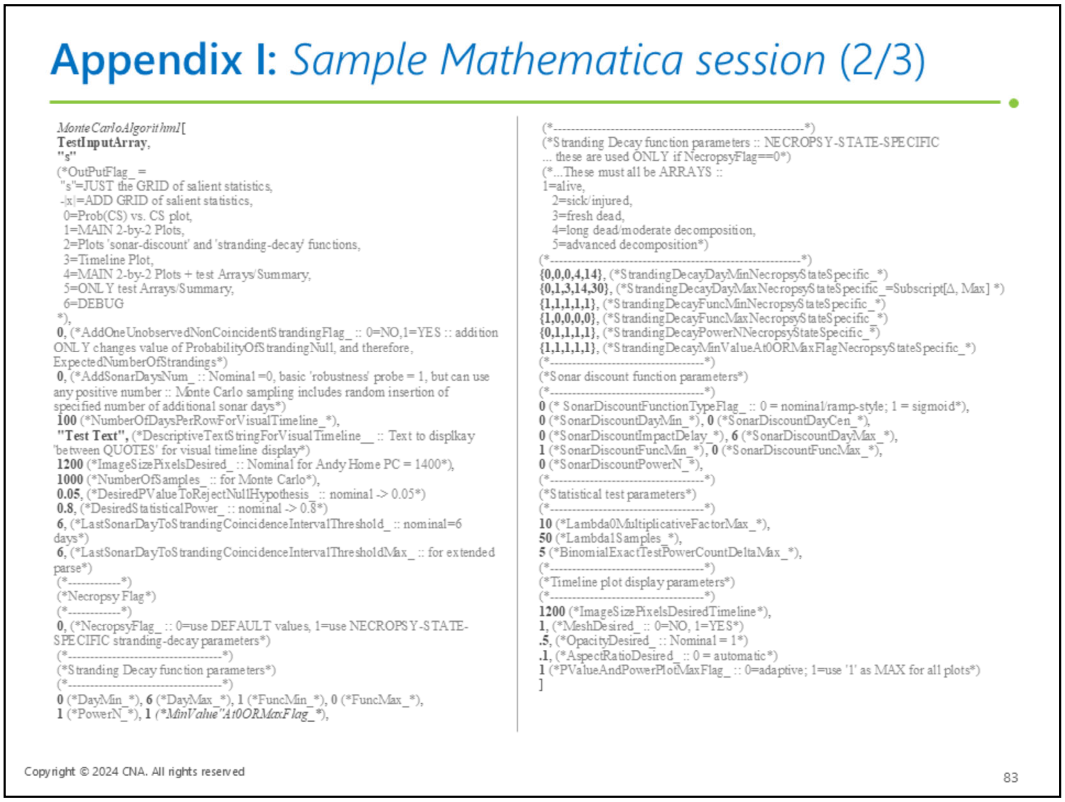
This slide shows how Mathematica can be used to easily generate notional datasets of arbitrary size for experimentation (such as those that appear on slides 30 to 37 and in **Appendix D**). The function ***GenerateRandomDataSet*[…]** takes the following as input:

1. *TypeFlag*, which instructs the function to generate a dataset either by assuming a fixed number of strandings (*TypeFlag*=1) or by assuming a stranding rate (*TypeFlag*=2).
2. *NumberOfDay*s, which specifies the total number of days to include in the notional dataset.
3. *NumberOfSonarDay*s, which specifies the total number of days that include active sonar.
4. *NumberOfStrandings*, which specifies the total number of strandings.

The dataset is randomly generated and saved as a variable array called *TestInputArray*.

The second function, ***PlotTimelineData*[…]**, generates a graphic view of the data in *TestInputArray***.** The timeline's appearance may be defined or altered by specifying the values of the following six optional parameters:

1. *SonarCoincidenceTimeDelta* defines the time delay δ_x (as defined on slide 28) between the stranding and the last sonar day.
2. *NumberOfDaysPerRow* specifies how many days will be rendered per row (e.g., the user may specify 365 days for datasets that contains several years' worth of information).
3. *ImageSizeDesired* defines how many total pixels will be used to render the image.
4. *MeshDesired* instructs the function whether to display a thin, light gray "mesh" to seperate the days ('0' = no, and '1' = yes).
5. *OpacityDesired* $\in$ [0,1] defines the opacity of the mesh (the closer the value is to '1', the more visible will be the mesh).
6. *AspectRatioDesired* specifies the desired aspect ratio (though a value of '0' yields a default value, which may be suboptimal, depending on the size of the dataset).



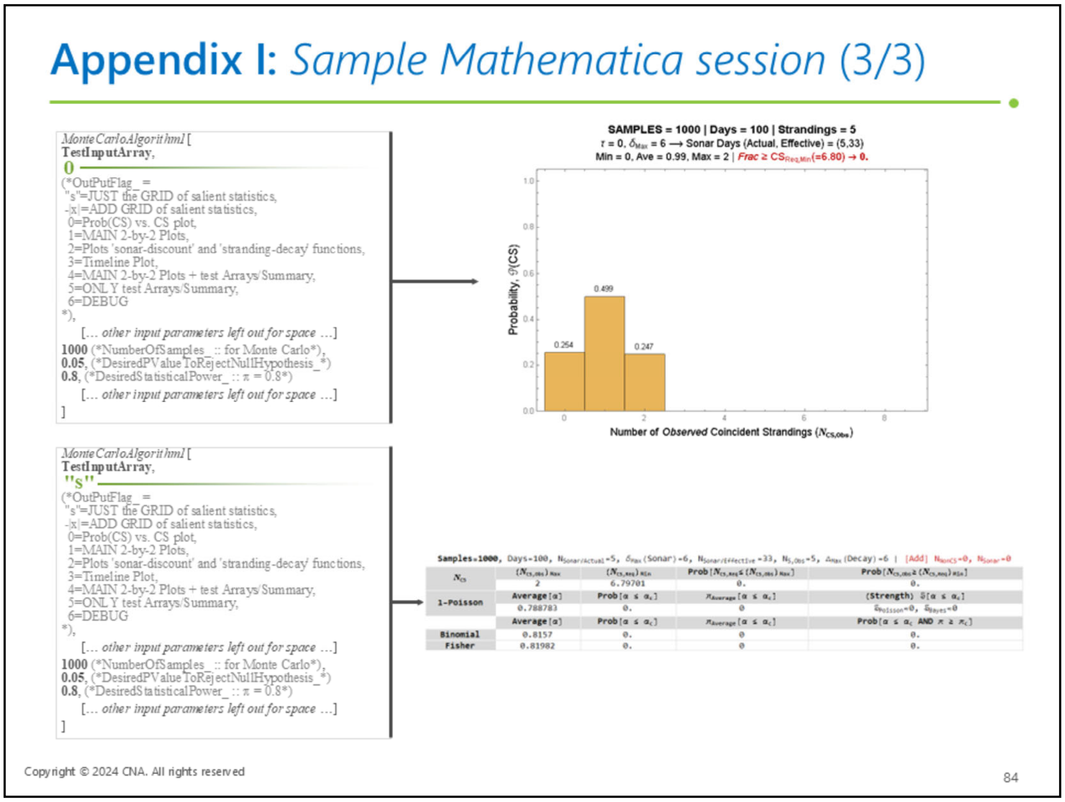
Appendix I: *Sample Mathematica session (2/3)*

```
MonteCarloAlgorithm1[
TestInputArray,
"s"
(*OutPutFlag_ =
 "s"=JUST the GRID of salient statistics,
 -|x|=ADD GRID of salient statistics,
  0=Prob(CS) vs. CS plot,
  1=MAIN 2-by-2 Plots,
  2=Plots 'sonar-discount' and 'stranding-decay' functions,
  3=Timeline Plot,
  4=MAIN 2-by-2 Plots + test Arrays/Summary,
  5=ONLY test Arrays/Summary,
  6=DEBUG
*),
 0, (*AddOneUnobservedNonCoincidentStrandingFlag_ :: 0=NO,1=YES :: addition
ONLY changes value of ProbabilityOfStrandingNull, and therefore,
ExpectedNumberOfStrandings*)
 0, (*AddSonarDaysNum_ :: Nominal =0, basic 'robustness' probe = 1, but can use
any positive number :: Monte Carlo sampling includes random insertion of
specified number of additional sonar days*)
 100 (*NumberOfDaysPerRowForVisualTimeline_*),
 "Test Text", (*DescriptiveTextStringForVisualTimeline__ :: Text to displkay
'between QUOTES' for visual timeline display*)
 1200 (*ImageSizePixelsDesired_ :: Nominal for Andy Home PC = 1400*),
 1000 (*NumberOfSamples_ :: for Monte Carlo*),
 0.05, (*DesiredPValueToRejectNullHypothesis_ :: nominal -> 0.05*)
 0.8, (*DesiredStatisticalPower_ :: nominal -> 0.8*)
 6, (*LastSonarDayToStrandingCoincidenceIntervalThreshold_ :: nominal=6
days*)
 6, (*LastSonarDayToStrandingCoincidenceIntervalThresholdMax_ :: for extended
parse*)
 (*-----------*)
 (*Necropsy Flag*)
 (*-----------*)
 0, (*NecropsyFlag_ :: 0=use DEFAULT values, 1=use NECROPSY-STATE-
SPECIFIC stranding-decay parameters*)
 (*--------------------------------*)
 (*Stranding Decay function parameters*)
 (*--------------------------------*)
 0 (*DayMin_*), 6 (*DayMax_*), 1 (*FuncMin_*), 0 (*FuncMax_*),
 1 (*PowerN_*), 1 (*MinValue"At0ORMaxFlag_*),
 (*--------------------------------------------------------*)
 (*Stranding Decay function parameters :: NECROPSY-STATE-SPECIFIC
 ... these are used ONLY if NecropsyFlag==0*)
 (*...These must all be ARRAYS ::
 1=alive,
   2=sick/injured,
   3=fresh dead,
   4=long dead/moderate decomposition,
   5=advanced decomposition*)
(*--------------------------------------------------------*)
{0,0,0,4,14}, (*StrandingDecayDayMinNecropsyStateSpecific_*)
{0,1,3,14,30}, (*StrandingDecayDayMaxNecropsyStateSpecific_=Subscript[Δ, Max] *)
{1,1,1,1,1}, (*StrandingDecayFuncMinNecropsyStateSpecific_*)
{1,0,0,0,0}, (*StrandingDecayFuncMaxNecropsyStateSpecific_*)
{0,1,1,1,1}, (*StrandingDecayPowerNNecropsyStateSpecific_*)
{1,1,1,1,1}, (*StrandingDecayMinValueAt0ORMaxFlagNecropsyStateSpecific_*)
(*--------------------------------*)
(*Sonar discount function parameters*)
(*--------------------------------*)
0 (* SonarDiscountFunctionTypeFlag_ :: 0 = nominal/ramp-style; 1 = sigmoid*),
0 (*SonarDiscountDayMin_*), 0 (*SonarDiscountDayCen_*),
0 (*SonarDiscountImpactDelay_*), 6 (*SonarDiscountDayMax_*),
1 (*SonarDiscountFuncMin_*), 0 (*SonarDiscountFuncMax_*),
0 (*SonarDiscountPowerN_*),
(*--------------------------------*)
(*Statistical test parameters*)
(*--------------------------------*)
10 (*Lambda0MultiplicativeFactorMax_*),
50 (*Lambda1Samples_*),
5 (*BinomialExactTestPowerCountDeltaMax_*),
(*--------------------------------*)
(*Timeline plot display parameters*)
(*--------------------------------*)
1200 (*ImageSizePixelsDesiredTimeline*),
1, (*MeshDesired_ :: 0=NO, 1=YES*)
.5, (*OpacityDesired_ :: Nominal = 1*)
.1, (*AspectRatioDesired_ :: 0 = automatic*)
1 (*PValueAndPowerPlotMaxFlag_ :: 0=adaptive; 1=use '1' as MAX for all plots*)
]
```

This slide shows the complete input parameter list for the function ***MonteCarloAlgorithm1***[…], the pseudocode for which is given on slide 32.

In perusing this code, note that any text between symmetric instances of "(*" and "*)" are comments. For example, the string "(*Stranding Decay function parameters*)" that appears on the bottom left of the slide merely identifies that the parameters that follow are all associated with (and used to define) the stranding decay function (see slide 30).

Other clusters of related parameters are grouped accordingly, such as *sonar discount function parameters*, *statistical test parameters*, and *timeline plot display parameters* (the latter of which replicates the parameter list used by the function ***PlotTimelineData***[…], as discussed on the previous slide).



Appendix I: *Sample Mathematica session* (3/3)

```
MonteCarloAlgorithm1[
TestInputArray,
0
(*OutPutFlag_ =
 "s"=JUST the GRID of salient statistics,
 -|x|=ADD GRID of salient statistics,
  0=Prob(CS) vs. CS plot,
  1=MAIN 2-by-2 Plots,
  2=Plots 'sonar-discount' and 'stranding-decay' functions,
  3=Timeline Plot,
  4=MAIN 2-by-2 Plots + test Arrays/Summary,
  5=ONLY test Arrays/Summary,
  6=DEBUG
*),
    [... other input parameters left out for space ...]
1000 (*NumberOfSamples_ :: for Monte Carlo*),
0.05, (*DesiredPValueToRejectNullHypothesis_*)
0.8, (*DesiredStatisticalPower_ :: π = 0.8*)
    [... other input parameters left out for space ...]
]
```

SAMPLES = 1000 | Days = 100 | Strandings = 5
$\tau = 0$, $\delta_{Max} = 6 \longrightarrow$ Sonar Days (Actual, Effective) = (5,33)
Min = 0, Ave = 0.99, Max = 2 | *Frac* ≥ $CS_{Req,Min}$(=6.80) → 0.
Probability, $\mathcal{P}$(CS)
1.0
0.8
0.6
0.4
0.2
0.0
0.254
0.499
0.247
0
2
4
6
8
Number of *Observed* Coincident Strandings ($N_{CS,Obs}$)

```
MonteCarloAlgorithm1[
TestInputArray,
"s"
(*OutPutFlag_ =
 "s"=JUST the GRID of salient statistics,
 -|x|=ADD GRID of salient statistics,
  0=Prob(CS) vs. CS plot,
  1=MAIN 2-by-2 Plots,
  2=Plots 'sonar-discount' and 'stranding-decay' functions,
  3=Timeline Plot,
  4=MAIN 2-by-2 Plots + test Arrays/Summary,
  5=ONLY test Arrays/Summary,
  6=DEBUG
*),
    [... other input parameters left out for space ...]
1000 (*NumberOfSamples_ :: for Monte Carlo*),
0.05, (*DesiredPValueToRejectNullHypothesis_*)
0.8, (*DesiredStatisticalPower_ :: π = 0.8*)
    [... other input parameters left out for space ...]
]
```

Samples=1000, Days=100, $N_{Sonar/Actual}$=5, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=33, $N_{S,Obs}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob$[N_{CS,Req} \le (N_{CS,Obs})_{Max}]$	Prob$[N_{CS,Obs} \ge (N_{CS,Req})_{Min}]$
	2	6.79701	0.	0.
1-Poisson	**Average[α]**	**Prob[α ≤ α_c]**	$\pi_{Average}[\alpha \le \alpha_c]$	**(Strength)** $\mathbb{S}[\alpha \le \alpha_c]$
	0.788783	0.	0	$\mathbb{S}_{Poisson}=0$, $\mathbb{S}_{Bayes}=0$
	Average[α]	**Prob[α ≤ α_c]**	$\pi_{Average}[\alpha \le \alpha_c]$	**Prob[α ≤ α_c AND π ≥ π_c]**
Binomial	0.8157	0.	0	0.
Fisher	0.81982	0.	0	0.

84

The function ***MonteCarloAlgorithm1***[…] provides nine optional forms of output, two of which are shown on this slide.

Setting the "(*OutPutFlag*)" parameter equal to "0" (as highlighted in green at the top left) instructs Mathematica to output a histogram of the fraction of runs that yield a given number of observed coincident strandings (as determined probabilistically using the value of other run-time parameters).

Setting the "(*OutPutFlag*)" parameter equal to "s" (as highlighted in green at the bottom left) instructs Mathematica to output the matrix of summary statistics described in **Appendix C**.

Other options (not shown on the slide) include additional histograms (such as those that appear on slide 30, which were generated by setting OutPutFlag = 1), a timeline plot (which uses the function ***PlotTimelineData***[…], as discussed earlier, and is activated by setting OutPutFlag = 3), and options to display the values of various interim variables and test arrays while the simulation is running (but which are designed more as debugging aids).

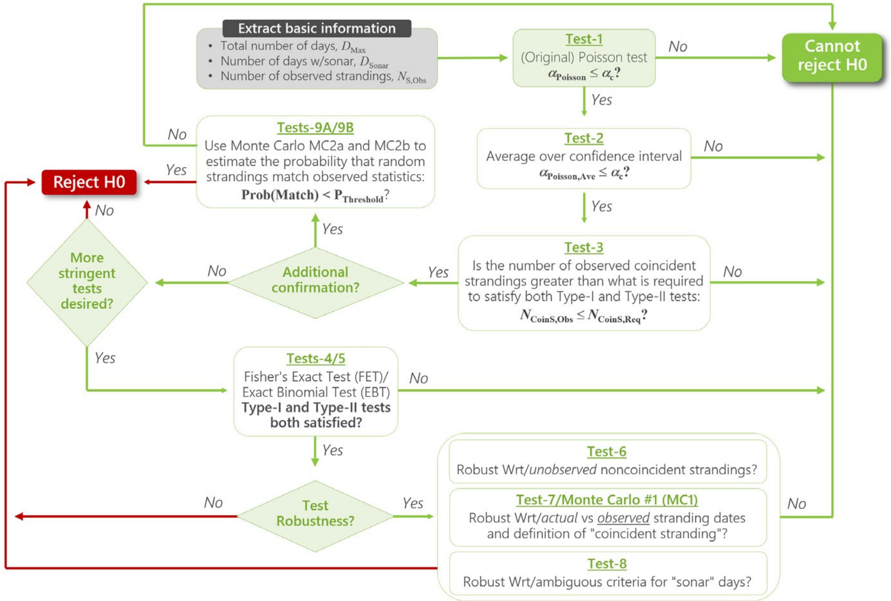
Full-resolution Slide deck

CNA

Statistical Analysis of Marine Mammal Stranding Events

Toward a Stranding Correlation Analysis Playbook

Andy Ilachinski, Ron Filadelfo



Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\mathrm{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\mathrm{Poisson}} \leq \alpha_{\mathrm{c}}$?

Cannot reject H0

Test-2
Average over confidence interval
$\alpha_{\mathrm{Poisson,Ave}} \leq \alpha_{\mathrm{c}}$?

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\mathrm{CoinS,Obs}} \leq N_{\mathrm{CoinS,Req}}$?

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\mathrm{Prob(Match)} < \mathrm{P}_{\mathrm{Threshold}}$?

Reject H0

Additional confirmation?

More stringent tests desired?

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Test Robustness?

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

Goals

- **Fundamental goal of this project**
  - To determine the likelihood that stranding events are *correlated* with—not necessarily *caused* by—the use of sonar
    - The statistical analysis we describe is based solely on ambiguously defined binary-valued "sparse event" datasets
    - Basic statistics provides only limited insight into a complex, multidimensional problem
- **More practical goal of this project**
  - To develop a step-by-step analysis framework that the Navy can use to quickly determine whether the statistical evidence is sufficient and supports the assertion that a series of stranding events are correlated with sonar

Problem with the current approach

- **Past analyses have typically inferred a correlation (or lack thereof) based on a *single* statistical means test—the null stranding rate**
  - This rate is determined by dividing the number of *observed* stranding events that took place on days when sonar was not active by the number of such days
  - It is used to estimate the number of *expected* stranding events coincident with sonar and compared to the *actual* number recorded for days with active sonar
  - If the actual number of coincident strandings greatly exceeds the expected number, a correlation between strandings and sonar is inferred to exist
- **Although statistically valid, this approach is significantly limited in two key respects**
  - The *observed* null stranding rate is a proxy for the *unknown true mean* of a random process
  - The inference is based solely on whether the data passes a so-called *significance* test; however, inferences cannot be credibly drawn if the statistical *power* of the test is too small

Results of this study

- **Refined existing analysis methodology**
  - E.g., added estimates of Poisson confidence intervals and statistical power
- **Developed additional statistical tests to strengthen the veracity of inferences**
  - More stringent tests generally *increase* the minimum number of observed coincident strandings required to infer a positive correlation
- **Developed methods to account for underlying uncertainties in the data**
  - Such as ambiguity in how *coincident stranding* is defined, uncertainty of the actual stranding date, the possibility of existing but unreported stranding events, or the presence of other (non–US Navy) sonar
- **Introduced a draft Stranding Correlation Analysis Playbook (SCAP)**
  - The SCAP serves as an inference flowchart for stepping though the battery of statistical tests developed for this study
  - Multiple pathways through this flowchart are possible, depending on individual stakeholder preferences and requirements

Recommendations

- Use both *significance* and *power* tests to reject the null hypothesis (i.e., that stranding events are not correlated with sonar) and not just significance alone, as is currently done
- Use Monte Carlo sampling to determine the robustness of single test inferences with respect to uncertainties in the data
- Follow the guidelines in the SCAP flowchart to apply a sufficient battery of tests to achieve the desired level of inferential veracity

Outline

- **Review of past work**
  - Peer-reviewed academic journals | NOAA and Navy reports documenting strandings
- **Basic questions motivating this study**
  - The fundamental statistical analysis problem
- **The existing approach**
  - Nontechnical walkthrough | Technical details
- **Easiest first-cut solution**
  - Statistical inference lookup table → Accept/Reject Criteria Chart
- **Mutually confirming battery of statistical tests**
- **Mitigating uncertainties**
  - Reported vs. actual stranding dates | Definition of “coincidence” | Monte Carlo simulations
- **Case studies**
  - Real-world datasets
- **Pulling everything together**
  - Decision flowchart → Stranding Correlation Analysis Playbook
- **Recommendations | Next steps**
- **References**
- **Appendices**

Past CNA work correlating sonar use and strandings

Aquatic Mammals 2009, *35*(4), 435-444, DOI 10.1578/AM.35.4.2009.435

Correlating Military Sonar Use with Beaked Whale Mass Strandings: What Do the Historical Data Show?

Ronald Filadelfo,[1] Jonathon Mintz,[1] Edward Michlovich,[1] Angela D'Amico,[2]
Peter L. Tyack,[3] and Darlene R. Ketten[4]

[1]Center for Naval Analyses, 482
[2]Space and Naval Warfare Sy
[3]Woods Hole Oceanographic Institutic
[4]Woods Hole Oceanographic Institutic
and Harvard Medical School, Depa

Abstract

There have been several inciden
Navy sonar operations at sea c
and location with a mass stra
mammals, particularly beaked v
a conclusive cause-and-effect
not been established, there is str
scientific concern that use of m
resulted in beaked whale mass
previous attempts to determine
sonar use and whale stranding
have looked at mass stranding r
whales and have singled out th
which military operations appe
time and location with a mass st
this study, historical data on be
strandings and military exercises
to include active sonar use were c
tistical analyses were performed
level of correlation between thes
geographic regions. Strandings w
correlated with naval activity in t
and Caribbean Seas, but not off th
and southern California.

Key Words: beaked whales, Zi
Mesoplodon, *Stenella*, *Bala*
strandings, military sonar, boo
estimation, correlation analysis

Introduction

Concern about the effects of anthropogenic noise on marine life has grown over the last decade. Early concerns about ocean noise focused on an increase in the low-frequency noise from increased shipping traffic (Payne & Webb, 1971) as well as increased speed and size of large commercial vessels impacting large whales (National Research Council [NRC], 2003). In the 1990s, the debate about

several warships employing active sonar for protracted periods. Within hours of the ships' passages, 14 beaked whales (nine *Z. cavirostris*, three *Mesoplodon densirostris*, and two in which species could not be identified) were found stranded along the shores of Abaco and Grand Bahama to the north, and three single animal strandings of other species were reported nearby. Beaked whales

Aquatic Mammals 2009, *35*(4), 445-451, DOI 10.1578/AM.35.4.2009.445

Correlating Whale Strandings with Navy Exercises off Southern California

Ronald Filadelfo,[1] Yevgeniya K. Pinelis,[1] Scott Davis,[1] Robert Chase,[1]
Jonathon Mintz,[1] Jessica Wolfanger,[1] Peter L. Tyack,[2] Darlene R. Ketten,[3]
and Angela D'Amico[4]

SA; E-mail: filadelr@cna.org
, Woods Hole, MA 02543-1049, USA
, Woods Hole, MA 02543-1049, USA,
es Street, Boston, MA 02114, USA
iego, CA 92152-5001, USA

what conditions does use of
ine mammals, and under what
MFAS impel whales to strand?
ploring possible links between
hale strandings, we examined
ng occurrences and Navy exer-
California. As in Filadelfo et al.
on sonar use and whale strand-
ed for correlations in time and
er studies examined potential
een beaked whale mass strand-
ercises. This study focuses on
data from all cetacean species,
l data such as decomposition
ailable for southern California,
thodology that has not been used
on the state of decomposition of
nal, we estimated the time delay
l time of the stranding and the
l as a random variable. We then
elation between Navy exercises
th a Monte Carlo model.

erials and Methods

hern California Navy exercises
d using the procedures described
(this issue). We focused on exer-
l anti-submarine warfare (ASW)
g which, at some time, MFAS
been used.

Scientific and legal debate continues regarding the effect of military mid-frequency active sonars (MFAS) on marine mammals. Cox et al. (2006), Filadelfo et al. (this issue), and D'Amico et al. (this issue) compiled historical data on large-scale naval exercises and whale mass strandings. The fundamental questions remaining are as follows:

In Filadelfo et al. (this issue), Evans & England (2001), Freitas (2004), and D'Amico et al. (this issue), only mass strandings were examined for coincidence with naval operations. Individual strandings are far more common than mass strandings and thus produce "noisier" data, making the analyses more complex than those for mass strandings. In addition, although mass strandings occur

Statistical test → Assume a Poisson process

- If there is no significant seasonal effect to strandings, we can calculate rates for sonar and non-sonar days:
  - 5 regions x 4745 days = 23725 region-days
  - 822 sonar region-days

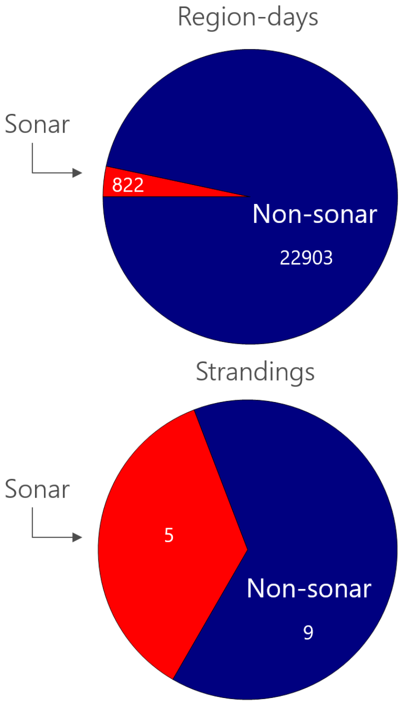
$$\mu = \frac{822}{23725} \cdot 14 \approx 0.485$$

$$\text{Probability}\left(n \geq 5 \mid \mu = 0.485\right) = \sum_{x=5}^{\infty} \frac{e^{-\mu} \cdot \mu^{x}}{x!} = 1 - \sum_{x=0}^{4} \frac{e^{-\mu} \cdot \mu^{x}}{x!}$$

$$\approx 0.00015$$

Statistical test of proportions

- Shows this difference to be statistically significant at >99% confidence level



Review of past work (1/2)

- Many (many!) research papers on strandings and accompanying analyses
  - Many discuss or compile instances of coincidence with sonar
- Some studies looked for conditions common to mass stranding events
  - Such as deep water near shore, surface ducting acoustic propagation conditions, wind direction, and shoreline
- Some used regression analysis to examine correlations of strandings with respect to various environmental variables
  - Such as seasonality, seismic events, proximity to a naval base, and presence of fringing reefs
- Others reviewed distant history, noting that strandings were extremely rare during the pre-sonar era

Review of past work (2/2)

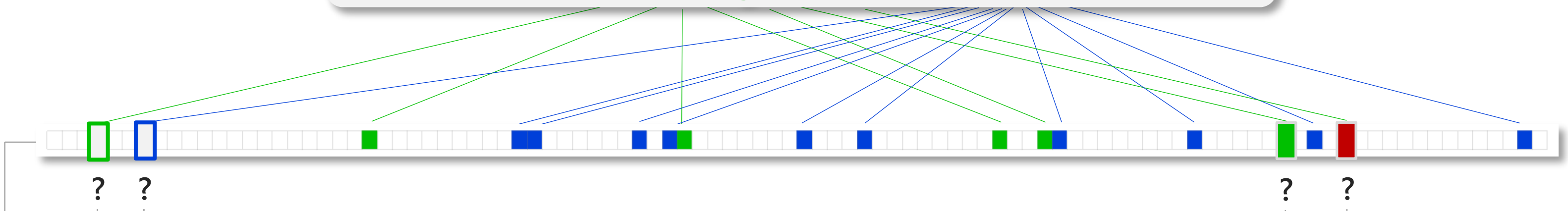
- Virtually no past research efforts have performed objective statistical analysis
  - CNA (2008)
    - Examined beaked whale strandings in the Mediterranean
  - CNA (2009)
    - Examined beaked whale strandings in southern California
  - Simonis et al. (2020)
    - Examined the level of event correlation between active sonar use and strandings in the Mariana Islands
  - Frantzis et al. (2003)
    - Examined the May 1996 Greece event
  - D'Amico et al. (2009)
    - Compiled a list of 126 beaked whale mass strandings from the 1870s to 2004
  - Quiros et al. (2019)
    - Noted that beaked whale mass strandings were very rare in the days before the advent of mid-frequency military sonars in the 1960s
  - Parsons et al. (2017)
    - Simply counted instances of coincidence
  - Foord et al. (2019)
    - Did not attempt to correlate strandings with military sonar or any particular cause

Review of past work (2/2) - *continued*

- Virtually no past research efforts have performed objective statistical analysis
  - CNA (2008)
    - Examined beaked whale strandings in the Mediterranean
  - CNA (2009)
    - Examined beaked whale strandings in southern California
  - Simonis et al. (2020)
    - Examined the level of event correlation between active sonar use and strandings in the Mariana Islands
  - Frantzis et al. (2003)
    - Examined the May 1996 Greece event
  - D'Amico et al. (2009)
    - Compiled a list of 126 beaked whale mass strandings from the 1870s to 2004
  - Quiros et al. (2019)
    - Noted that beaked whale mass strandings were very rare in the days before the advent of mid-frequency military sonars in the 1960s
  - Parsons et al. (2017)
    - Simply counted instances of coincidence
  - Foord et al. (2019)
    - Did not attempt to correlate strandings with military sonar or any particular cause

Basic questions motivating this study

Given a dataset that consists of a day-indexed time-series of **stranding events** and **sonar**



Q_1 How was the dataset prepared?

Q_2 How are the "first day" and length of time-series determined?

Q_3 What is the quality of the sonar data?

Q_4 Was non-US sonar active in same operating area?

Q_5 If a stranding event is observed and reported on a given day, when did it actually occur?

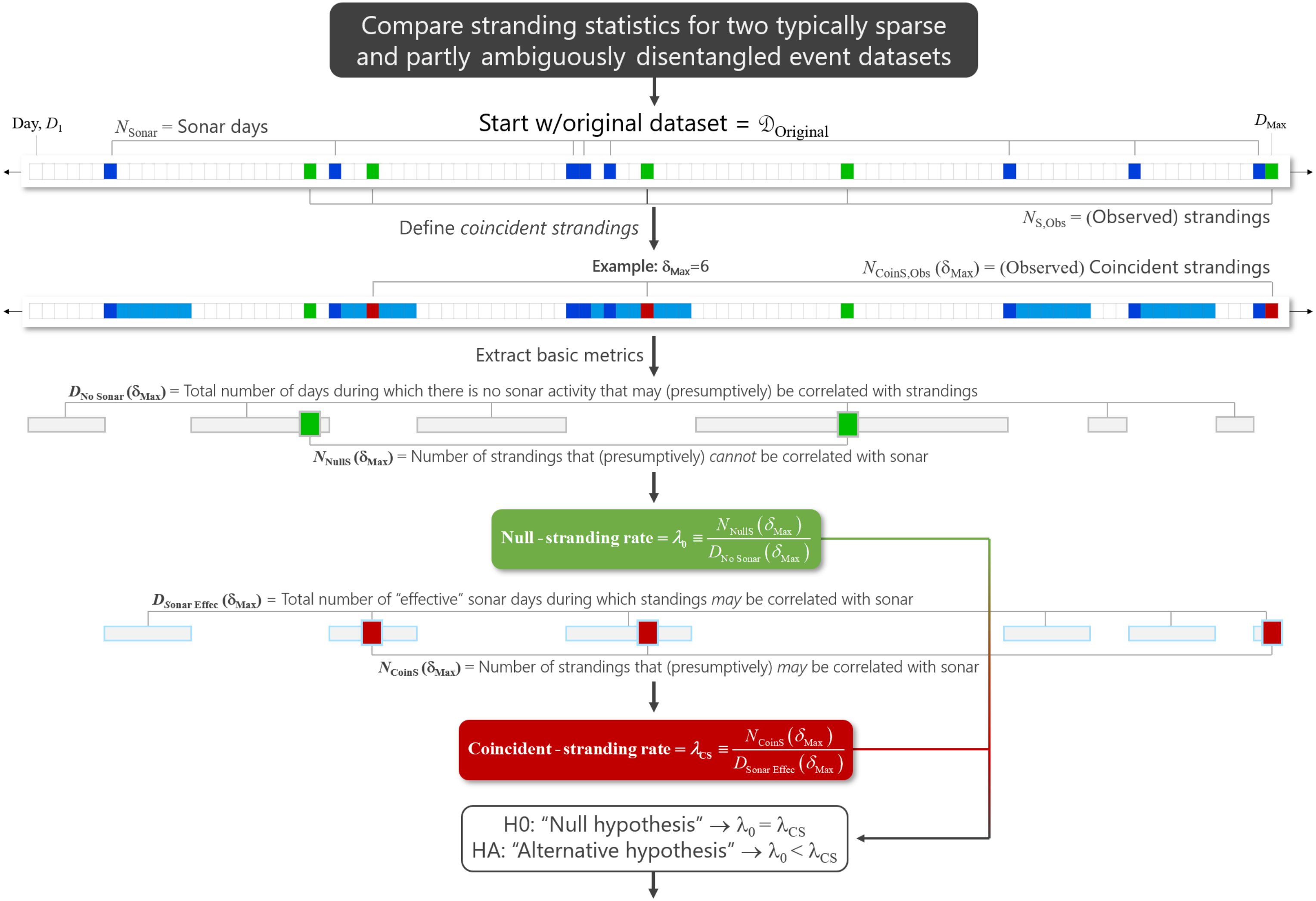
Q_6 Do other stranding events go unreported?

Q_7 How are coincident strandings determined?

Q_8 What information is lost by pigeonholing three-dimensional data (two space dimensions + time) into a one-dimensional time-series?

Q_9 How may the uncertainties in data and correlation analysis be best communicated to allow informed regulatory rulemaking?

The fundamental statistical analysis problem



Compare stranding statistics for two typically sparse and partly ambiguously disentangled event datasets

Start w/original dataset = $\mathcal{D}_{\text{Original}}$

Day, D_1

N_{Sonar} = Sonar days

D_{Max}

Define *coincident strandings*

$N_{\text{S,Obs}}$ = (Observed) strandings

Example: δ_{Max}=6

$N_{\text{CoinS,Obs}}(\delta_{\text{Max}})$ = (Observed) Coincident strandings

Extract basic metrics

$D_{\text{No Sonar}}(\delta_{\text{Max}})$ = Total number of days during which there is no sonar activity that may (presumptively) be correlated with strandings

$N_{\text{NullS}}(\delta_{\text{Max}})$ = Number of strandings that (presumptively) *cannot* be correlated with sonar

Null-stranding rate $= \lambda_0 \equiv \dfrac{N_{\text{NullS}}(\delta_{\text{Max}})}{D_{\text{No Sonar}}(\delta_{\text{Max}})}$

$D_{\text{Sonar Effec}}(\delta_{\text{Max}})$ = Total number of "effective" sonar days during which standings *may* be correlated with sonar

$N_{\text{CoinS}}(\delta_{\text{Max}})$ = Number of strandings that (presumptively) *may* be correlated with sonar

Coincident-stranding rate $= \lambda_{CS} \equiv \dfrac{N_{\text{CoinS}}(\delta_{\text{Max}})}{D_{\text{Sonar Effec}}(\delta_{\text{Max}})}$

H0: "Null hypothesis" → $\lambda_0 = \lambda_{CS}$
HA: "Alternative hypothesis" → $\lambda_0 < \lambda_{CS}$

Various statistical hypothesis tests are available

Each entails certain assumptions and limitations: no de facto "best" test
We recommend applying more than one test (for mutual confirmation), and
using Monte Carlo sampling to account for underlying uncertainties in the data

Nontechnical walkthrough of existing approach (1/2)

- **Virtually all researchers (inside and outside of CNA) have traditionally based the decision to either accept or reject the null hypothesis (i.e., that strandings and sonar are uncorrelated) on the results of applying a single Poisson means test**
- In this test, the significance (or P-value) of observing a given number of coincident strandings is compared to the number that one expects to see based on how many strandings occur on days without sonar
  - The P-value estimates the probability that two means (the *observed* and *expected* number of coincident strandings) fall outside an acceptance region within which the two means are assumed equal
  - Small P-values that are less than some critical threshold (typically 0.05) are interpreted as providing sufficient evidence to reject the null hypothesis
- However, two potential issues arise by following this approach
- **The first issue is that the method assumes that the observed null stranding rate (estimated by dividing the number of observed strandings on days without sonar by the number of no-sonar days) is the true average of an underlying random Poisson process**
  - In fact, the true average may be any number that lies within *a range of numbers* (called the confidence interval) that may be estimated by assuming an underlying Poisson process

Nontechnical walkthrough of existing approach (2/2)

- An immediate consequence is that one must compute not a *single* P-value (as almost all current strandings analyses do) but *rather a range of possible P-values* predicated on the possible null stranding rates that fall within the confidence interval
- For some scenarios, the difference between using a single mean estimate of P-values and averaging over a range of coincident rates falling within confidence interval will not effectively matter—in the sense that both tests may result in the same final inference
  - However, for other scenarios, significant differences may arise, typically resulting in more stringent criteria for rejecting the null hypothesis; for example, though a single mean estimate may, by itself, suggest that strandings and sonar are correlated (P-value < 0.05), averaging over a range of coincident rates within the confidence interval may push the P-value over the critical threshold (i.e., P-value > 0.05), which means the null hypothesis cannot be rejected
- **The second issue is that existing stranding analysis mitigates only so-called Type I (i.e., false positive) errors**
  - However, by itself, this approach is insufficient because we must simultaneously minimize the probability of making Type II (false negative) errors; that is, we must also minimize the probability that the null hypothesis is *false* but is erroneously *accepted*
  - Unfortunately, this test of power (of not making Type II errors) is seldom, if ever, applied
  - An alternative, more stringent test for accepting/rejecting the null hypothesis would be to estimate the minimum number of coincident strandings required to satisfy *both* Type I *and* Type II tests

Technical details of existing approach (1/6)

- Historically, the null hypothesis is accepted or rejected by ***applying a single means test***

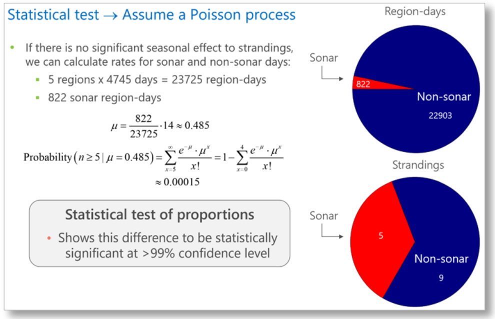
Typically, $\alpha_c = 0.01, 0.03, \text{or } 0.05$

Poisson

Significance Test: $\alpha_{\text{Poisson}} \le \alpha_c$? Y/N

P-value

$$\overbrace{\begin{cases} \alpha_{\text{Poisson}} = \text{Probablity}\left[N_{\text{CoinS}} \ge \overbrace{N_{\text{CoinS,Exp}}\left(\lambda_0\right)}^{\lambda_0 \cdot D_{\text{Sonar Effec}} = \text{Expected \# of coincident standings assuming the } null \text{ stranding rate}}\right] \approx \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right] \\ \approx \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \frac{e^{-N_{\text{CoinS,Exp}}(\lambda_0)} \cdot \left[N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right]^n}{n!} = 1 - \sum_{n=0}^{N_{\text{CoinS,Obs}}-1} \frac{e^{-N_{\text{CoinS,Exp}}(\lambda_0)} \cdot \left[N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right]^n}{n!} \end{cases}}$$



Statistical test → Assume a Poisson process

- If there is no significant seasonal effect to strandings, we can calculate rates for sonar and non-sonar days:
  - 5 regions x 4745 days = 23725 region-days
  - 822 sonar region-days

$$\mu = \frac{822}{23725} \cdot 14 \approx 0.485$$

$$\text{Probability}\left(n \ge 5 \mid \mu = 0.485\right) = \sum_{x=5}^{\infty} \frac{e^{-\mu} \cdot \mu^x}{x!} = 1 - \sum_{x=0}^{4} \frac{e^{-\mu} \cdot \mu^x}{x!} \approx 0.00015$$

Statistical test of proportions

- Shows this difference to be statistically significant at >99% confidence level

- There are two potential issues with this prima facie laudable approach
- **Issue #1** **It implicitly assumes that the *observed* null stranding rate = the *true* mean**
  - To emphasize: this estimate of the mean is based on a *single observation* of null strandings!
- Problem: Given that $N_{\text{NullS,Obs}}$ strandings have been observed on non-sonar days, determine ***confidence interval*** (CI) for the expected mean, $\mu_0 \approx D_{\text{Sonar Effec}} \times (N_{\text{NullS,Obs}} / D_{\text{No Sonar}})$
  - Find μ_{Lower} and μ_{Lower} such that: $Prob\underbrace{\left(\mu_{\text{Lower}} \le \mu_0 \le \mu_{\text{Upper}}\right)}=1-\alpha_c$

$$\begin{cases} \mu_{\text{Lower}} \approx \dfrac{1}{2 \cdot D_{\text{No Sonar}}} \cdot \chi^2\left[\alpha_c / 2, 2 \cdot N_{\text{NullS,Obs}}\right] \\ \mu_{\text{Upper}} \approx \dfrac{1}{2 \cdot D_{\text{No Sonar}}} \cdot \underbrace{\chi^2\left[1-\alpha_c / 2, 2 \cdot \left(N_{\text{NullS,Obs}} + 1\right)\right]} \end{cases}$$

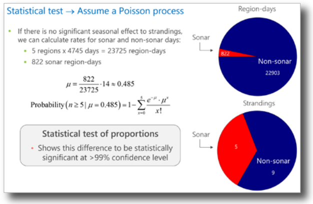
$\chi^2\left[\alpha, n\right] =$ the α^{th} percentile of the Chi-Square distribution with n degrees of freedom

Technical details of existing approach (2/6)

- Rather than use a *single* means test to adjudicate accepting/rejecting H0, may instead use a P-Value ***averaged* over *all* coincident rates falling within confidence interval** that are consistent with $D_{\text{Sonar Effec}}$, $N_{\text{NullS,Obs}}$, and $D_{\text{No Sonar}}$: $\mu_{\text{Lower}} \le \mu_i \le \mu_{\text{Lower}}$

$$\alpha_{\text{Poisson}}\left(\mu\right) = 1 - \sum_{n=0}^{N_{\text{CoinS,Obs}}-1} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}\left(\lambda_{\text{CS}}\right)\right] \rightarrow \boxed{\alpha_{\text{Poisson,Ave}} = \sum_{\mu=\mu_{\text{Lower}}}^{\mu_{\text{Lower}}} \rho\left(\mu\right) \cdot \alpha_{\text{Poisson}}\left(\mu\right)}$$

- For some scenarios, the difference between α_{Poisson} and $\alpha_{\text{Poisson,Ave}}$ may not matter
  - E.g.,



Statistical test → Assume a Poisson process

- If there is no significant seasonal effect to strandings, we can calculate rates for sonar and non-sonar days:
  - 5 regions x 4745 days = 23725 region-days
  - 822 sonar region-days

$$\mu = \frac{822}{23725} \cdot 14 \approx 0.485$$

$$\text{Probability}\left(n \ge 5 \mid \mu = 0.485\right) = 1 - \sum_{x=0}^{5} \frac{e^{-\mu} \cdot \mu^x}{x!}$$

Statistical test of proportions

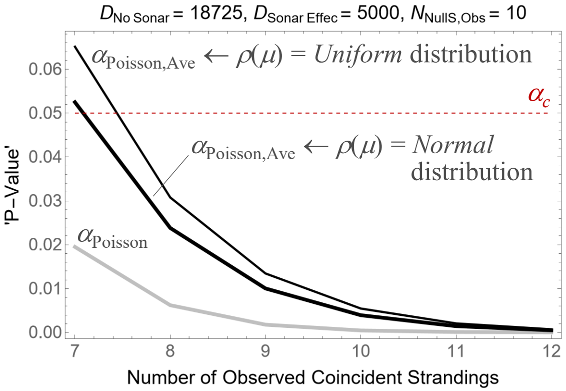
- Shows this difference to be statistically significant at >99% confidence level

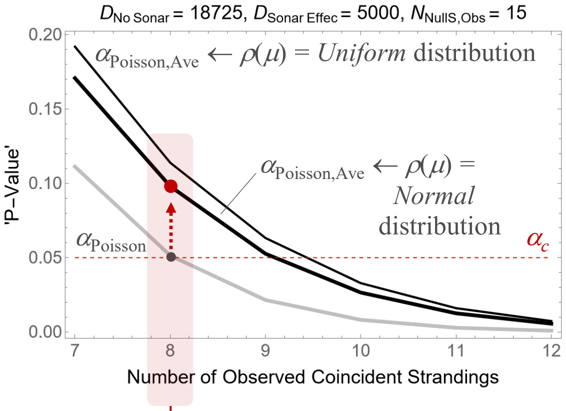
$$\rightarrow \alpha_{\text{Poisson}} \approx 0.0016 \text{ and } \alpha_{\text{Poisson,Ave}} \approx 0.0060$$

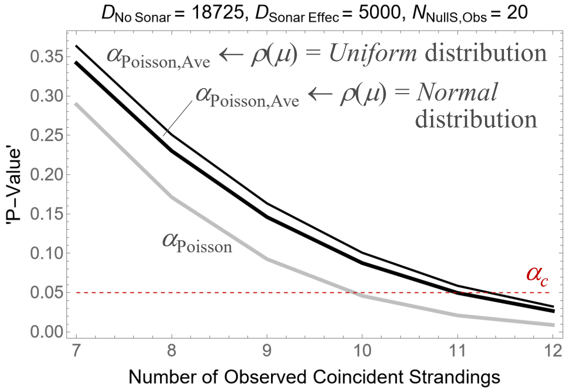
$$\rightarrow \textbf{Both are} \le \alpha_c = \mathbf{0.05}$$

Technical details of existing approach (3/6)

- However, for other scenarios, significant differences may arise; e.g.,



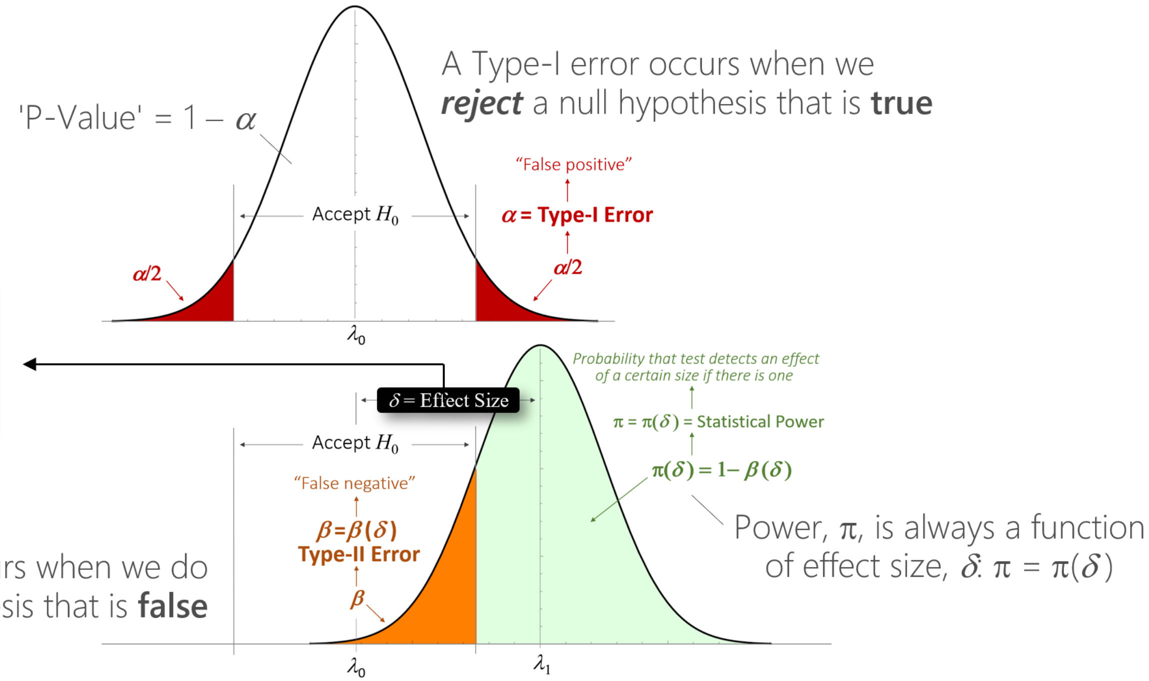




"OLD" test $\rightarrow$ **Reject** at 8 (since P-value ≤ α_c),
but $\alpha_{\text{Poisson,Ave}}(8) > \alpha_c \rightarrow$ ***Cannot* reject**

Technical details of existing approach (4/6)

- Issue #2 $\alpha \leq \alpha_c$ **test mitigates only Type I errors (false positives)**
  - Data may pass the *significance* test, but an inference cannot be credibly drawn if the *power* of the test is too small ← **this additional criterion is seldom, if ever, applied**
    - *Power*, $\pi \equiv$ probability of not making Type II errors (false negatives)



A Type-I error occurs when we ***reject*** a null hypothesis that is **true**

Power, π, is always a function of effect size, δ: $\pi = \pi(\delta)$

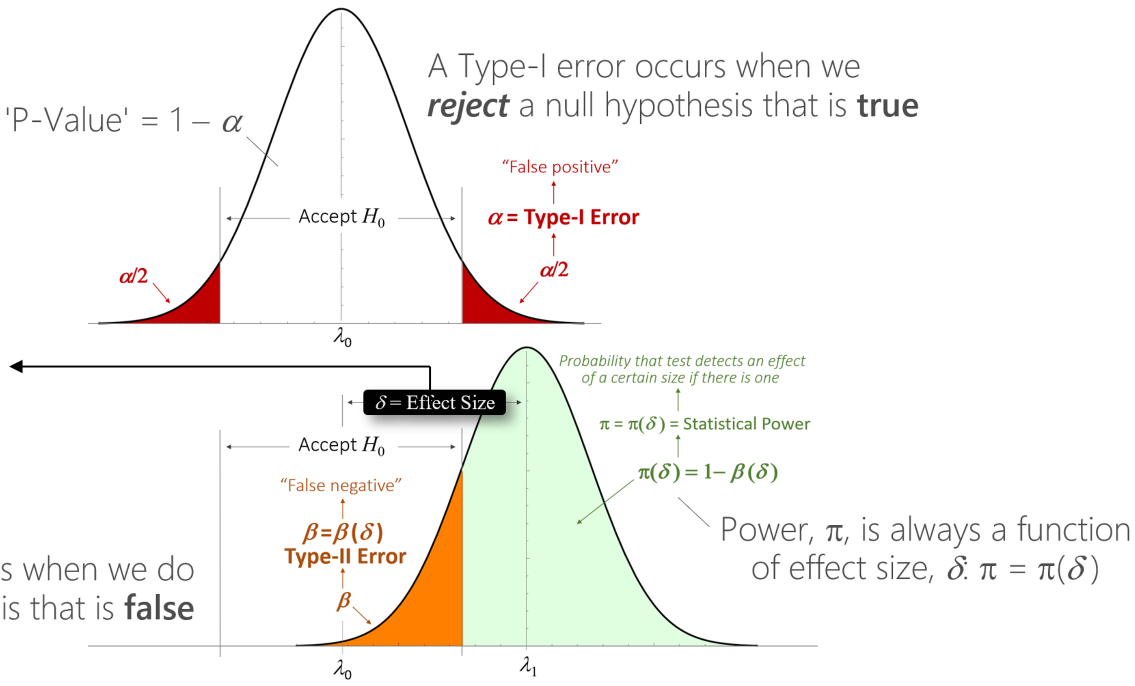
δ_{Min} represents the smallest statistically discernable difference between the null and alternative hypotheses, $\delta = \lambda_1 - \lambda_0$, for which $\alpha \leq \alpha_{\mathrm{Max}}$ and $\pi \geq \pi_{\mathrm{Min}}$

A Type-II error occurs when we do ***not reject*** a null hypothesis that is **false**

Technical details of existing approach (4/6) – *continued*

- **Issue #2** $\alpha \le \alpha_c$ **test mitigates only Type I errors (false positives)**
  - Data may pass the *significance* test, but an inference cannot be credibly drawn if the *power* of the test is too small ← **this additional criterion is seldom, if ever, applied**
    - *Power*, π ≡ probability of not making Type II errors (false negatives)

δ_{Min} represents the smallest statistically discernable difference between the null and alternative hypotheses, $\delta = \lambda_1 - \lambda_0$, for which $\alpha \le \alpha_{\text{Max}}$ and $\pi \ge \pi_{\text{Min}}$



A Type-I error occurs when we ***reject*** a null hypothesis that is **true**

A Type-II error occurs when we do ***not reject*** a null hypothesis that is **false**

Power, π, is always a function of effect size, δ: π = π(δ)

Technical details of existing approach (5/6)

- Estimate Type I and Type II errors → **reject null hypothesis only if *both* test positive**

	Poisson
Significance Test: $\alpha \leq \alpha_c$?	Y/N
Power Test: $\pi \geq \pi_c$?	Y/N

Typically, $\pi_c = 0.75, 0.8, \text{or } 0.85$

$$\begin{cases} \pi_{\text{Poisson}} = \text{Probablity}\left[N_{\text{CoinS}} \geq \overbrace{N_{\text{CoinS,Exp}}\left(\lambda_{\text{CS}}\right)}^{\boxed{\lambda_{\text{CS}}}\cdot D_{\text{Sonar Effec}} = \text{Expected \# of coincident standings assuming } \textit{coincident stranding rate}}\right] \approx \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}\left(\lambda_{\text{CS}}\right)\right] \\ \approx \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \frac{e^{-N_{\text{CoinS,Exp}}\left(\lambda_{\text{CS}}\right)} \cdot \left[N_{\text{CoinS,Exp}}\left(\lambda_{\text{CS}}\right)\right]^n}{n!} \approx 1 - \sum_{n=N_{\text{CoinS,Obs}}}^{N_{\text{CoinS,Obs}}-1} \frac{e^{-N_{\text{CoinS,Exp}}\left(\lambda_{\text{CS}}\right)} \cdot \left[N_{\text{CoinS,Exp}}\left(\lambda_{\text{CS}}\right)\right]^n}{n!} \end{cases}$$

- This equation for π is formally identical to the equation that defines α, except that λ_0 (= null stranding rate in the equation for α) is replaced by λ_{CS} (= coincident stranding rate), as determined by the ***observed*** number of coincident strandings, $N_{\text{CoinS,Obs}}$
- However, $\pi(N_{\text{CoinS,Obs}}+1) < \pi(N_{\text{CoinS,Obs}})$, and $\pi_{\text{Max}} \equiv \text{Max}[\pi(x)] \sim 0.63$ at $N_{\text{CoinS,Obs}} = 1$
  - The *Power* of rejecting H0 (at α as determined by comparing $N_{\text{CoinS,Obs}}$ to the ***expected*** number of coincident strandings, $N_{\text{CoinS,Exp}}$ (defined by λ_0), cannot pass the Type II test for *any* $\pi_c > 0.63$
  - This is true even if a scenario yields $\alpha \leq \alpha_c$; that is, the Type I test *alone* is satisfied!

    Example: expected # coincident strandings = 3 and observed # = 7 → *a* ≈ 0.03, but $\pi \approx 0.55$

Remains generally true even if one averages over all (possible) coincident strandings within a confidence interval

$$\pi_{\text{Poisson}} = 1 - \sum_{n=1}^{N_{\text{CoinS,Obs}}-1} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}\left(\lambda_{\text{CS}}\right)\right] \rightarrow \boxed{\pi_{\text{Poisson,Ave}} = \sum_{\mu=\mu_{\text{Lower}}}^{\mu_{\text{Lower}}} \rho(\mu) \cdot \pi_{\text{Poisson}}(\mu)}$$

$$\text{where} \begin{cases} \mu_{\text{Lower}} \approx \dfrac{1}{2 \cdot D_{\text{Sonar Effec}}} \cdot \chi^2\left[\alpha_c/2, 2 \cdot N_{\text{CoinS,Obs}}\right] \\ \mu_{\text{Upper}} \approx \dfrac{1}{2 \cdot D_{\text{Sonar Effec}}} \cdot \chi^2\left[1 - \alpha_c/2, 2 \cdot \left(N_{\text{CoinS,Obs}} + 1\right)\right] \end{cases}$$

Technical details of existing approach (6/6)

- An alternative statistical test to accept or reject the null hypothesis (H0)

Estimate the *minimum* number of coincident strandings required to satisfy *both* Type I and Type II tests, $N_{\mathrm{CoinS,Req}}$

- Step 1
  - Find the minimum null coincident stranding rate, H0 $= (\lambda_0)_{\mathrm{Min}}$, that yields $\alpha \leq \alpha_c$
- Step 2
  - Find the minimum coincident stranding rate, HA $= (\lambda_{\mathrm{CS}})_{\mathrm{Min}}$, that satisfies $\pi \geq \pi_c$
- Step 3
  - Estimate the minimum required number of coincident strandings: $N_{\mathrm{CoinS,Req}}[(\lambda_{\mathrm{CS}})_{\mathrm{Min}}]$

In the example above, the minimum number of observed coincidences required to satisfy both tests is 10 because the *power* for nine coincident strandings (≈ 0.79) is just shy of $\pi_c = 0.8$

→ **Reject H0 if the observed number of coincident strandings > $N_{\mathrm{CoinS,Req}}[(\lambda_{\mathrm{CS}})_{\mathrm{Min}}]$**

Easiest first-cut solution: *Poisson Mean Lookup Table* (1/3)

Number = $\alpha \geq \alpha_c = 0.05$, ***Number*** = $\alpha \leq \alpha_c = 0.05$, ■ = $\alpha \sim 0$, $\pi < \pi_c$, √ = $\alpha \leq \alpha_c$, $\pi \geq \pi_c$

← Number of ***observed*** of coincident strandings, $N_{\mathrm{CoinS,Obs}}$ →

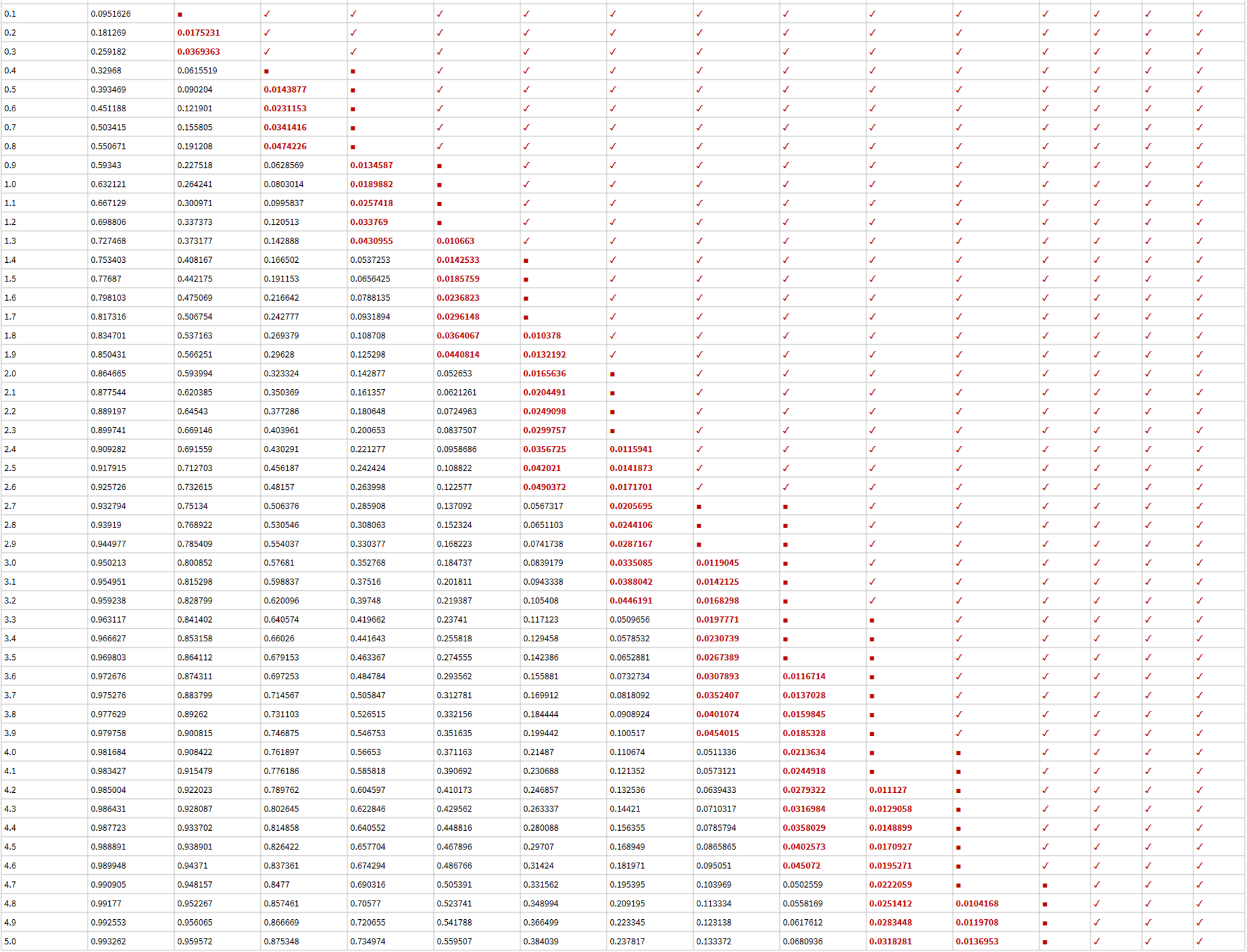
Number of ***expected*** coincident strandings under null hypothesis, $N_{\mathrm{CoinS,Exp}}$

	1	2	3	4	5	6	7
0.1	0.0951626	■	✓	✓	✓	✓	✓
0.2	0.181269	**0.0175231**	✓	✓	✓	✓	✓
0.3	0.259182	**0.0369363**	✓	✓	✓	✓	✓
0.4	0.32968	0.0615519	■	■	✓	✓	✓
0.5	0.393469	0.090204	**0.0143877**	■	✓	✓	✓
0.6	0.451188	0.121901	**0.0231153**	■	✓	✓	✓
0.7	0.503415	0.155805	**0.0341416**	■	✓	✓	✓
0.8	0.550671	0.191208	**0.0474226**	■	✓	✓	✓
0.9	0.59343	0.227518	0.0628569	**0.0134587**	■	✓	✓
1.0	0.632121	0.264241	0.0803014	**0.0189882**	■	✓	✓

Easiest first-cut solution: *Poisson Mean Lookup Table* (2/3)

← Number of ***observed*** of coincident strandings, $N_{\mathrm{CoinS,Obs}}$ →

Number of ***expected*** coincident strandings under null hypothesis, $N_{\mathrm{CoinS,Exp}}$

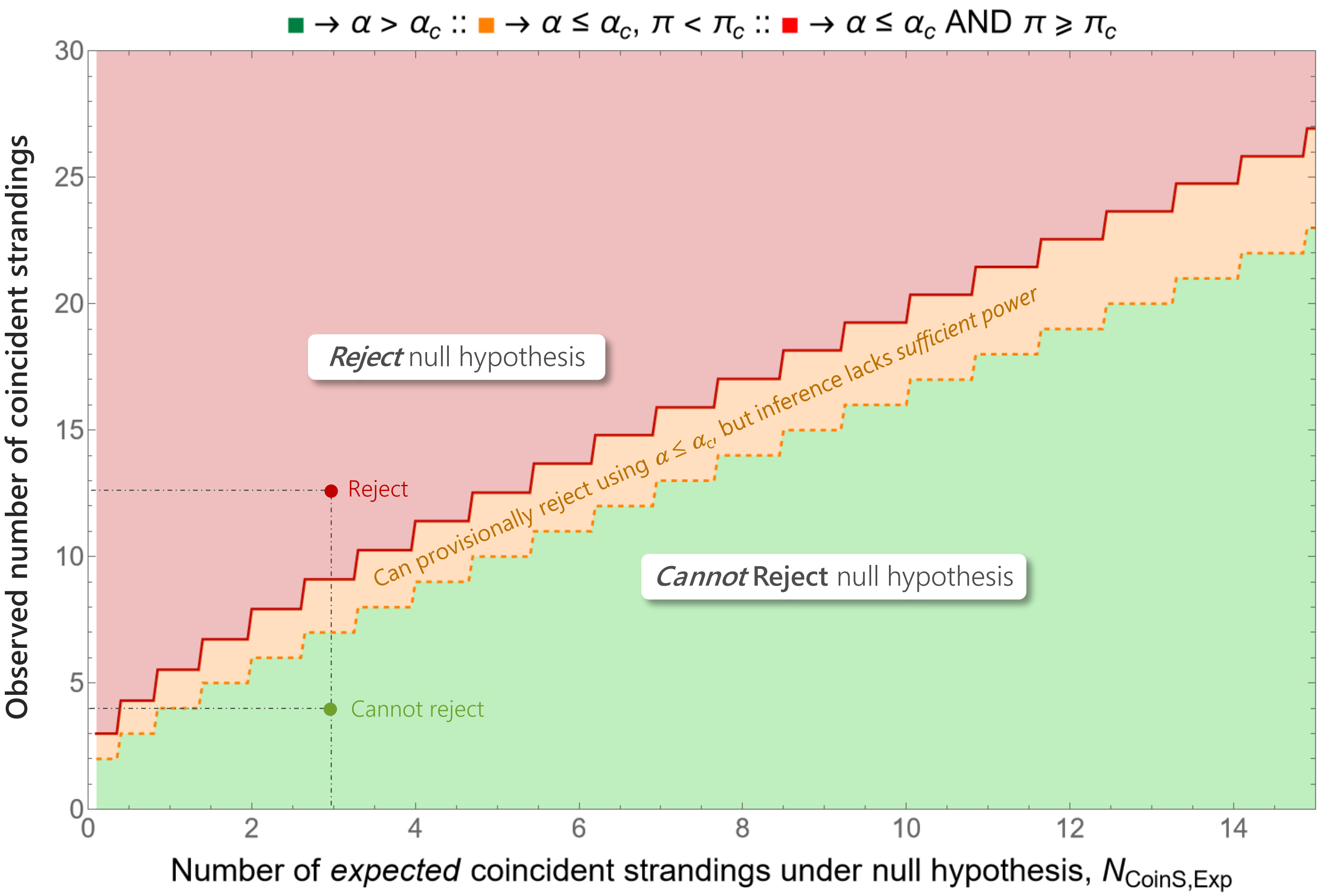


		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.1	0.1	0.0951626	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.2	0.2	0.181269	0.0175231	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.3	0.3	0.259182	0.0369363	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.4	0.4	0.32968	0.0615519	■	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.5	0.5	0.393469	0.090204	0.0143877	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.6	0.6	0.451188	0.121901	0.0231153	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.7	0.7	0.503415	0.155805	0.0341416	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.8	0.8	0.550671	0.191208	0.0474226	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
0.9	0.9	0.59343	0.227518	0.0628569	0.0134587	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.0	1.0	0.632121	0.264241	0.0803014	0.0189882	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.1	1.1	0.667129	0.300971	0.0995837	0.0257418	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.2	1.2	0.698806	0.337373	0.120513	0.033769	■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.3	1.3	0.727468	0.373177	0.142888	0.0430955	0.010663	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.4	1.4	0.753403	0.408167	0.166502	0.0537253	0.0142533	■	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.5	1.5	0.77687	0.442175	0.191153	0.0656425	0.0185759	■	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.6	1.6	0.798103	0.475069	0.216642	0.0788135	0.0236823	■	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.7	1.7	0.817316	0.506754	0.242777	0.0931894	0.0296148	■	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.8	1.8	0.834701	0.537163	0.269379	0.108708	0.0364067	0.010378	✓	✓	✓	✓	✓	✓	✓	✓	✓
1.9	1.9	0.850431	0.566251	0.29628	0.125298	0.0440814	0.0132192	✓	✓	✓	✓	✓	✓	✓	✓	✓
2.0	2.0	0.864665	0.593994	0.323324	0.142877	0.052653	0.0165636	■	✓	✓	✓	✓	✓	✓	✓	✓
2.1	2.1	0.877544	0.620385	0.350369	0.161357	0.0621261	0.0204491	■	✓	✓	✓	✓	✓	✓	✓	✓
2.2	2.2	0.889197	0.64543	0.377286	0.180648	0.0724963	0.0249098	■	✓	✓	✓	✓	✓	✓	✓	✓
2.3	2.3	0.899741	0.669146	0.403961	0.200653	0.0837507	0.0299757	■	✓	✓	✓	✓	✓	✓	✓	✓
2.4	2.4	0.909282	0.691559	0.430291	0.221277	0.0958686	0.0356725	0.0115941	✓	✓	✓	✓	✓	✓	✓	✓
2.5	2.5	0.917915	0.712703	0.456187	0.242424	0.108822	0.042021	0.0141873	✓	✓	✓	✓	✓	✓	✓	✓
2.6	2.6	0.925726	0.732615	0.48157	0.263998	0.122577	0.0490372	0.0171701	✓	✓	✓	✓	✓	✓	✓	✓
2.7	2.7	0.932794	0.75134	0.506376	0.285908	0.137092	0.0567317	0.0205695	■	■	✓	✓	✓	✓	✓	✓
2.8	2.8	0.93919	0.768922	0.530546	0.308063	0.152324	0.0651103	0.0244106	■	■	✓	✓	✓	✓	✓	✓
2.9	2.9	0.944977	0.785409	0.554037	0.330377	0.168223	0.0741738	0.0287167	■	■	✓	✓	✓	✓	✓	✓
3.0	3.0	0.950213	0.800852	0.57681	0.352768	0.184737	0.0839179	0.0335085	0.0119045	■	✓	✓	✓	✓	✓	✓
3.1	3.1	0.954951	0.815298	0.598837	0.37516	0.201811	0.0943338	0.0388042	0.0142125	■	✓	✓	✓	✓	✓	✓
3.2	3.2	0.959238	0.828799	0.620096	0.39748	0.219387	0.105408	0.0446191	0.0168298	■	✓	✓	✓	✓	✓	✓
3.3	3.3	0.963117	0.841402	0.640574	0.419662	0.23741	0.117123	0.0509656	0.0197771	■	■	✓	✓	✓	✓	✓
3.4	3.4	0.966627	0.853158	0.66026	0.441643	0.255818	0.129458	0.0578532	0.0230739	■	■	✓	✓	✓	✓	✓
3.5	3.5	0.969803	0.864112	0.679153	0.463367	0.274555	0.142386	0.0652881	0.0267389	■	■	✓	✓	✓	✓	✓
3.6	3.6	0.972676	0.874311	0.697253	0.484784	0.293562	0.155881	0.0732734	0.0307893	0.0116714	■	✓	✓	✓	✓	✓
3.7	3.7	0.975276	0.883799	0.714567	0.505847	0.312781	0.169912	0.0818092	0.0352407	0.0137028	■	✓	✓	✓	✓	✓
3.8	3.8	0.977629	0.89262	0.731103	0.526515	0.332156	0.184444	0.0908924	0.0401074	0.0159845	■	✓	✓	✓	✓	✓
3.9	3.9	0.979758	0.900815	0.746875	0.546753	0.351635	0.199442	0.100517	0.0454015	0.0185328	■	✓	✓	✓	✓	✓
4.0	4.0	0.981684	0.908422	0.761897	0.56653	0.371163	0.21487	0.110674	0.0511336	0.0213634	■	■	✓	✓	✓	✓
4.1	4.1	0.983427	0.915479	0.776186	0.585818	0.390692	0.230688	0.121352	0.0573121	0.0244918	■	■	✓	✓	✓	✓
4.2	4.2	0.985004	0.922023	0.789762	0.604597	0.410173	0.246857	0.132536	0.0639433	0.0279322	0.011127	■	✓	✓	✓	✓
4.3	4.3	0.986431	0.928087	0.802645	0.622846	0.429562	0.263337	0.14421	0.0710317	0.0316984	0.0129058	■	✓	✓	✓	✓
4.4	4.4	0.987723	0.933702	0.814858	0.640552	0.448816	0.280088	0.156355	0.0785794	0.0358029	0.0148899	■	✓	✓	✓	✓
4.5	4.5	0.988891	0.938901	0.826422	0.657704	0.467896	0.29707	0.168949	0.0865865	0.0402573	0.0170927	■	✓	✓	✓	✓
4.6	4.6	0.989948	0.94371	0.837361	0.674294	0.486766	0.31424	0.181971	0.095051	0.045072	0.0195271	■	✓	✓	✓	✓
4.7	4.7	0.990905	0.948157	0.8477	0.690316	0.505391	0.331562	0.195395	0.103969	0.0502559	0.0222059	■	■	✓	✓	✓
4.8	4.8	0.99177	0.952267	0.857461	0.70577	0.523741	0.348994	0.209195	0.113334	0.0558169	0.0251412	0.0104168	■	✓	✓	✓
4.9	4.9	0.992553	0.956065	0.866669	0.720655	0.541788	0.366499	0.223345	0.123138	0.0617612	0.0283448	0.0119708	■	✓	✓	✓
5.0	5.0	0.993262	0.959572	0.875348	0.734974	0.559507	0.384039	0.237817	0.133372	0.0680936	0.0318281	0.0136953	■	✓	✓	✓

Alternatively, consult the "Accept/Reject Criteria Chart" (A/R-CC)

Easiest first-cut solution: *Poisson Mean Lookup Table* (3/3)

Poisson Mean "Accept/Reject Criteria Chart" (PM-ARCC)



Additional statistical tests: nontechnical walkthrough

- **Best not to accept or reject the null hypothesis based on just a *single* test**
  - Accept or reject the null hypothesis only when "yea/nay" inferences of multiple tests *all* agree
- We recommend two additional statistical tests for two populations that may be used to mutually confirm the results of the Poisson means test:
  - **The exact binomial test** looks for differences between two Poisson means
  - **Fisher's exact test** is a significance test used to help analyze contingency tables
    - Contingency tables are matrices that contain the frequency distributions for combinations of two categorical variables (such as the presence or absence of sonar and strandings).

Additional statistical tests: technical details

- **Exact binomial test**
  - An exact test for analyzing the difference between two Poisson means

$$
\begin{cases}
\alpha_{\text{Binomial}} = \displaystyle\sum_{n=0}^{N_{\text{NullS}}} \binom{N_{\text{NullS}} + N_{\text{CoinS}}}{n} \cdot \left(\frac{D_{\text{No Sonar}}}{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}} \right)^{n} \cdot \left(1 - \frac{D_{\text{No Sonar}}}{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}} \right)^{N_{\text{NullS}} + N_{\text{CoinS}} - n} \\
\pi_{\text{Binomial}} = \underbrace{\displaystyle\sum_{n=0}^{\infty} \sum_{m=0}^{\infty}} \delta\left(\alpha_{\text{Binomial}} \le \alpha_{\text{c}}\right) \cdot Poisson\left(n, \lambda_0 \cdot D_{\text{No Sonar}}\right) \cdot Poisson\left(m, \lambda_{\text{CS}} \cdot D_{\text{Sonar Effec}}\right)
\end{cases}
$$

In practice, only terms for which the probability of n and m is significant need to be included (i.e., near $\lambda_0 \cdot D_{\text{No Sonar}}$ and $\lambda_{\text{CS}} \cdot D_{\text{Sonar Effec}}$, respectively)

- **Fisher's exact test**
  - "Exact" in the sense that α and π can both be calculated without approximations
  - FET designed to analyze the relative statistics of 2-by-2 contingency tables if the event size (i.e., number of strandings) is small compared to the sample (i.e., number of days)

	Sonar	No Sonar	Row Total
Stranding	N_{CoinS}	N_{NullS}	$N_{\text{S,Obs}}$
No Stranding	$D_{\text{Sonar Effec}} - N_{\text{CoinS}}$	$D_{\text{Max}} - N_{\text{NullS}} - D_{\text{Sonar Effec}}$	$D_{\text{Max}} - N_{\text{S,Obs}}$
Column Total	$D_{\text{Sonar Effec}}$	$D_{\text{No Sonar}}$	D_{Max}

Key assumption: row/column totals are fixed

The hypergeometric distribution describes the probability of having k successes in n draws without replacement from a population of size N that contains exactly K objects with the given feature (and such that each draw is either a success or failure)

$$
\begin{cases}
\alpha_{\text{Fisher}} = \displaystyle\sum_{n=0}^{N_{\text{NullS}}} \overbrace{h}^{h[x] = \textit{Hypergeometric probability distribution}}\left[n; D_{\text{No Sonar}}, D_{\text{Sonar Effec}}, N_{\text{NullS}} + N_{\text{CoinS}}\right] = \sum_{n=0}^{N_{\text{NullS}}} \frac{\binom{D_{\text{No Sonar}}}{n} \cdot \binom{D_{\text{Sonar Effec}}}{N_{\text{NullS}} + N_{\text{CoinS}} - n}}{\binom{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}}{N_{\text{NullS}} + N_{\text{CoinS}}}} \\
\pi_{\text{Fisher}} = \displaystyle\sum_{n=0}^{D_{\text{No Sonar}}} \sum_{m=0}^{D_{\text{Sonar Effec}}} \overbrace{\delta\left(\alpha_{\text{Fisher}} \le \alpha_{\text{c}}\right)}^{\delta(x) = \begin{cases} 1 \text{ if } x \text{ is } \textit{True} \\ 0 \text{ else} \end{cases}} \cdot \underbrace{f\left(n; D_{\text{No Sonar}}, \frac{N_{\text{NullS}}}{D_{\text{No Sonar}}}\right)}_{f(x;a,b) = \binom{a}{x} \cdot b^{x} \cdot (1-b)^{a-x}} \cdot f\left(m; D_{\text{Sonar Effec}}, \frac{N_{\text{CoinS}}}{D_{\text{Sonar Effec}}}\right)
\end{cases}
$$

Plethora of uncertainties (1/2)

- Sonar
  - Pre-SPORTS (2006) data very sparse | deployment of non–US Navy sonar
  - 🟥 (Possibly) ambiguous or inconsistent criteria for including in datasets
- Stranding events
  - Data completeness / randomness of observations
    - 🟥 Specter of existing but *unreported* strandings
  - 🟥 State of decay (actual stranding date versus observation date)
  - Size of stranding is rarely taken into account (as part of analysis)
    - Typically, $n > 1$ → “single stranding event”
- 🟥 Definition of *coincident stranding*
  - Presumes ability to do (approximate) space-time reconstruction
- Data size
  - Drawing inferences from very small sample sizes
    - Null hypothesis stranding rate (typically) based on only a few observed strandings
  - 🟥 Arbitrariness of time windows (defined by data *availability*)
- Confounding effects of other factors; specter of Simpson’s paradox
  - Such as seasonality, seismic events, and presence of fringing reefs

Plethora of uncertainties (2/2)

- Underlying distributions of strandings, sonar use
  - Previous analyses assume, but do not test for, Poisson statistics

Plethora of uncertainties: *example 1 (of 3)*

- Sonar
  - Pre-SPORTS (2006) data very sparse | Deployment or non U.S. Navy sonar
  - (Possibly) Ambiguous and/or inconsistent criteria for including in datasets
- Stranding events
  - – Specter of existing but *unreported* strandings
  - State of decay (actual stranding date vs. observation date)
  - Size of stranding is rarely taken into account (as part of analysis)
    - Typically, *n* > 1 → "single stranding event"
- Definition of "coincident stranding"
  - Presumes ability to do (approximate) space-time reconstruction
- Data "size"
  - Drawing inferences from very small sample sizes
    - Null-hypothesis stranding rate (typically) based on only a few observed strandings
  - Arbitrariness of "time windows" (defined by data *availability*)
- Confounding effects of other factors – specter of "Simpson's Paradox"
  - E.g., Seasonality, seismic events, presence of fringing reefs, …
- Underlying distributions of strandings, sonar use
  - Previous analyses assume, but do not test for, Poisson statistics

How robust is H0 rejection?

Ask: Given a dataset that satisfies the Type I significance test ($\alpha \leq \alpha_c$), how many *additional unobserved* non-coincident strandings are required for this test to fail (such that the null hypothesis cannot be rejected?

Assume there exists an *additional unobserved* non-coincident stranding

$$\mu_0 = \lambda_0 \cdot D_{\text{Sonar Effec}} \approx N_{\text{NullS}} \cdot f \,, \mu_1 = \left[N_{\text{NullS}} + \boxed{1} \; \right] \cdot f$$

$$\underbrace{\text{where } f = f\left(D_{\text{Sonar Effec}}, D_{\text{No Sonar}}\right) \equiv D_{\text{Sonar Effec}} / D_{\text{No Sonar}}}$$

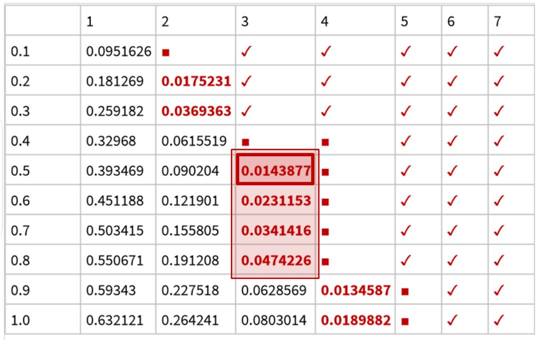
The value of which depends on the ratio of sonar days to non-sonar days (and, implicitly, on δ_{Max})

Each additional unobserved non-coincident stranding effectively increases the expected number of coincident strandings by $\delta\mu$

$$\delta_\mu \equiv \mu_1 - \mu_0 = f\left(D_{\text{Sonar Effec}}, D_{\text{No Sonar}}\right)$$

Number of ***observed*** of coincident strandings, $N_{\text{CoinS,Obs}}$

Number of ***expected*** coincident strandings under null hypothesis, $N_{\text{CoinS,Exp}}$



	1	2	3	4	5	6	7
0.1	0.0951626	■	✓	✓	✓	✓	✓
0.2	0.181269	0.0175231	✓	✓	✓	✓	✓
0.3	0.259182	0.0369363	✓	✓	✓	✓	✓
0.4	0.32968	0.0615519	■	■	✓	✓	✓
0.5	0.393469	0.090204	0.0143877	■	✓	✓	✓
0.6	0.451188	0.121901	0.0231153	■	✓	✓	✓
0.7	0.503415	0.155805	0.0341416	■	✓	✓	✓
0.8	0.550671	0.191208	0.0474226	■	✓	✓	✓
0.9	0.59343	0.227518	0.0628569	0.0134587	■	✓	✓
1.0	0.632121	0.264241	0.0803014	0.0189882	■	✓	✓

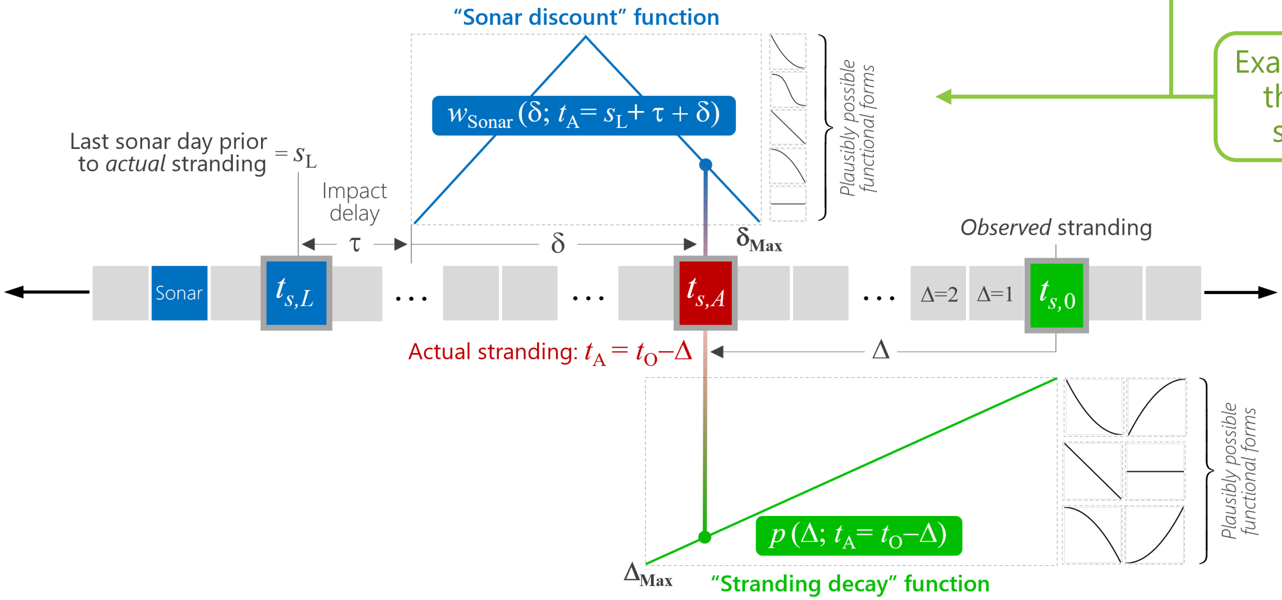
- With each pair of observed ($=N_{\text{CoinS,Obs}}$) and expected ($=N_{\text{CoinS,Exp}}$) coincident strandings for which $\alpha \leq \alpha_c$ is associated a **range of expected coincident strandings**, entailed by the presence of *unobserved* noncoincident strandings for which ***α remains*** $\leq \alpha_c$
- Each additional *unobserved* noncoincident stranding entails an effective increase in the *expected* number of coincident strandings, $\delta_\mu = D_{\text{Sonar-Effec}} \; / \; D_{\text{No Sonar}}$
- Apply the same logic to all elements highlighted in red (including pairs that satisfy both Type I and Type II errors); each such pair entails an associated "region of robustness"

Plethora of uncertainties: *example 2 (of 3)*

- Sonar
  - Pre-SPORTS (2006) data very sparse | Deployment or non U.S. Navy sonar
  - (Possibly) Ambiguous and/or inconsistent criteria for including in datasets
- Stranding events
  - Data completeness / randomness of observations
  - State of decay (actual stranding date vs. observation date)
  - Size of stranding is rarely taken into account (as part of analysis)
    - Typically, $n > 1 \rightarrow$ "single stranding event"
- Definition of "coincident stranding"
  - Presumes ability to do (approximate) space-time reconstruction
- Data "size"
  - Drawing inferences from very small sample sizes
    - Null-hypothesis stranding rate (typically) based on only a few observed strandings
  - Arbitrariness of "time windows" (defined by data *availability*)
- Confounding effects of other factors – specter of "Simpson's Paradox"
  - E.g., Seasonality, seismic events, presence of fringing reefs, ...
- Underlying distributions of strandings, sonar use
  - Previous analyses assume, but do not test for, Poisson statistics

A heuristic approach

Example of why this may be significant


"Sonar discount" function
$w_{\text{Sonar}}(\delta;\, t_A = s_L + \tau + \delta)$
Plausibly possible functional forms
Last sonar day prior to *actual* stranding $= s_L$
Impact delay
τ
δ
δ_{Max}
Observed stranding
Sonar
$t_{s,L}$
$t_{s,A}$
$\Delta=2$
$\Delta=1$
$t_{s,0}$
Actual stranding: $t_A = t_O - \Delta$
Δ
$p(\Delta;\, t_A = t_O - \Delta)$
Δ_{Max}
"Stranding decay" function
Plausibly possible functional forms

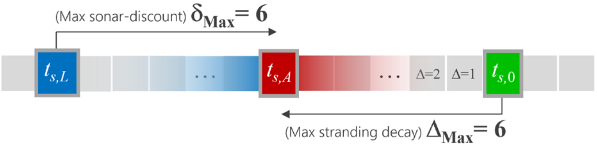
Fractional coincident stranding function

Fractional Coincident Stranding (FCS)

$$0 \le C_f\left(t_0\right) = \sum_{\Delta=0}^{\Delta_{\text{Max}}} p\left(t_0-\Delta\right)\cdot w\left(t_0-\Delta-s_{\text{L}}-\tau\right) \le 1$$

Compare to → (Typical) Existing Method

$$C_f = \left\{ \begin{array}{l} p\left(\Delta_{\text{Max}}=0\right)=1 \\ w_{\text{Sonar}}\left(\delta\right) = \begin{cases} 1 \text{ if } \delta \le \delta_{\text{Max}}=6 \\ 0 \text{ else} \end{cases} \end{array} \right\} \to C_f \in \{0,1\}$$


(Max sonar-discount) $\delta_{\text{Max}}=6$
$t_{s,L}$
...
$t_{s,A}$
...
Δ=2
Δ=1
$t_{s,0}$
(Max stranding decay) $\Delta_{\text{Max}}=6$

Salient takeaway

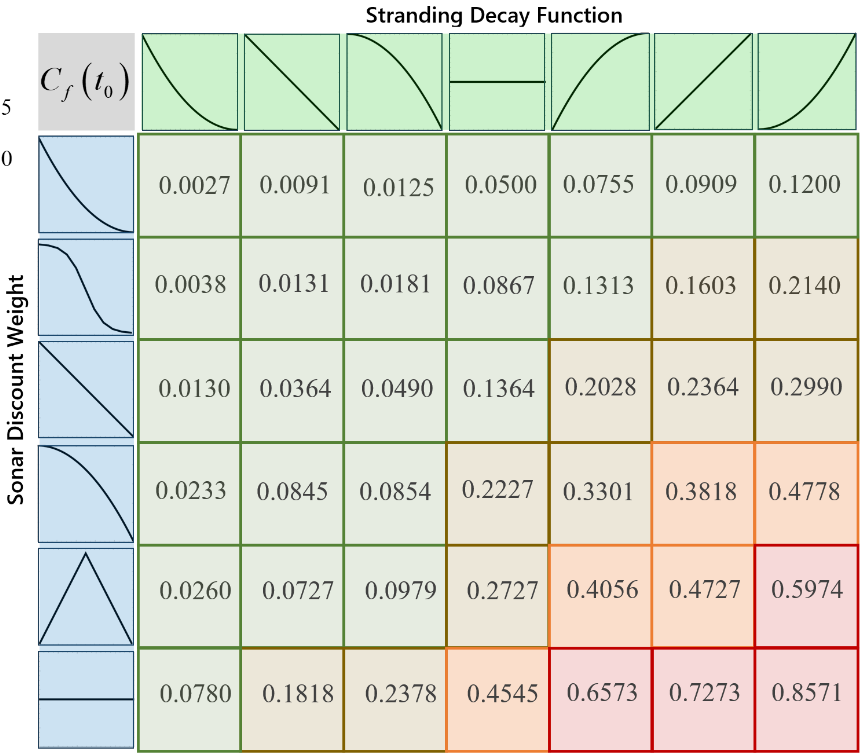
Wide range of possible scenarios in which any given observed stranding may be called "coincident" with sonar

🟩 $0 \le C_f(t_0) \le 0.15$

🟫 $0.15 < C_f(t_0) \le 0.35$

🟧 $0.35 < C_f(t_0) \le 0.50$

🟥 $0.50 < C_f(t_0) \le 1$

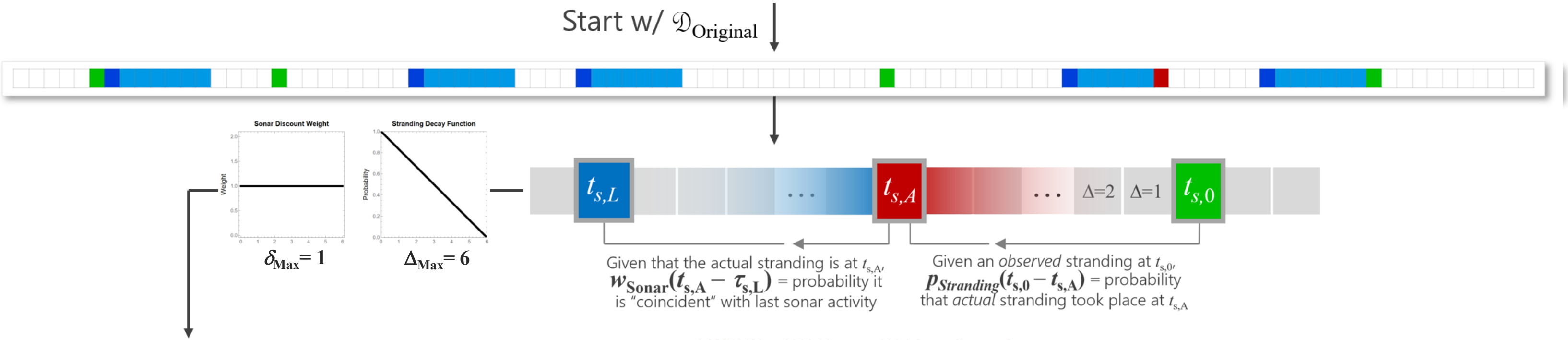


$C_f(t_0)$	Stranding Decay Function						
Sonar Discount Weight	0.0027	0.0091	0.0125	0.0500	0.0755	0.0909	0.1200
	0.0038	0.0131	0.0181	0.0867	0.1313	0.1603	0.2140
	0.0130	0.0364	0.0490	0.1364	0.2028	0.2364	0.2990
	0.0233	0.0845	0.0854	0.2227	0.3301	0.3818	0.4778
	0.0260	0.0727	0.0979	0.2727	0.4056	0.4727	0.5974
	0.0780	0.1818	0.2378	0.4545	0.6573	0.7273	0.8571

Use Monte Carlo simulation to explore distribution of possible outcomes for different assumptions and scenarios

Accounting for uncertainties: *Monte Carlo sampling*

Estimate probability distribution of *coincident* strandings



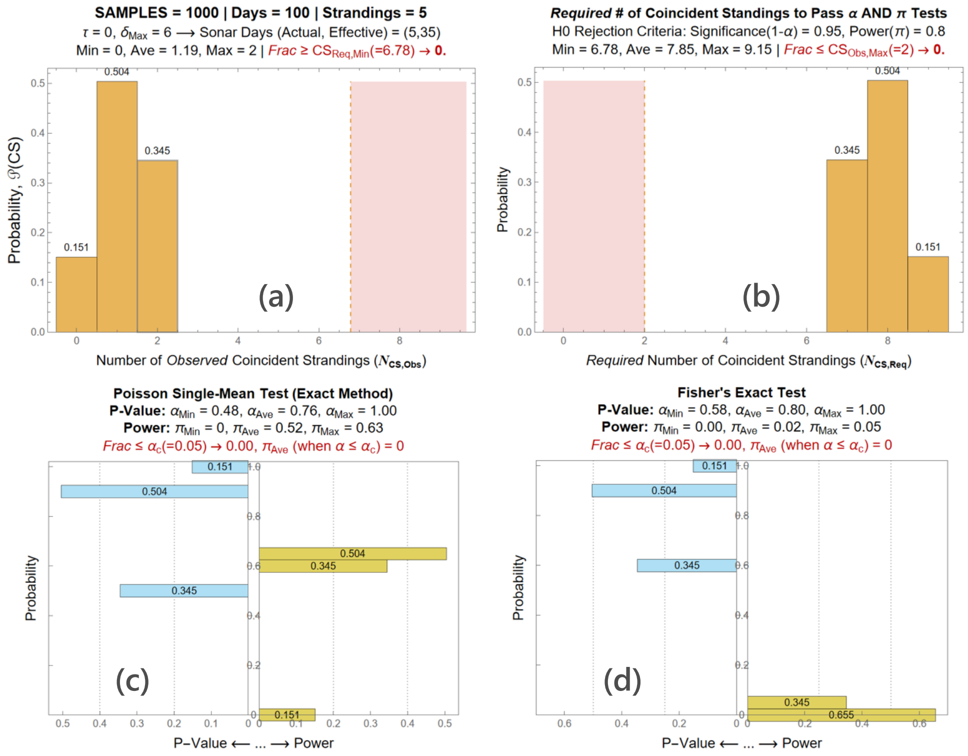
Start w/ $\mathcal{D}_{\text{Original}}$

Given that the actual stranding is at $t_{s,A}$, $w_{\text{Sonar}}(t_{s,A} - \tau_{s,L})$ = probability it is "coincident" with last sonar activity

Given an *observed* stranding at $t_{s,0}$, $p_{Stranding}(t_{s,0} - t_{s,A})$ = probability that *actual* stranding took place at $t_{s,A}$

Monte Carlo Simulation #1

```
$\tau, \delta_{\text{Max}}, \Delta_{\text{Max}}, w_{\text{Sonar}}$ and $p_{\text{Stranding}} \leftarrow$ Initialize
Coin_array ={0,0,...,0} // $N_{\text{S,Obs}}$+1 elements
for sample = 1,2,…,$N_{\text{Samples}}$
    Coin_counter = 0
    for stranding = 1, 2, …, $N_{\text{S,Obs}}$
        if Necropsy_flag = True, then
            Use necropsy-dependent parameters
        else
            Use default stranding-decay function
        end if
        $t_{\text{S,A}} \leftarrow p_{\text{Stranding}}$
        $t_{\text{S,L}} \leftarrow \tau, \delta_{\text{Max}}$
        if Random(0,1) $\leq w_{\text{Sonar}}(t_{\text{S,A}} - t_{\text{S,L}}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}}$ ← Coin_counter
    Apply statistics tests
        Poisson test :: $P_{\text{P\_value}}(\text{s}), P_{\text{Power}}(\text{s})$
        $\text{CS}_{\text{Min}}$ required to satisfy BOTH Type-I/II error tests
        Fisher's exact test :: $F_{\text{p\_value}}(\text{s}), F_{\text{power}}(\text{s})$
end for // sample
$\{\mathcal{P}(0), ..., \mathcal{P}(\text{CS}_{\text{Max}})\} \leftarrow$ Coin_array[[...]] / $N_{\text{Samples}}$
$<P> = P(\text{s}) / N_{\text{Samples}} \leftarrow$ Poisson test
$<F> = F(\text{s}) / N_{\text{Samples}} \leftarrow$ Fisher's exact test
```



SAMPLES = 1000 | Days = 100 | Strandings = 5
τ = 0, δ_{Max} = 6 ⟶ Sonar Days (Actual, Effective) = (5,35)
Min = 0, Ave = 1.19, Max = 2 | *Frac* ≥ $\text{CS}_{\text{Req,Min}}$(=6.78) → **0.**

***Required* # of Coincident Standings to Pass α AND π Tests**
H0 Rejection Criteria: Significance(1-α) = 0.95, Power(π) = 0.8
Min = 6.78, Ave = 7.85, Max = 9.15 | *Frac* ≤ $\text{CS}_{\text{Obs,Max}}$(=2) → **0.**

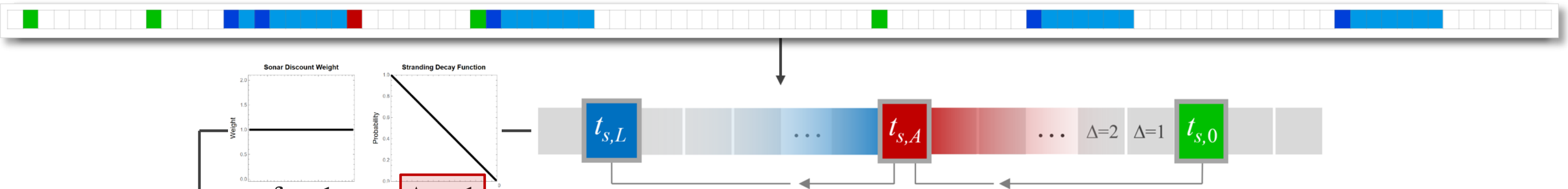
Poisson Single-Mean Test (Exact Method)
P-Value: α_{Min} = 0.48, α_{Ave} = 0.76, α_{Max} = 1.00
Power: π_{Min} = 0, π_{Ave} = 0.52, π_{Max} = 0.63
Frac ≤ α_c(=0.05) → 0.00, π_{Ave} (when $\alpha \leq \alpha_c$) = 0

Fisher's Exact Test
P-Value: α_{Min} = 0.58, α_{Ave} = 0.80, α_{Max} = 1.00
Power: π_{Min} = 0.00, π_{Ave} = 0.02, π_{Max} = 0.05
Frac ≤ α_c(=0.05) → 0.00, π_{Ave} (when $\alpha \leq \alpha_c$) = 0

Accounting for uncertainties: *Monte Carlo sampling*

Estimate probability distribution of *coincident* strandings

Start w/ $\mathcal{D}_{\text{Original}}$



$\delta_{\text{Max}} = 1$

$\Delta_{\text{Max}} = 1$

Given that the actual stranding is at $t_{s,A}$, $w_{\text{Sonar}}(t_{s,A} - \tau_{s,L})$ = probability it is "coincident" with last sonar activity

Given an *observed* stranding at $t_{s,0}$, $p_{Stranding}(t_{s,0} - t_{s,A})$ = probability that *actual* stranding took place at $t_{s,A}$

Test

Observed stranding = *Actual* stranding

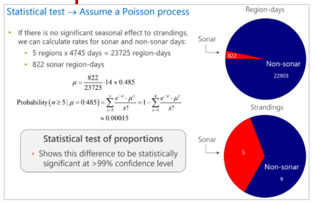
Monte Carlo Simulation #1

```
$\tau$, $\delta_{\text{Max}}$, $\Delta_{\text{Max}}$, $w_{\text{Sonar}}$ and $p_{\text{Stranding}}$ ← Initialize
Coin_array ={0,0,...,0} // $N_{\text{S,Obs}}$+1 elements
for sample = 1,2,…,$N_{\text{Samples}}$
    Coin_counter = 0
    for stranding = 1, 2, …, $N_{\text{S,Obs}}$
        if Necropsy_flag = True, the
            Use necropsy-dependent
        else
            Use default stranding-de
        end if
        $t_{\text{S,A}}$ ← $p_{\text{Stranding}}$
        $t_{\text{S,L}}$ ← $\tau$, $\delta_{\text{Max}}$
        if Random(0,1) ≤ $w_{\text{Sonar}}(t_{\text{S,A}} - t_{\text{S,L}}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}}$ ← Coin_counter
    Apply statistics tests
        Poisson test :: $P_{\text{P\_value}}$(s), $P_{\text{Power}}$(s)
        $\text{CS}_{\text{Min}}$ required to satisfy BOTH Type-I/II error tests
        Fisher's exact test :: $F_{\text{p\_value}}$(s), $F_{\text{power}}$(s)
end for // sample
{$\mathcal{P}$(0), ..., $\mathcal{P}(\text{CS}_{\text{Max}})$ } ← Coin_array[[...]] / $N_{\text{Samples}}$
<P> = P(s) / $N_{\text{Samples}}$ ← Poisson test
<F> = F(s) / $N_{\text{Samples}}$ ← Fisher's exact test
```

Summary of statistical tests

Samples=1000, Days=100, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$ =30, $N_{\text{S,Obs}}$=5, Δ_{Max}(Decay)=1 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{cs}	$(N_{\text{cs,Obs}})_{\text{Max}}$	$(N_{\text{cs,Req}})_{\text{Min}}$	Prob[$N_{\text{cs,Req}} \leq (N_{\text{cs,Obs}})_{\text{Max}}$]	Prob[$N_{\text{cs,Obs}} \geq (N_{\text{cs,Req}})_{\text{Min}}$]
	1	6.85714	0.	0.
1–Poisson	**Average[α]**	**Prob[$\alpha \leq \alpha_c$]**	**$\pi_{\text{Average}}[\alpha \leq \alpha_c]$**	**(Strength) $\mathcal{S}[\alpha \leq \alpha_c]$**
	0.819908	0.	0	$\mathcal{S}_{\text{Poisson}}$≈0, $\mathcal{S}_{\text{Bayes}}$≈0
	Average[α]	**Prob[$\alpha \leq \alpha_c$]**	**$\pi_{\text{Average}}[\alpha \leq \alpha_c]$**	**Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]**
Binomial	0.83193	0.	0	0.
Fisher	0.839243	0.	0	0.



Statistical test → Assume a Poisson process

- If there is no significant seasonal effect to strandings, we can calculate rates for sonar and non-sonar days:
  - 5 regions x 4745 days = 23725 region-days
  - 822 sonar region-days

$$\mu = \frac{822}{23725} \cdot 14 \approx 0.485$$

$$\text{Probability}(n \geq 5 \mid \mu = 0.485) = \sum_{x=5}^{\infty} \frac{e^{-\mu} \cdot \mu^x}{x!} = 1 - \sum_{x=0}^{4} \frac{e^{-\mu} \cdot \mu^x}{x!} \approx 0.00015$$

Statistical test of proportions

- Shows this difference to be statistically significant at >99% confidence level

(a)

This is the P-value that is conventionally estimated

(b)

The ***strongest*** test statistic is the probability that both Type I and II tests are simultaneously satisfied

Appendixes C and D contain illustrative examples

Plethora of uncertainties: *example 3 (of 3)*

- Sonar
  - (Possibly) Ambiguous and/or inconsistent criteria for including in datasets
- Stranding events
  - Data completeness / randomness of observations
    - Specter of existing but *unreported* strandings
  - State of decay (actual stranding date vs. observation date)
  - Size of stranding is rarely taken into account (as part of analysis)
    - Typically, $n > 1$ → "single stranding event"
- Definition of "coincident stranding"
  - Presumes ability to do (approximate) space-time reconstruction
- Data "size"
  - Drawing inferences from very small sample sizes
    - Null-hypothesis stranding rate (typically) based on only a few observed strandings
  - Arbitrariness of "time windows" (defined by data *availability*)
- Confounding effects of other factors – specter of "Simpson's Paradox"
  - E.g., Seasonality, seismic events, presence of fringing reefs, …
- Underlying distributions of strandings, sonar use
  - Previous analyses assume, but do not test for, Poisson statistics

Additional issues identified in Appendix E

How robust is H0 rejection?

Ask

Given a dataset that satisfies Type I "significance" ($\alpha \le \alpha_c$) and Type II "power" ($\pi \ge \pi_c$) tests, how many *additional* sonar days does it take for these tests to fail?

Example: Sonar Days = 5 (Total) and 25 ("Padded") using d_{Max} = 6

Samples=1000, Days=400, $N_{Sonar/Actual}$=5, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=35, $N_{S,Obs}$=7, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, $\Delta_{Sonar} = 0$

N_{CS}	$(N_{CS,Obs})$ Max	$(N_{CS,Req})$ Min	Prob[$N_{CS,Req} \le (N_{CS,Obs})$ Max]	Prob[$N_{CS,Obs} \ge (N_{CS,Req})$ Min]
	5	3.04932	1.	0.31
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	(Strength) $\mathbb{S}$[α ≤ α_c]
	0.0453349	0.691	0.571731	$\mathbb{S}_{Poisson}$≈0.761897, $\mathbb{S}_{Bayes}$≈0.98024
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.0608465	0.691	0.699535	0.31
Fisher	0.0601528	0.691	0.729257	0.31

Samples=1000, Days=400, $N_{Sonar/Actual}$=10, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=66, $N_{S,Obs}$=7, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, $\Delta_{Sonar} = 5$

N_{CS}	$(N_{CS,Obs})$ Max	$(N_{CS,Req})$ Min	Prob[$N_{CS,Req} \le (N_{CS,Obs})$ Max]	Prob[$N_{CS,Obs} \ge (N_{CS,Req})$ Min]
	6	2.99535	1.	0.756
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	(Strength) $\mathbb{S}$[α ≤ α_c]
	0.0993511	0.591	0.568614	$\mathbb{S}_{Poisson}$≈0.516903, $\mathbb{S}_{Bayes}$≈0.899377
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.135062	0.4	0.67885	0.099
Fisher	0.134245	0.4	0.697413	0.099

Samples=1000, Days=400, $N_{Sonar/Actual}$=15, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=96, $N_{S,Obs}$=7, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, $\Delta_{Sonar} = 10$

N_{CS}	$(N_{CS,Obs})$ Max	$(N_{CS,Req})$ Min	Prob[$N_{CS,Req} \le (N_{CS,Obs})$ Max]	Prob[$N_{CS,Obs} \ge (N_{CS,Req})$ Min]
	7	3.00192	0.978979	0.446
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	(Strength) $\mathbb{S}$[α ≤ α_c]
	0.156896	0.446	0.564288	$\mathbb{S}_{Poisson}$≈0.490057, $\mathbb{S}_{Bayes}$≈0.888575
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.213489	0.188	0.658547	0.015
Fisher	0.212706	0.188	0.677854	0.015

Prob[$\alpha \le \alpha_c, \pi \ge \pi_c$] decreases as Δ_{Sonar} increases

Modified Monte Carlo #1

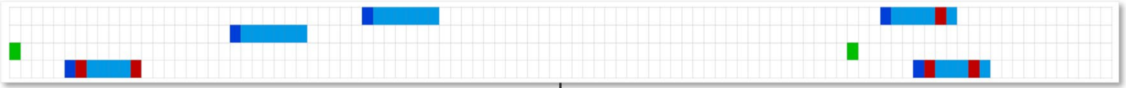
```
$\tau$, $\delta_{Max}$, $\Delta_{Max}$, $w_{Sonar}$ and $p_{Stranding}$, $\Delta_{Sonar}$ ← Initialize
Coin_array = {0,0,...,0} // $N_{S,Obs}$+1 elements
$\mathcal{D}_{Seed}$ ← $\mathcal{D}_{Original}$
for sample = 1,2,…,$N_{Samples}$
    for stranding = 1, 2, ..., $\Delta_{Sonar}$
        $\mathcal{D}_{Sample}$ ← random unused sonar date
    end if
    Coin_counter = 0
        …
        [same as original Monte Carlo]
        …
end for // sample
{$\mathcal{P}(0)$, ..., $\mathcal{P}(CS_{Max})$ } ← Coin_array[[...]] / $N_{Samples}$
<P> = P(s) / $N_{Samples}$ ← Poisson test
<F> = F(s) / $N_{Samples}$ ← Fisher's exact test
<B> = B(s) / $N_{Samples}$ ← Exact Binomial test
```

Additional tests: *bootstrapped vs. observed strandings* (1/5)

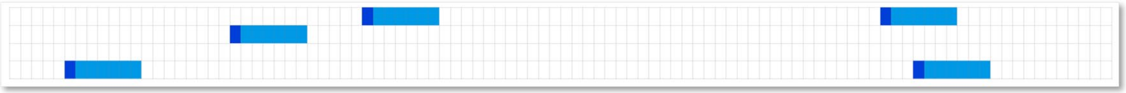
Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test A Randomly distribute a ***fixed*** number of total strandings

Start w/ $\mathfrak{D}_{\text{Original}}$



Strip $\mathfrak{D}_{\text{Original}}$ *of all strandings*

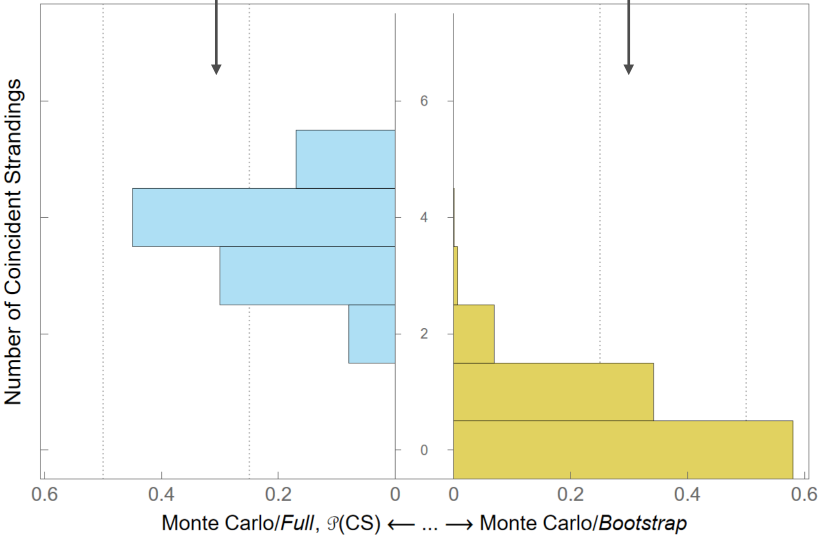


Monte Carlo Algorithm #2a

```
$\tau, \delta_{\text{Max}}, \Delta_{\text{Max}}, w_{\text{Sonar}}$ and $p_{\text{Stranding}} \leftarrow$ Initialize
**#Observed Strandings** $\leftarrow \mathfrak{D}_{\textbf{Original}}$
$\mathfrak{D}_{\textbf{Seed}} \leftarrow \mathfrak{D}_{\textbf{Original}} - \{1, \ldots, N_{\text{S,Obs}}\}$ // strip of strandings
for sample = $1, 2, \ldots, N_{\text{Samples}}$
    if Necropsy_flag = True, then
        Use necropsy-dependent parameters
    else
        Use default stranding-decay function
    end if
    $\mathfrak{D}_{\textbf{Sample}} \leftarrow$ Random[#Observed Strandings]
    Apply Monte Carlo Algorithm #1 to $\mathfrak{D}_{\textbf{Sample}}$
    Coin_counter = 0
    for stranding = $1, 2, \ldots, N_{\text{S,Obs}}$
        $t_{\text{S,A}} \leftarrow p_{\text{Stranding}}$
        $t_{\text{S,L}} \leftarrow \tau, \delta_{\text{Max}}$
        if $Random(0,1) \leq w_{\text{Sonar}}(t_{\text{S,A}} - t_{\text{S,L}}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}} \leftarrow$ Coin_counter
    end for // sample
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

Distribution of coincident strandings, as determined using **Monte Carlo #1**

Bootstrapped distribution of coincident strandings with random stranding dates



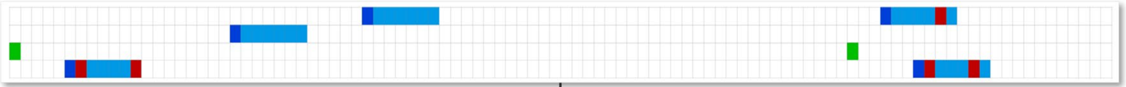
Additional tests: *bootstrapped vs. observed strandings* (2/5)

Ask →

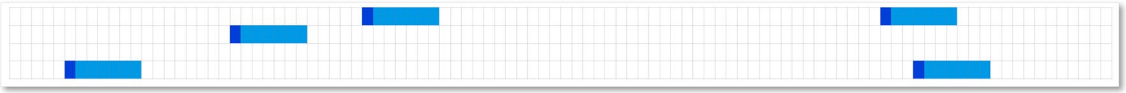
How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test A Randomly distribute a ***fixed*** number of total strandings

Start w/ $\mathfrak{D}_{\text{Original}}$ ↓

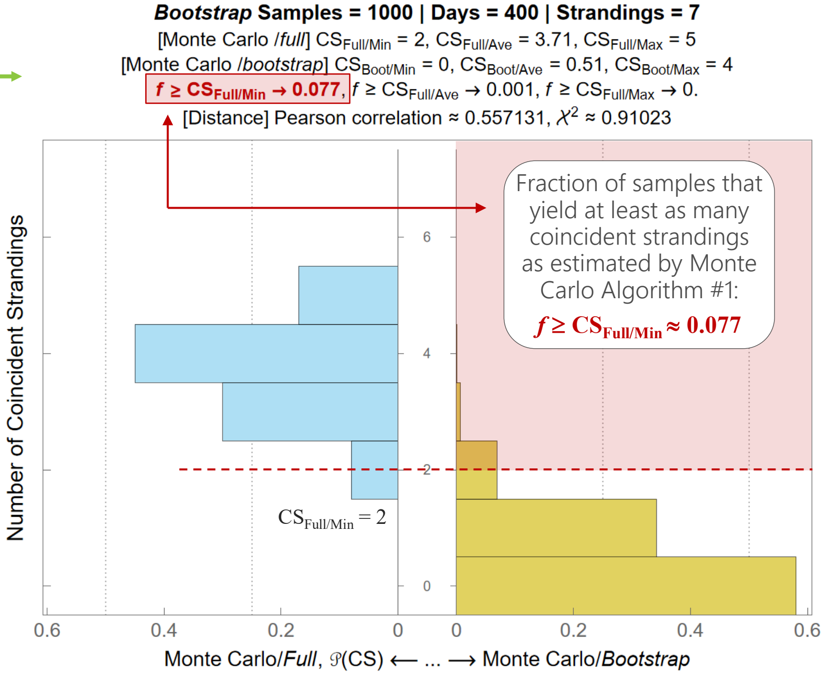


Strip $\mathfrak{D}_{\text{Original}}$ *of all strandings* ↓



Monte Carlo Algorithm #2a

```
$\tau$, $\delta_{\text{Max}}$, $\Delta_{\text{Max}}$, $w_{\text{Sonar}}$ and $p_{\text{Stranding}}$ ← Initialize
#Observed Strandings ← $\mathfrak{D}_{\textbf{Original}}$
$\mathfrak{D}_{\textbf{Seed}}$ ← $\mathfrak{D}_{\textbf{Original}}$ − {1, ..., $N_{\text{S,Obs}}$} // strip of strandings
for sample = 1,2,…,$N_{\text{Samples}}$
    if Necropsy_flag = True, then
        Use necropsy-dependent parameters
    else
        Use default stranding-decay function
    end if
    $\mathfrak{D}_{\textbf{Sample}}$ ← Random[#Observed Strandings]
    Apply Monte Carlo Algorithm #1 to $\mathfrak{D}_{\textbf{Sample}}$
    Coin_counter = 0
    for stranding = 1, 2, …, $N_{\text{S,Obs}}$
        $t_{\text{S,A}} \leftarrow p_{\text{Stranding}}$
        $t_{\text{S,L}} \leftarrow \tau, \delta_{\text{Max}}$
        if Random(0,1) ≤ $w_{\text{Sonar}}(t_{\text{S,A}} - t_{\text{S,L}}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}}$ ← Coin_counter
end for // sample
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```



***Bootstrap* Samples = 1000 | Days = 400 | Strandings = 7**

[Monte Carlo /*full*] $CS_{\text{Full/Min}} = 2$, $CS_{\text{Full/Ave}} = 3.71$, $CS_{\text{Full/Max}} = 5$

[Monte Carlo /*bootstrap*] $CS_{\text{Boot/Min}} = 0$, $CS_{\text{Boot/Ave}} = 0.51$, $CS_{\text{Boot/Max}} = 4$

$\boldsymbol{f \geq CS_{\text{Full/Min}} \rightarrow 0.077}$, $f \geq CS_{\text{Full/Ave}} \rightarrow 0.001$, $f \geq CS_{\text{Full/Max}} \rightarrow 0$.

[Distance] Pearson correlation ≈ 0.557131, $\chi^2 \approx 0.91023$

Fraction of samples that yield at least as many coincident strandings as estimated by Monte Carlo Algorithm #1: $f \geq CS_{\text{Full/Min}} \approx 0.077$

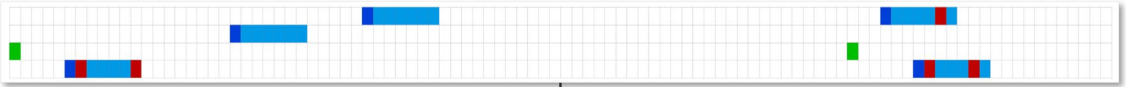
Number of Coincident Strandings
6
4
2
0
$CS_{\text{Full/Min}} = 2$
0.6
0.4
0.2
0
0
0.2
0.4
0.6
Monte Carlo/*Full*, $\mathcal{P}$(CS) ⟵ ... ⟶ Monte Carlo/*Bootstrap*

Additional tests: *bootstrapped vs. observed strandings* (3/5)

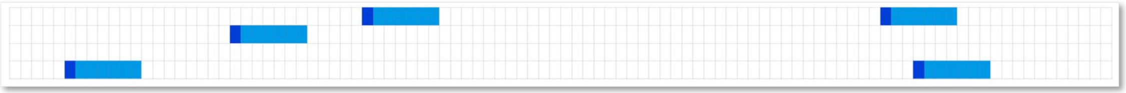
Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test A Randomly distribute a *fixed* number of total strandings

Start w/ $\mathfrak{D}_{\text{Original}}$



Strip $\mathfrak{D}_{\text{Original}}$ *of all strandings*



Monte Carlo Algorithm #2a

```
$\tau, \delta_{\text{Max}}, \Delta_{\text{Max}}, w_{\text{Sonar}}$ and $p_{\text{Stranding}} \leftarrow$ Initialize
#Observed Strandings $\leftarrow \mathfrak{D}_{\text{Original}}$
$\mathfrak{D}_{\text{Seed}} \leftarrow \mathfrak{D}_{\text{Original}} - \{1, \ldots, N_{\text{S,Obs}}\}$ // strip of strandings
for sample $= 1,2,\ldots,N_{\text{Samples}}$
    if Necropsy_flag = True, then
        Use necropsy-dependent parameters
    else
        Use default stranding-decay function
    end if
    $\mathfrak{D}_{\text{Sample}} \leftarrow$ Random[#Observed Strandings]
    Apply Monte Carlo Algorithm #1 to $\mathfrak{D}_{\text{Sample}}$
    Coin_counter = 0
    for stranding $= 1, 2, \ldots, N_{\text{S,Obs}}$
        $t_{\text{S,A}} \leftarrow p_{\text{Stranding}}$
        $t_{\text{S,L}} \leftarrow \tau, \delta_{\text{Max}}$
        if Random(0,1) $\leq w_{\text{Sonar}}(t_{\text{S,A}} - t_{\text{S,L}}; \tau)$
            Coin_counter = Coin_counter + 1
    end for // stranding
    Coin_array[[Coin_counter ]]/$N_{\text{Samples}} \leftarrow$ Coin_counter
    end for // sample
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

***Bootstrap* Samples = 1000 | Days = 400 | Strandings = 7**

[Monte Carlo /*full*] $\text{CS}_{\text{Full/Min}} = 2$, $\text{CS}_{\text{Full/Ave}} = 3.71$, $\text{CS}_{\text{Full/Max}} = 5$

[Monte Carlo /*bootstrap*] $\text{CS}_{\text{Boot/Min}} = 0$, $\text{CS}_{\text{Boot/Ave}} = 0.51$, $\text{CS}_{\text{Boot/Max}} = 4$

$f \geq \text{CS}_{\text{Full/Min}} \rightarrow 0.077$, $f \geq \text{CS}_{\text{Full/Ave}} \rightarrow 0.001$, $f \geq \text{CS}_{\text{Full/Max}} \rightarrow 0$.

[Distance] Pearson correlation ≈ 0.557131, $\mathcal{X}^2 \approx 0.91023$

$$\begin{cases} D_{\text{Pearson Corr}}\left(\vec{h}_1, \vec{h}_2\right) = \dfrac{\sum_i \left(h_{1,i} - \left\langle \vec{h}_1 \right\rangle\right) \cdot \left(h_{1,i} - \left\langle \vec{h}_2 \right\rangle\right)}{\sqrt{\sum_i \left(h_{1,i} - \left\langle \vec{h}_1 \right\rangle\right)^2} \cdot \sqrt{\sum_i \left(h_{2,i} - \left\langle \vec{h}_2 \right\rangle\right)^2}} \\ D_{\chi^2}\left(\vec{h}_1, \vec{h}_1\right) = \displaystyle\sum_{i=1}^{n} \frac{\left(h_{1,i} - h_{2,i}\right)^2}{h_{1,i} + h_{2,i}} \end{cases}$$

where $\vec{h}_k \equiv \left(h_{k,1}, h_{k,2}, \ldots, h_{k,N_k}\right)$ are histogram frequency vectors

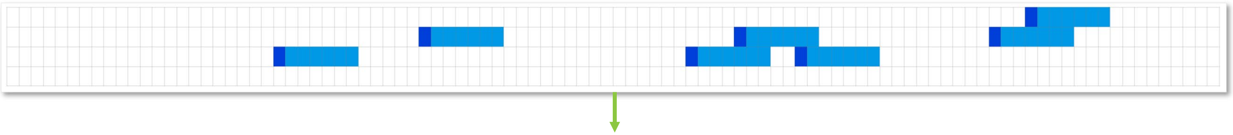
- Nothing sacrosanct about using the *Pearson correlation* and χ^2
- Many other metrics are possible; we have selected two common ones that have the added virtue of being symmetric in h_1 and h_2
- Other metrics include Bhattacharyya, Earth mover's distance, Euclidean, Intersection, and Kullback-Leibler divergence

Additional tests: *bootstrapped vs. observed strandings* (4/5)

Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test B Sample over random datasets using the estimated ***distribution*** of null stranding rates

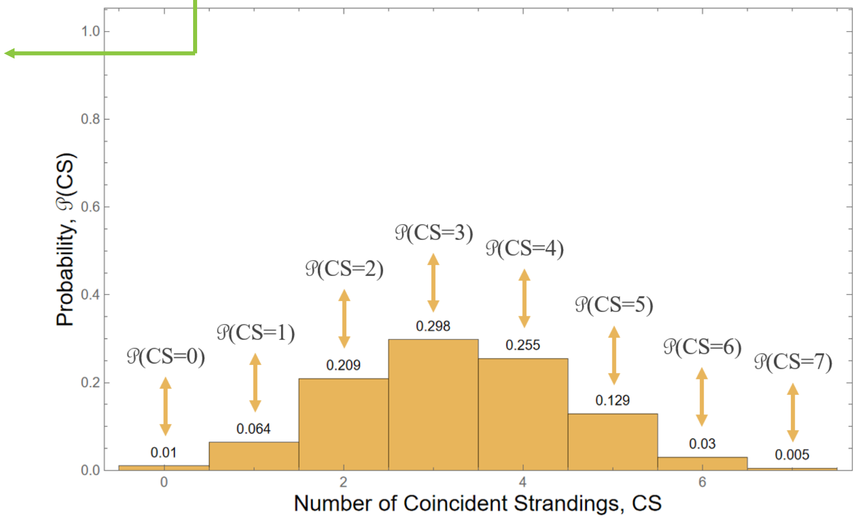
Start w/ $\mathfrak{D}_{\text{Original}}$ stripped of all observed strandings



Import data from Monte Carlo Algorithm #1

Monte Carlo Algorithm #2b

```
$\tau$, $\delta_{\text{Max}}$, $\Delta_{\text{Max}}$, $w_{\text{Sonar}}$ and $p_{\text{Stranding}}$ ← Initialize
for $CS = 0,1,\ldots,CS_{Max}$
    // strip away all strandings
    $\mathfrak{D}_{\text{CS}} \leftarrow \mathfrak{D}_{\text{Original}} - \{1, 2, \ldots, N_{\text{S,Obs}}\}$
    $\lambda_0$ ← assume # coincident strandings = $CS$
    Initialize all interim arrays and counters
    Number_of_strandings = 0
    for sample = $1,2,\ldots,N_{\text{Samples}}$
        for day = $1, 2, \ldots, D_{\text{Max}}$
            if Random(0,1) $\le \lambda_0$,
                $\mathfrak{D}_{\text{CS}}$ + Stranding@d ← $\mathfrak{D}_{\text{CS}}$
            end if
            Number_of_strandings =
                Number_of_strandings + 1
        end for // day
            …
            [same as original Monte Carlo]
            …
    end for // sample
end for // CS
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

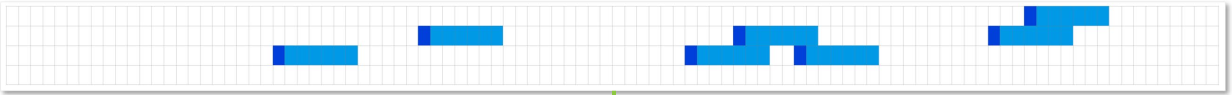


Additional tests: *bootstrapped vs. observed strandings* (5/5)

Ask → How likely is it that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?

Test B Sample over random datasets using the estimated ***distribution*** of null stranding rates

Start w/ $\mathcal{D}_{\text{Original}}$ stripped of all observed strandings

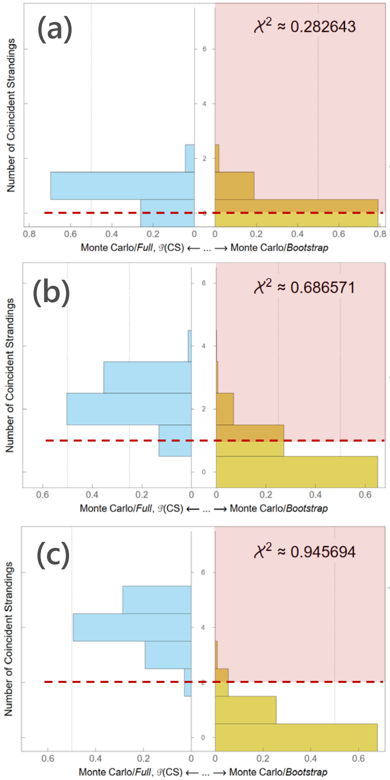


Import data from Monte Carlo Algorithm #1

Monte Carlo Algorithm #2b

```
$\tau$, $\delta_{\text{Max}}$, $\Delta_{\text{Max}}$, $w_{\text{Sonar}}$ and $p_{\text{Stranding}}$ ← Initialize
for $CS = 0,1,\ldots,CS_{Max}$
    // strip away all strandings
    $\mathcal{D}_{\text{CS}} \leftarrow \mathcal{D}_{\textbf{Original}} - \{1, 2, \ldots, N_{\text{S,Obs}}\}$
    $\lambda_0$ ← assume # coincident strandings = $CS$
    Initialize all interim arrays and counters
    Number_of_strandings = 0
    for sample = $1,2,\ldots,N_{\text{Samples}}$
        for day = $1, 2, \ldots, D_{\text{Max}}$
            if Random(0,1) $\leq \lambda_0$,
                $\mathcal{D}_{\text{CS}}$ + Stranding@d ← $\mathcal{D}_{\text{CS}}$
            end if
            Number_of_strandings =
                Number_of_strandings + 1
        end for // day
            …
            [same as original Monte Carlo]
            …
    end for // sample
end for // CS
Compare bootstrap CS distribution to MC#1
Pearson correlation
Chi-Squared distance
```

N_{CoinS}	N_{NullS}
$\delta_{\text{Max}} = 6$,	$\Delta_{\text{Max}} = 6$
1	6
3	3
6	1



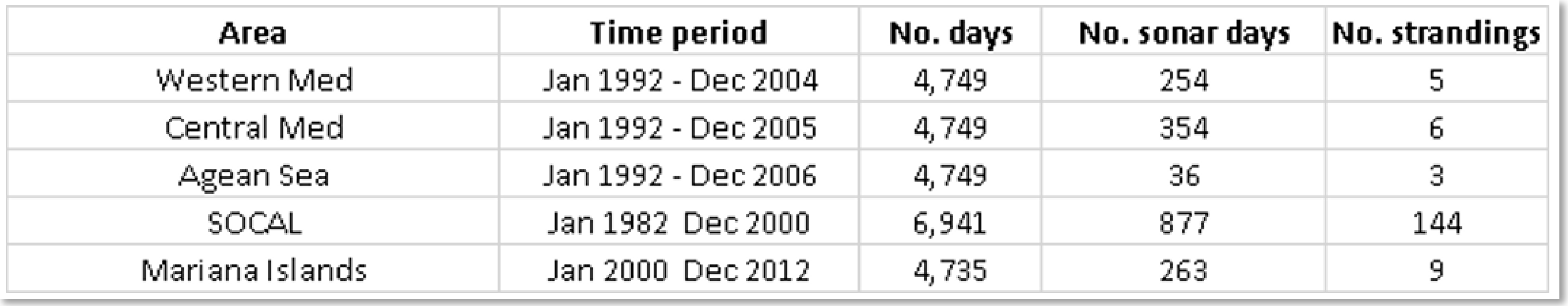
Fraction of samples that yield at least as many coincident strandings as estimated by MC #1

$f \geq \text{CS}_{\text{Full/Min}} \approx 1$

$f \geq \text{CS}_{\text{Full/Min}} \approx 0.349$

$f \geq \text{CS}_{\text{Full/Min}} \approx 0.064$

Real-world datasets



Area	Time period	No. days	No. sonar days	No. strandings
Western Med	Jan 1992 - Dec 2004	4,749	254	5
Central Med	Jan 1992 - Dec 2005	4,749	354	6
Agean Sea	Jan 1992 - Dec 2006	4,749	36	3
SOCAL	Jan 1982 Dec 2000	6,941	877	144
Mariana Islands	Jan 2000 Dec 2012	4,735	263	9

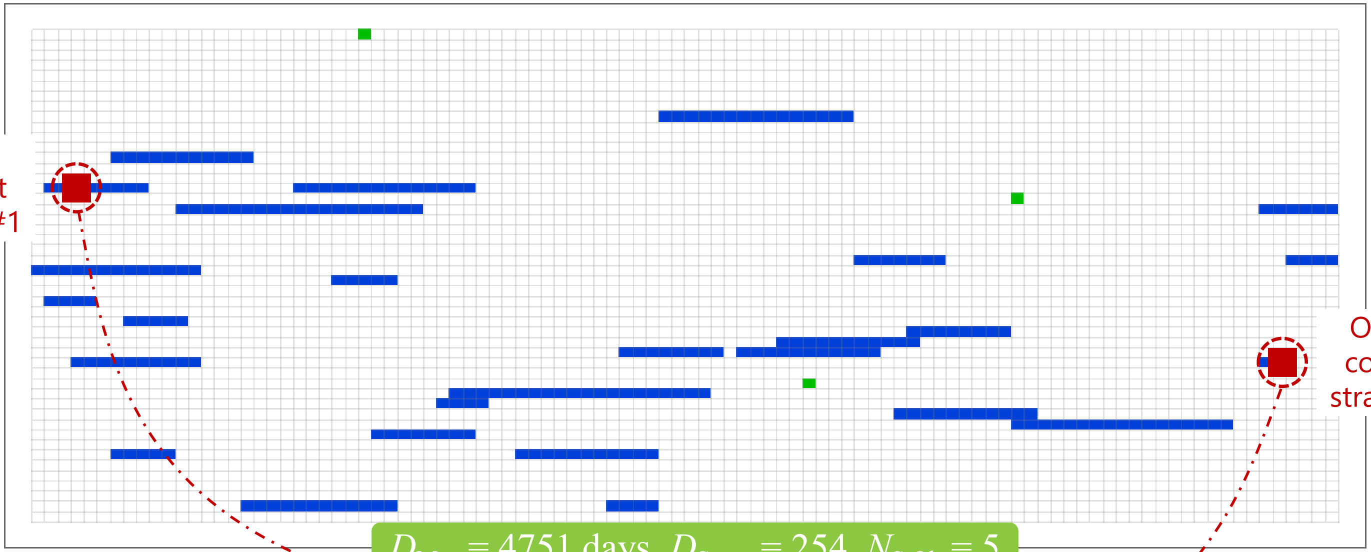
Real-world datasets – *continued*

- Mediterranean
  - Sonar use: 1992–2004, location by basin
  - Strandings: Beaked whale mass, rough geographic location
- SOCAL
  - Sonar use: 1982–2002; geographic coordinates (SPORTS)
  - Strandings: Multi-species singles, geographic coordinates (NOAA data)
- Hawai'i-Mariana Islands
  - Strandings only, 100 years, singles, rough location information
- Mariana Islands ("Simonis Study")
  - Sonar use: 2007–2019; geographic coordinates (SPORTS)
  - Strandings: Beaked whale single; geographic coordinates
- NOAA National Stranding Database data
  - SOCAL, HI, MidLant
  - Last 5 years

Case study: Western Mediterranean (1/2)

Western Mediterranean

Observed coincident stranding #1



Observed coincident stranding #2

$D_{\text{Max}} = 4751$ days, $D_{\text{Sonar}} = 254$, $N_{\text{S,Obs}} = 5$

Estimate expected number of coincident strandings

$$\lambda_0 = \frac{N_{\text{NullS}}}{D_{\text{No Sonar}}} \approx \frac{3}{4229} \approx 0.00069$$

$$N_{\text{CoinS,Expected}} \approx \lambda_0 \cdot D_{\text{Sonar Effec}} \approx \boxed{0.2800}$$

Number of *observed* of coincident strandings, $N_{\text{CoinS,Obs}}$

Number of ***expected*** coincident strandings under null hypothesis, $N_{\text{CoinS,Exp}}$

	1	2	3	4	5	6	7
0.1	0.09516			✓	✓	✓	✓
0.2	0.181269	**0.0175231**	✓	✓	✓	✓	✓
0.3	182	**0.0369363**	✓	✓	✓	✓	✓
0.4	0.3296				✓	✓	✓
0.5	0.393						
0.6	0.451						
0.7	0.503						
0.8	0.550671	0.191208	0.0474226	■	✓	✓	✓
0.9						✓	✓
1.0	0.632121	0.264241	0.0803014	0.0189882	■	✓	✓

$\boxed{0.2800}$

$0.018 < \alpha < 0.037$

Poisson Mean Lookup Table

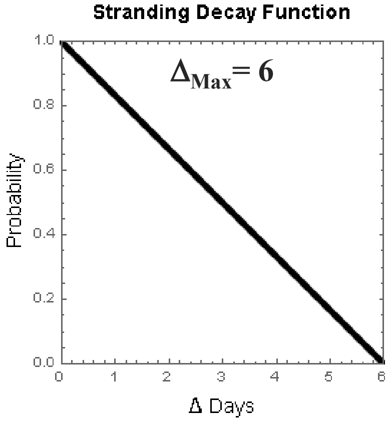
The conventional analysis suggests that the null hypothesis can be rejected!

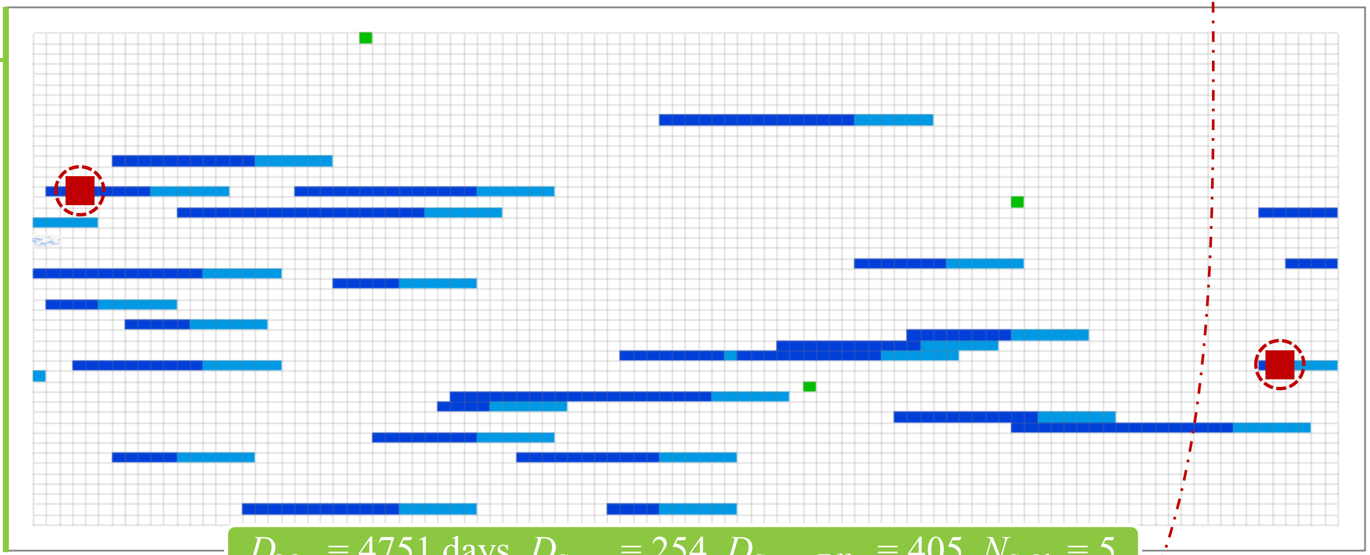
But is this really the case?

Case study: Western Mediterranean (2/2)

Western Mediterranean → Cannot Reject H0

Account for uncertainty in the *observed* vs. *actual* stranding dates

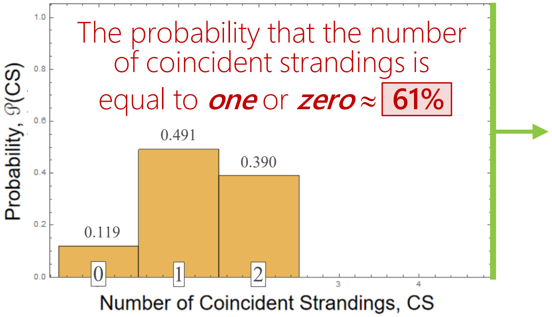




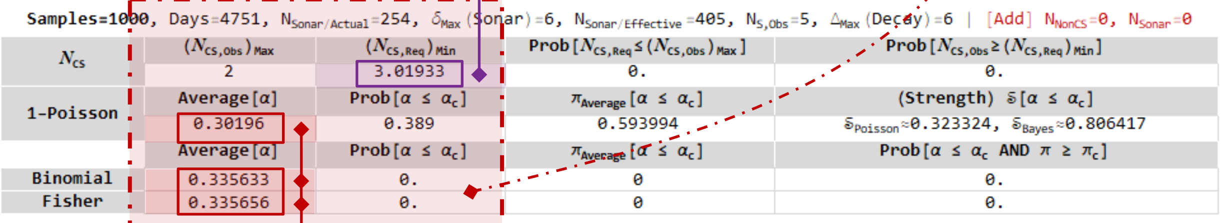
$D_{Max} = 4751$ days, $D_{Sonar} = 254$, $D_{Sonar\ Effec} = 405$, $N_{S,Obs} = 5$

Ave number of CS ≈ $\boxed{1.27}$

Observed CS < Required CS



The probability that the number of coincident strandings is equal to ***one*** or ***zero*** ≈ $\boxed{61\%}$



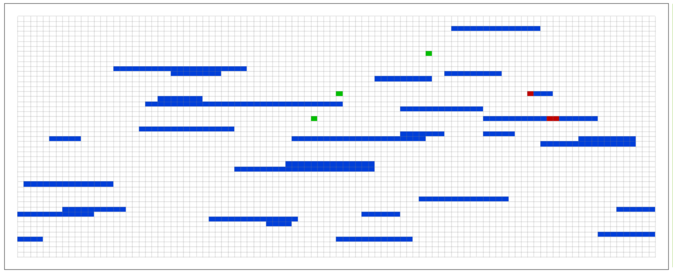
Samples=1000, Days=4751, $N_{Sonar/Actual}$=254, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=405, $N_{S,Obs}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob[$N_{CS,Req} \le (N_{CS,Obs})_{Max}$]	Prob[$N_{CS,Obs} \ge (N_{CS,Req})_{Min}$]
	2	3.01933	0.	0.
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	(Strength) δ[α ≤ α_c]
	0.30196	0.389	0.593994	$\delta_{Poisson}$≈0.323324, δ_{Bayes}≈0.806417
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.335633	0.	0	0.
Fisher	0.335656	0.	0	0.

Poisson test, BET, and FET *all* yield $\alpha > \alpha_c = 0.05$

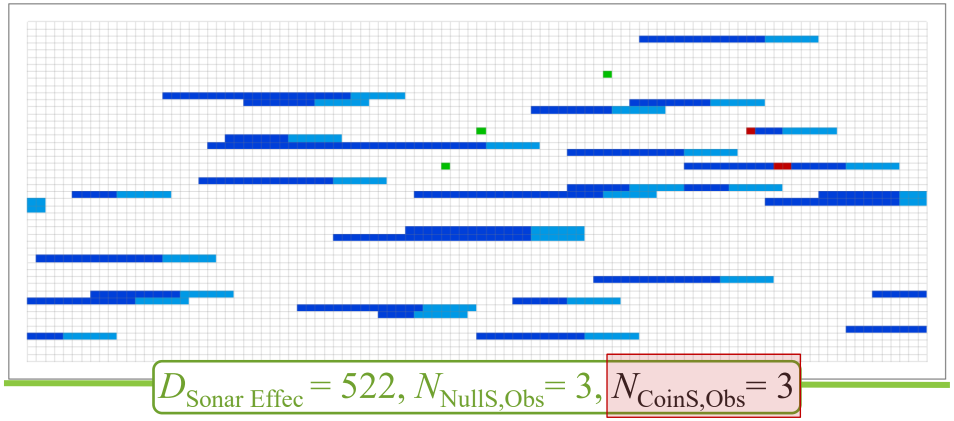
Case study: Central Mediterranean (1/3)

Central Mediterranean



$D_{\text{Max}} = 4751$ days, $D_{\text{Sonar}} = 354$, $N_{\text{S,Obs}} = 6$

$\delta_{\text{Max}} = 6$

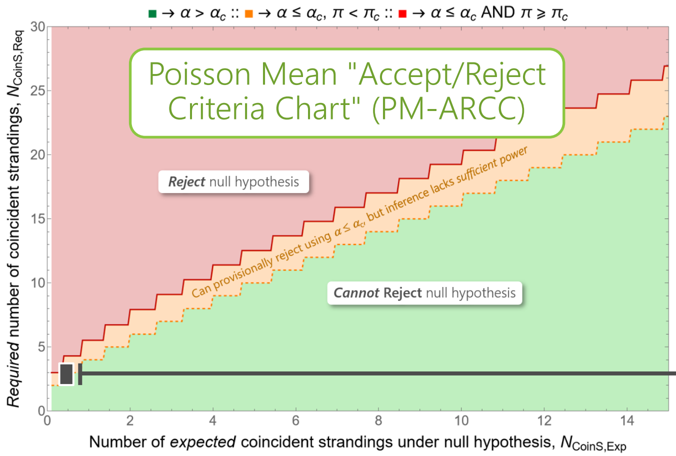


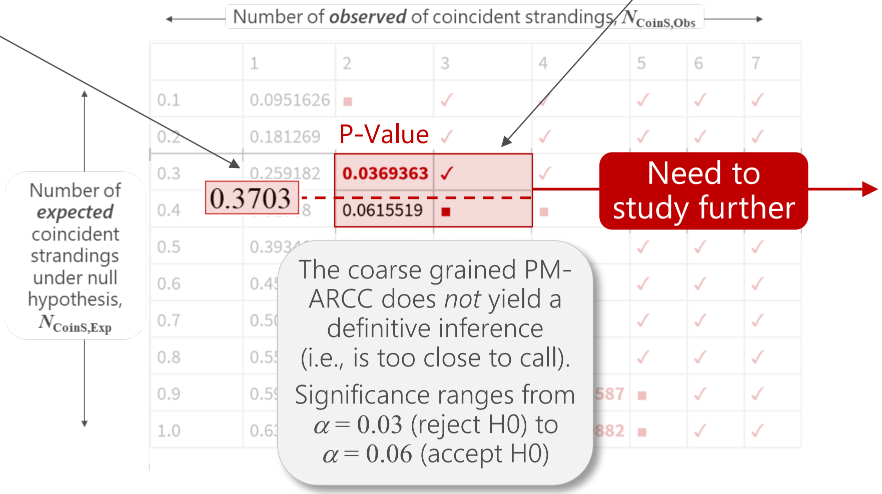
$D_{\text{Sonar Effec}} = 522$, $N_{\text{NullS,Obs}} = 3$, $N_{\text{CoinS,Obs}} = 3$

$$\lambda_0 = \frac{N_{\text{NullS}}}{D_{\text{No Sonar}}} \approx \frac{3}{4229} \approx 0.000709$$

Estimate expected number of coincident strandings

$$N_{\text{CoinS,Expected}} \approx \lambda_0 \cdot D_{\text{Sonar Effec}} \approx \boxed{0.3703}$$





Number of *observed* of coincident strandings, $N_{\text{CoinS,Obs}}$

Number of *expected* coincident strandings under null hypothesis, $N_{\text{CoinS,Exp}}$

	1	2	3	4	5	6	7
0.1	0.0951626	■	✓	✓	✓	✓	✓
0.2	0.181269		✓	✓	✓	✓	✓
0.3	0.259182	0.0369363	✓	✓			
0.4	8	0.0615519	■	■			
0.5	0.393				✓	✓	✓
0.6	0.45				✓	✓	✓
0.7	0.50				✓	✓	✓
0.8	0.55				✓	✓	✓
0.9	0.59			587	■	✓	✓
1.0	0.63			882	■	✓	✓

P-Value

0.3703

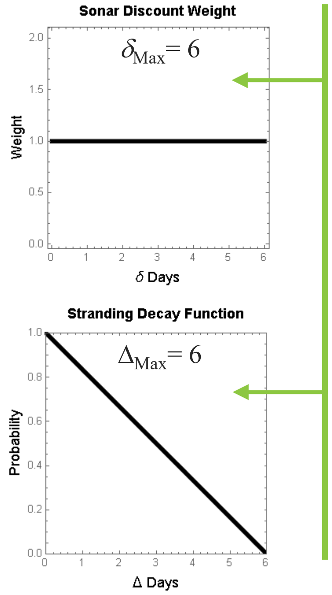
Need to study further

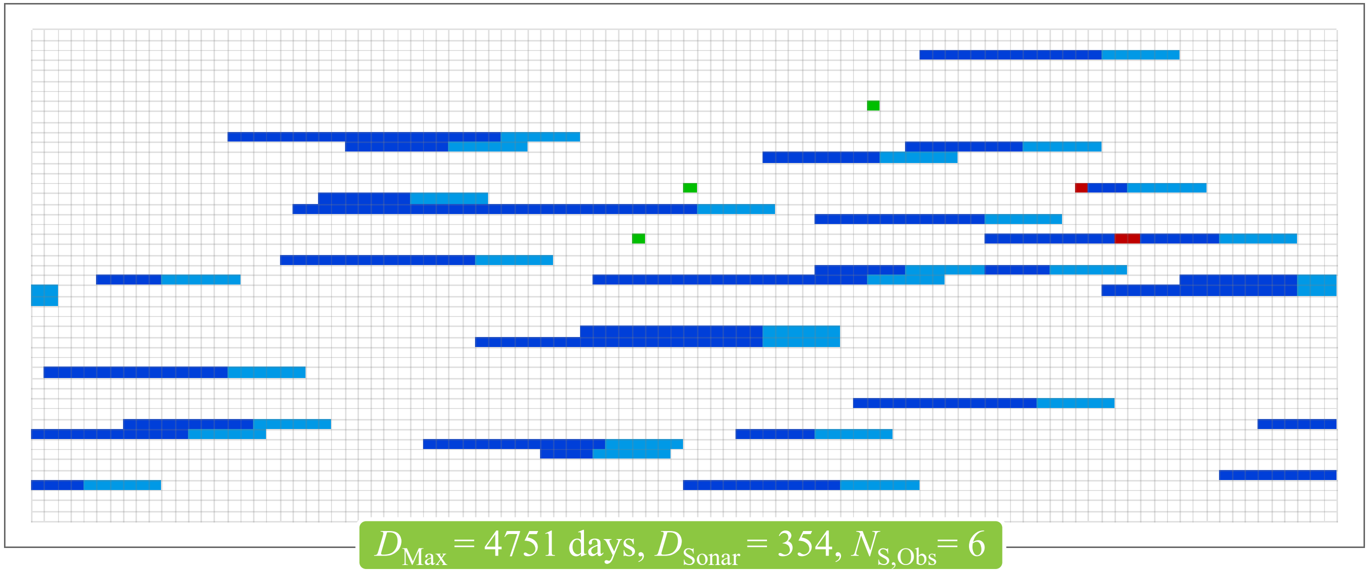
The coarse grained PM-ARCC does *not* yield a definitive inference (i.e., is too close to call).

Significance ranges from $\alpha = 0.03$ (reject H0) to $\alpha = 0.06$ (accept H0)

Case study: Central Mediterranean (2/3)

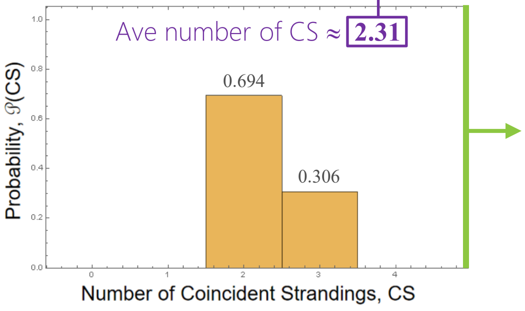
Central Mediterranean



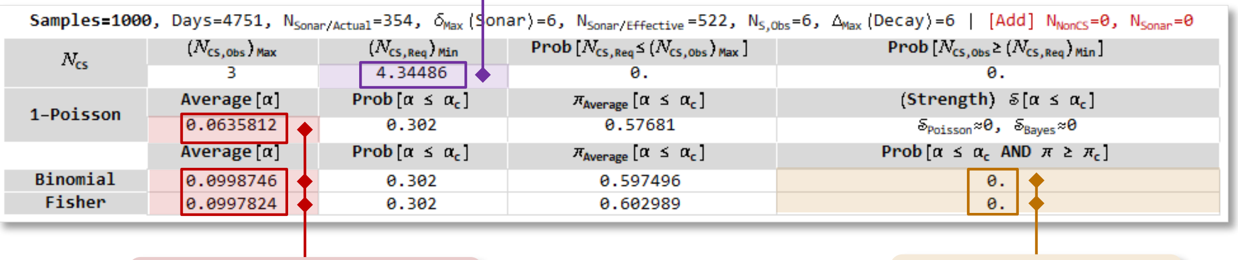


$D_{\text{Max}} = 4751$ days, $D_{\text{Sonar}} = 354$, $N_{\text{S,Obs}} = 6$

(a)



Observed CS < *Required* CS



Samples=1000, Days=4751, $N_{\text{Sonar/Actual}}$=354, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=522, $N_{\text{S,Obs}}$=6, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

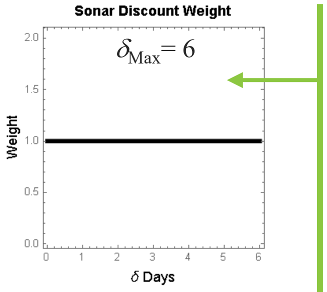
N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob$[N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}]$	Prob$[N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}]$
	3	4.34486	0.	0.
1-Poisson	Average[α]	Prob[α ≤ α_c]	π_{Average}[α ≤ α_c]	(Strength) $\mathcal{S}$[α ≤ α_c]
	0.0635812	0.302	0.57681	$\mathcal{S}_{\text{Poisson}}$≈0, $\mathcal{S}_{\text{Bayes}}$≈0
	Average[α]	Prob[α ≤ α_c]	π_{Average}[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.0998746	0.302	0.597496	0.
Fisher	0.0997824	0.302	0.602989	0.

(b) Poisson test, BET, and FET all yield $\alpha > \alpha_c = 0.05$

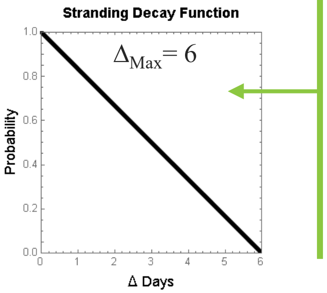
(c) No scenarios satisfy both Type I and II tests

Case study: Central Mediterranean (3/3)

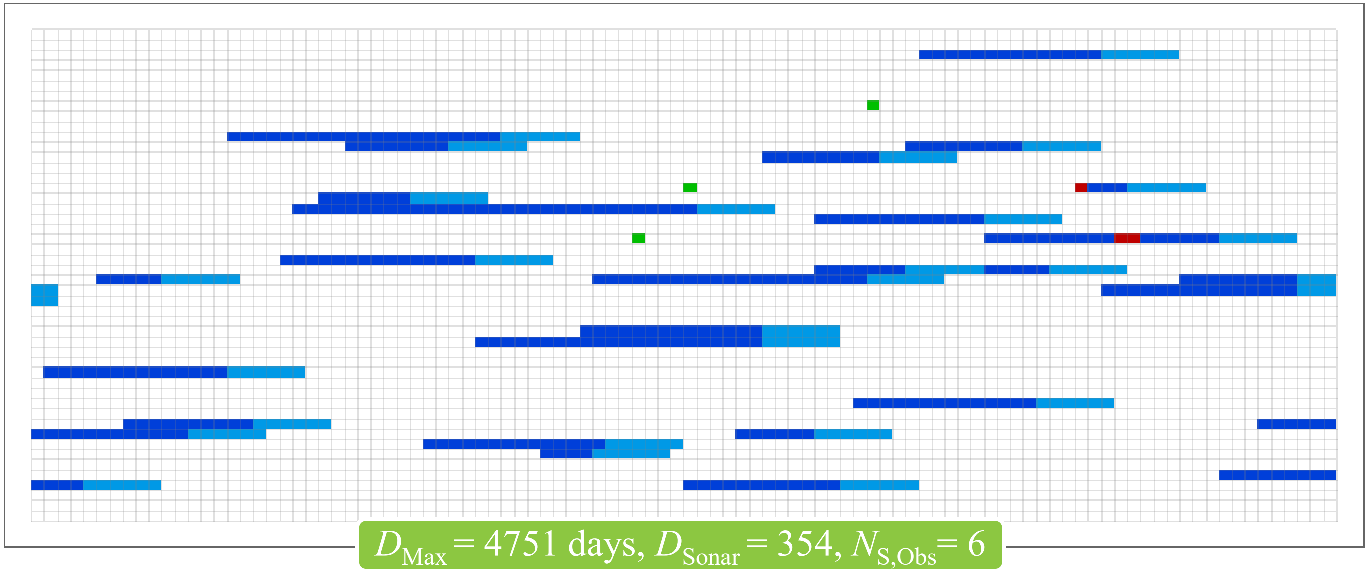
Central Mediterranean → Cannot Reject H0

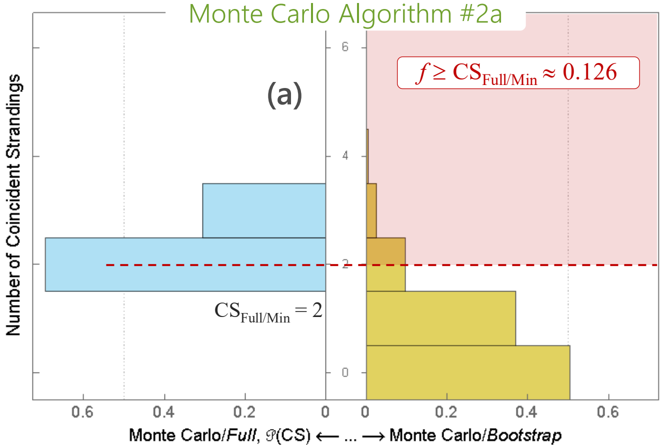

Sonar Discount Weight

$\delta_{\mathrm{Max}} = 6$


Stranding Decay Function

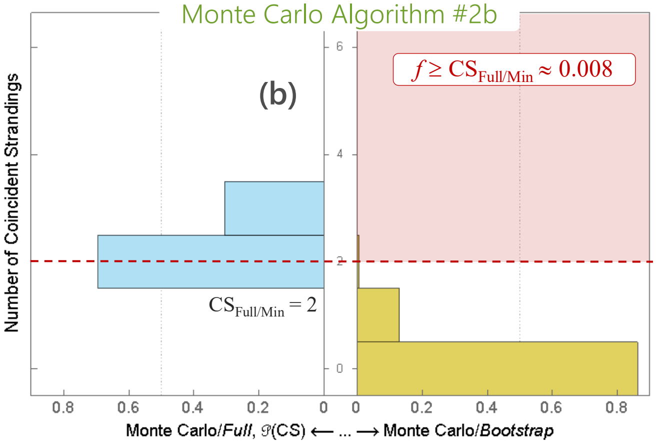
$\Delta_{\mathrm{Max}} = 6$


$D_{\mathrm{Max}} = 4751$ days, $D_{\mathrm{Sonar}} = 354$, $N_{\mathrm{S,Obs}} = 6$


Monte Carlo Algorithm #2a

(a)

$f \geq \mathrm{CS}_{\mathrm{Full/Min}} \approx 0.126$


Monte Carlo Algorithm #2b

(b)

$f \geq \mathrm{CS}_{\mathrm{Full/Min}} \approx 0.008$

Pulling everything together (1/2)

Towards a Stranding Correlation Analysis Playbook (SCAP)

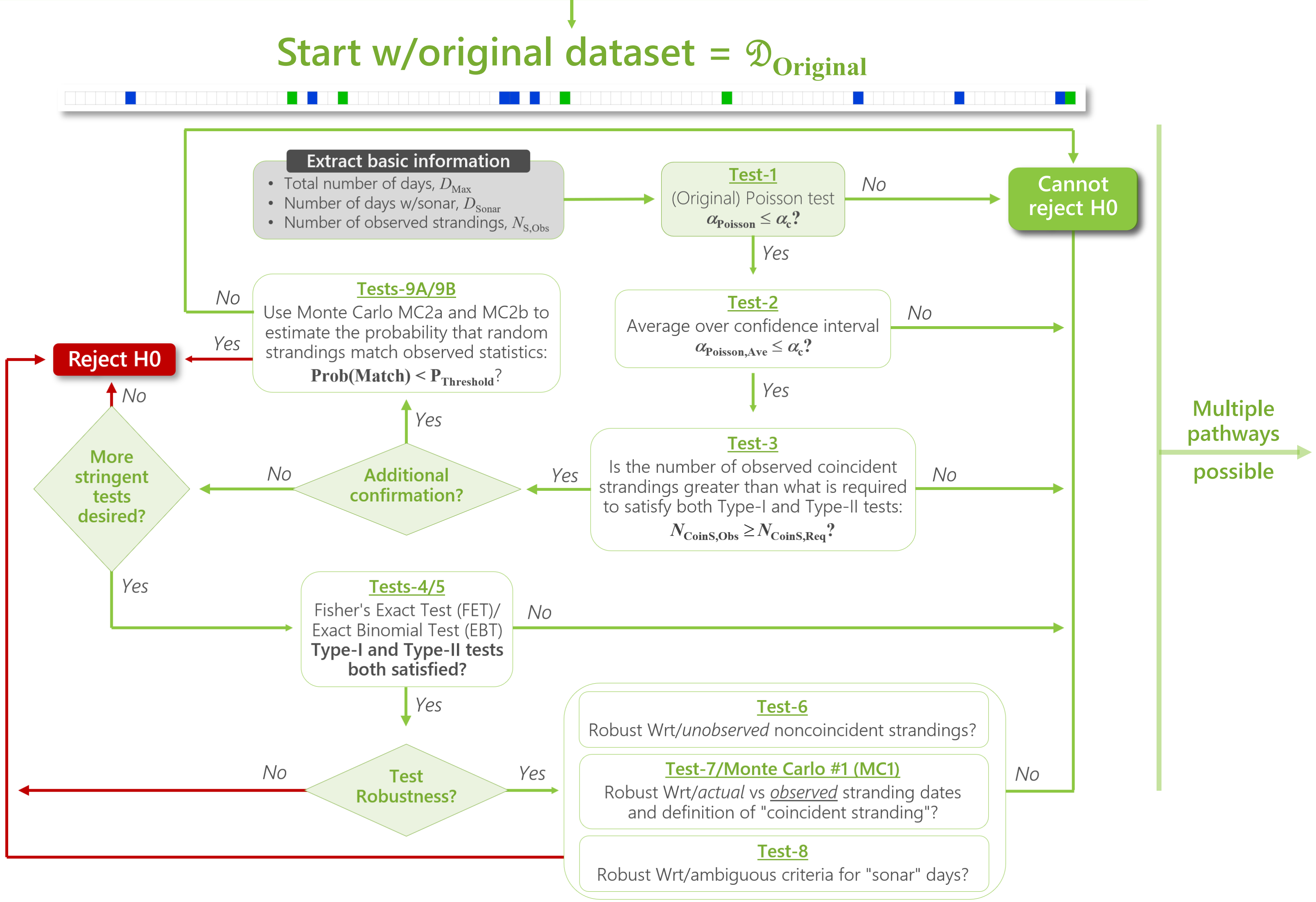
Summary of statistical tests and analysis tools

Possible refinements to account for uncertainties

- **Test 1:** (Original) single-means Poisson test, α_{Poisson}
  - Test 0/Strength: Poisson or Bayesian estimate
- **Test 2:** Averaged over all coincident rates in CI, $\alpha_{\text{Poisson,Ave}}$
- **Test 3:** Minimum # of CS required to satisfy both Type I and Type II tests
  - Use (as reference) the Poisson Mean "Accept/Reject Criteria Chart" (PM-ARCC)
- **Test 4:** Fisher's exact test
- **Test 5:** Exact binomial test
- **Test 6:** How robust is H0 rejection to unobserved non-coincident strandings?
  - Determine range of expected coincident strandings entailed by the presence of unobserved noncoincident strandings, for which α remains $\leq \alpha_c$
- **Test 7/Monte Carlo #1 (MCS 1)**
  - How robust is H0 rejection to uncertainties regarding actual versus observed stranding date and regarding labeling a given stranding event as "coincident" with sonar?
  - 4-by-4 matrix of test statistics
- **Test 8:** How robust is the rejection of the null hypothesis to ambiguous or inconsistent criteria for including a specific number of sonar days in dataset?
- **Tests 9A/9B:** What is the probability that a set of randomly assigned stranding dates (for a fixed number of total strandings) yields the observed number of coincident strandings?
  - **Test 9A/Monte Carlo #2a (MCS 2a):** Randomly distribute a fixed number of total strandings
  - **Test 9B/Monte Carlo #2b (MCS 2b):** Sample over random datasets using estimated distribution of null stranding rates

Pulling everything together (2/2)

Towards a Stranding Correlation Analysis Playbook (SCAP)



Start w/original dataset = $\mathfrak{D}_{\textbf{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\mathrm{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\mathrm{Poisson}} \le \alpha_c$?

No

Cannot reject H0

Yes

Test-2
Average over confidence interval
$\alpha_{\mathrm{Poisson,Ave}} \le \alpha_c$?

No

Yes

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\mathrm{CoinS,Obs}} \ge N_{\mathrm{CoinS,Req}}$?

No

Yes

Additional confirmation?

Yes

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\mathbf{Prob(Match)} < \mathbf{P}_{\mathrm{Threshold}}$?

No

Yes

Reject H0

No

More stringent tests desired?

No

Yes

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

No

Yes

Test Robustness?

No

Yes

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

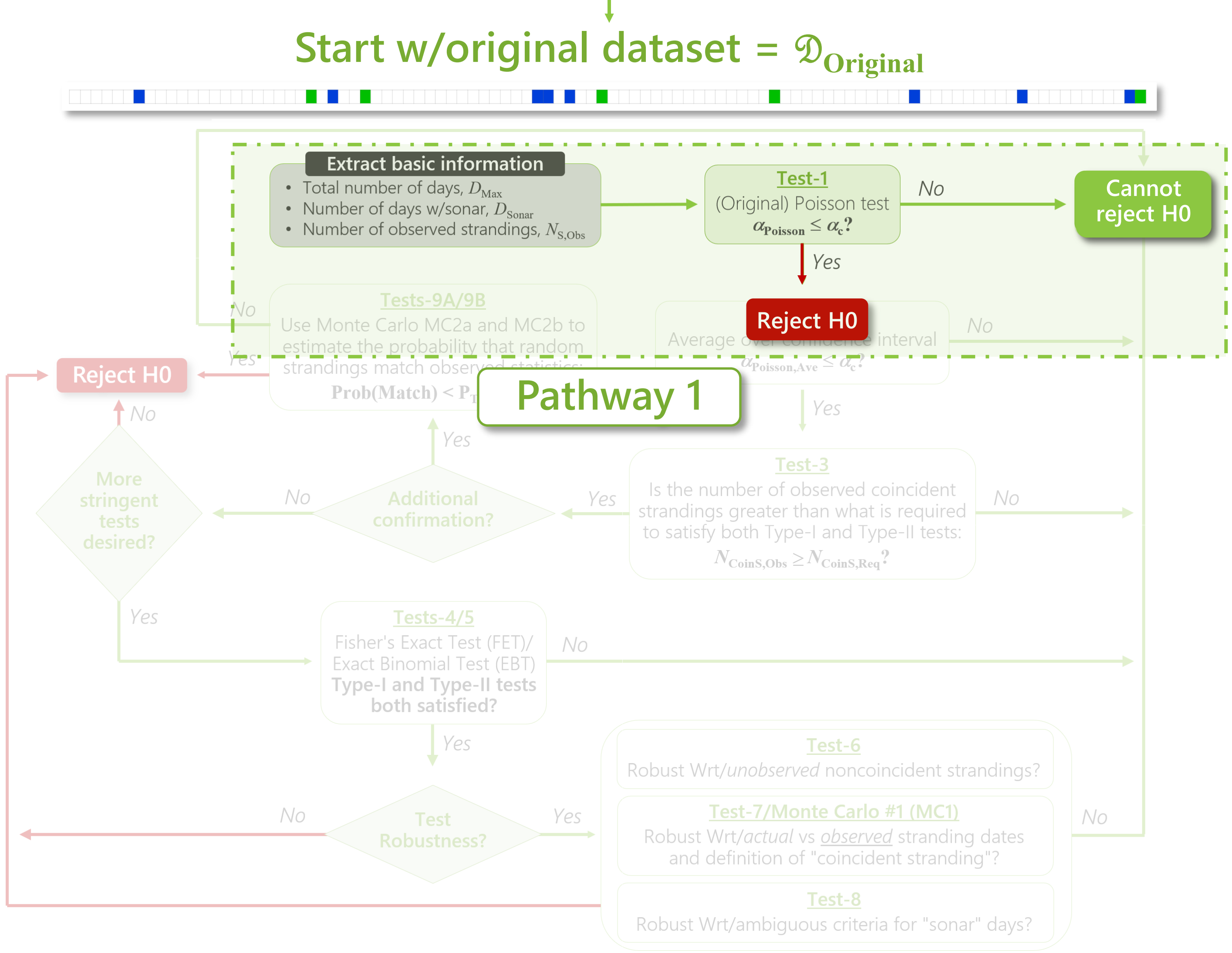
Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

No

Multiple pathways possible

SCAP: *pathway 1*

Towards a Stranding Correlation Analysis Playbook (SCAP)



Start w/original dataset = $\mathfrak{D}_{\mathbf{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\mathrm{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\mathrm{Poisson}} \le \alpha_c?$

No

Cannot reject H0

Yes

Reject H0

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\mathrm{Prob(Match)} < \mathrm{P_T}$

No

Yes

Average over confidence interval
$\alpha_{\mathrm{Poisson,Ave}} \le \alpha_c?$

No

Pathway 1

Reject H0

No

Yes

Yes

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\mathrm{CoinS,Obs}} \ge N_{\mathrm{CoinS,Req}}?$

No

More stringent tests desired?

No

Additional confirmation?

Yes

Yes

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

No

Yes

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

No

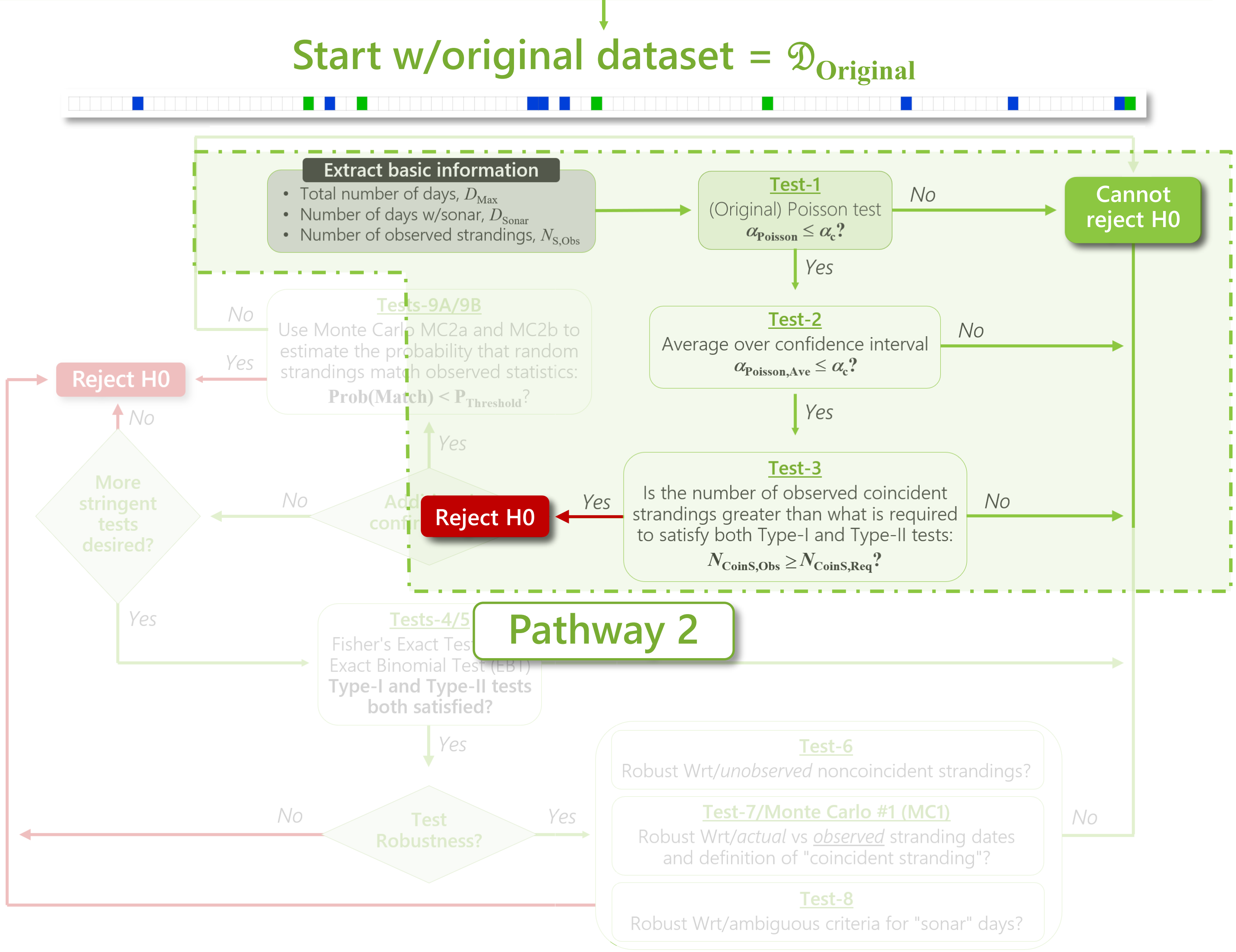
Test Robustness?

Yes

No

SCAP: *pathway 2*

Towards a Stranding Correlation Analysis Playbook (SCAP)



Start w/original dataset = $\mathcal{D}_{\text{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\text{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\text{Poisson}} \le \alpha_c$?

No → Cannot reject H0

Yes

Test-2
Average over confidence interval
$\alpha_{\text{Poisson,Ave}} \le \alpha_c$?

No

Yes

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\text{CoinS,Obs}} \ge N_{\text{CoinS,Req}}$?

Yes → Reject H0

No

Pathway 2

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\text{Prob(Match)} < \text{P}_{\text{Threshold}}$?

No

Yes → Reject H0

No

More stringent tests desired?

Yes

No

Add confi

Yes

Tests-4/5
Fisher's Exact Test
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Yes

Test Robustness?

No

Yes

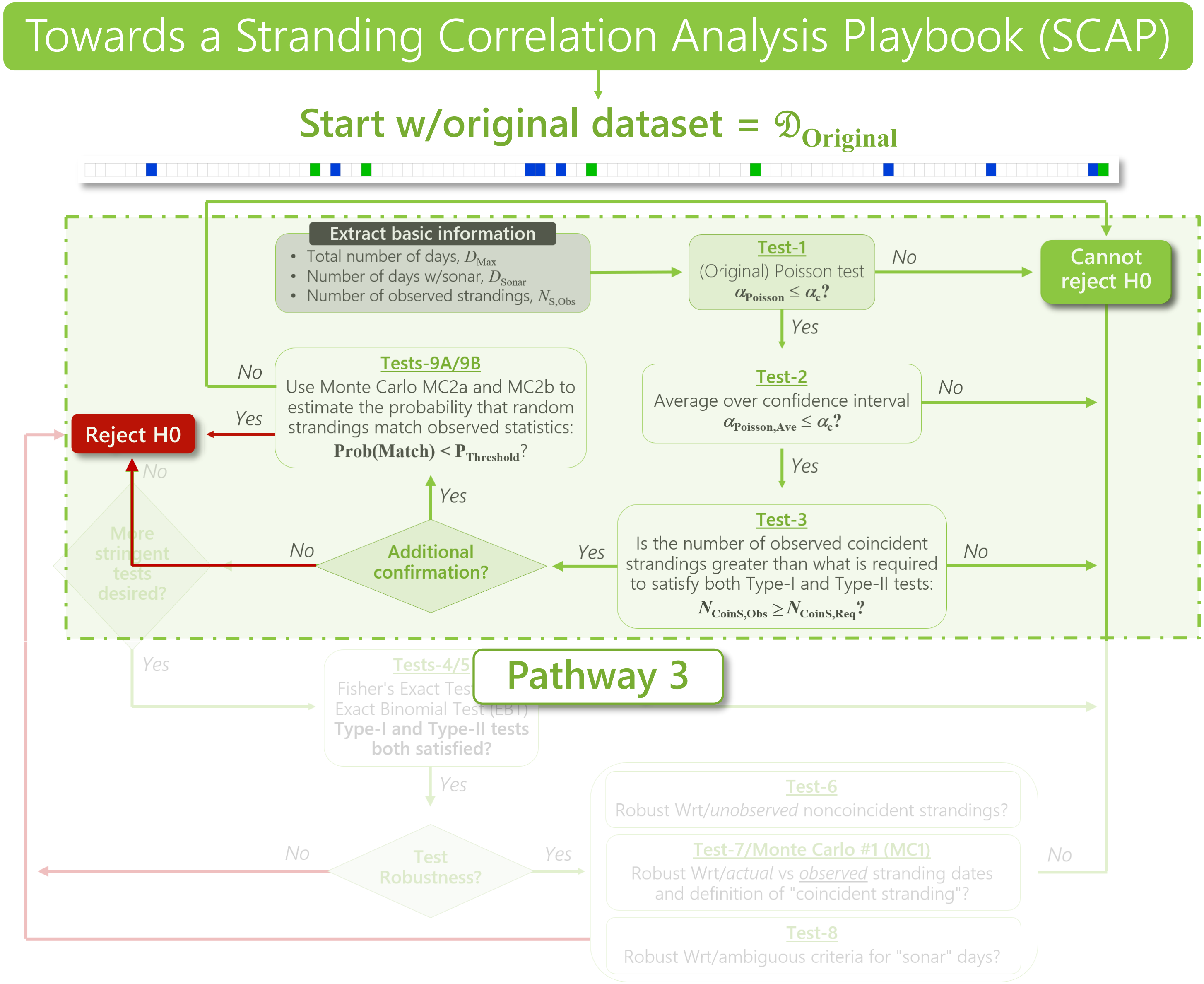
Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

No

SCAP: *pathway 3*



Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathfrak{D}_{\mathbf{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\text{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\text{Poisson}} \le \alpha_c$?

No → Cannot reject H0

Yes ↓

Test-2
Average over confidence interval
$\alpha_{\text{Poisson,Ave}} \le \alpha_c$?

No →

Yes ↓

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\text{CoinS,Obs}} \ge N_{\text{CoinS,Req}}$?

No →

Yes ←

Additional confirmation?

Yes ↑

No → Reject H0

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\mathbf{Prob(Match)} < \mathbf{P}_{\mathbf{Threshold}}$?

Yes → Reject H0

No → Cannot reject H0

Reject H0

No

More stringent tests desired?

Yes

Tests-4/5
Fisher's Exact Test
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

Yes

Test Robustness?

No

Yes

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

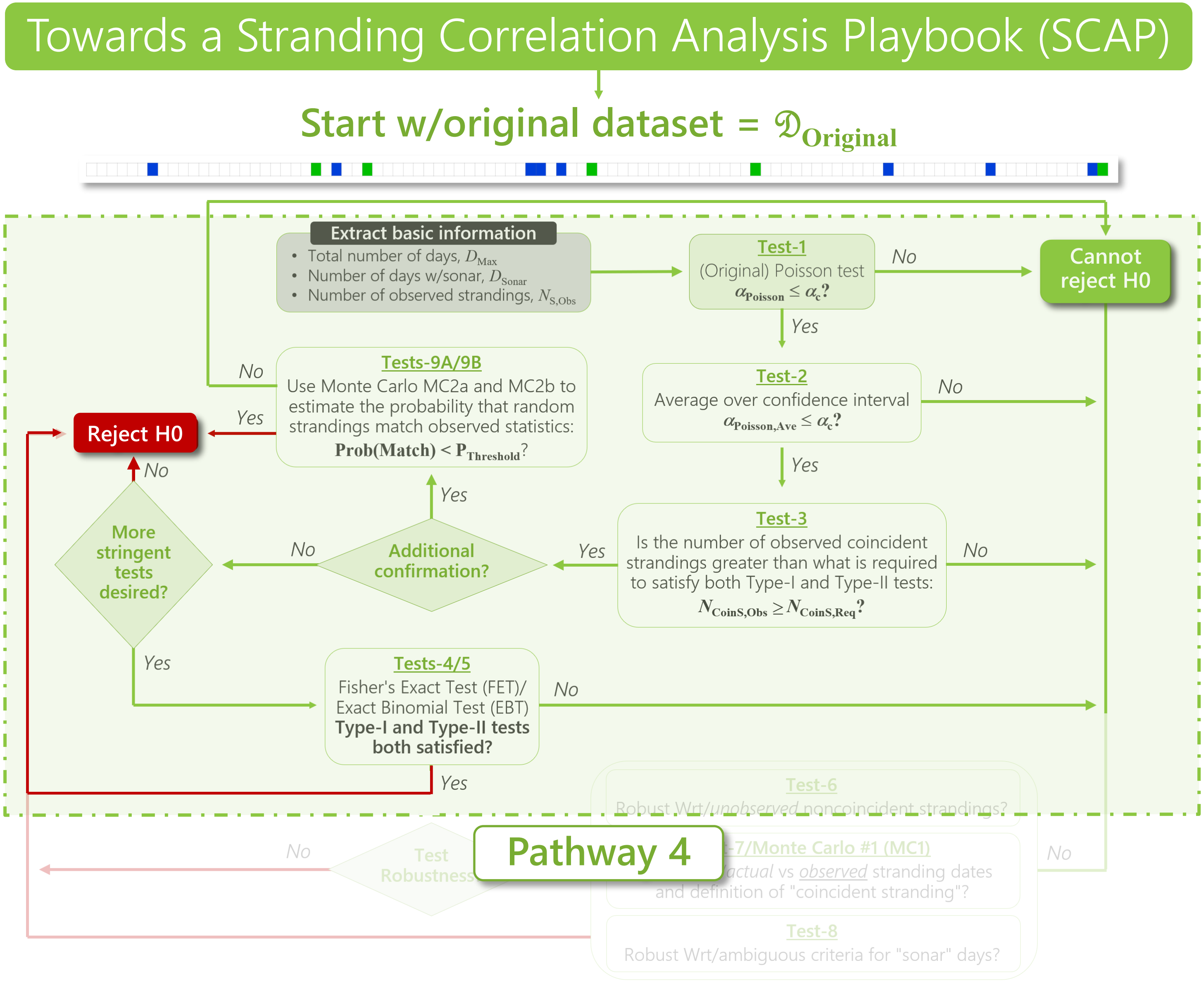
Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

No

Pathway 3

SCAP: *pathway 4*



Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathfrak{D}_{\textbf{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\text{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\text{Poisson}} \le \alpha_c$?

No

Cannot reject H0

Yes

Test-2
Average over confidence interval
$\alpha_{\text{Poisson,Ave}} \le \alpha_c$?

No

Yes

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\text{CoinS,Obs}} \ge N_{\text{CoinS,Req}}$?

No

Yes

Additional confirmation?

Yes

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\text{Prob(Match)} < \text{P}_{\text{Threshold}}$?

No

Yes

Reject H0

No

More stringent tests desired?

No

Yes

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

No

Yes

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

-7/Monte Carlo #1 (MC1)
actual vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

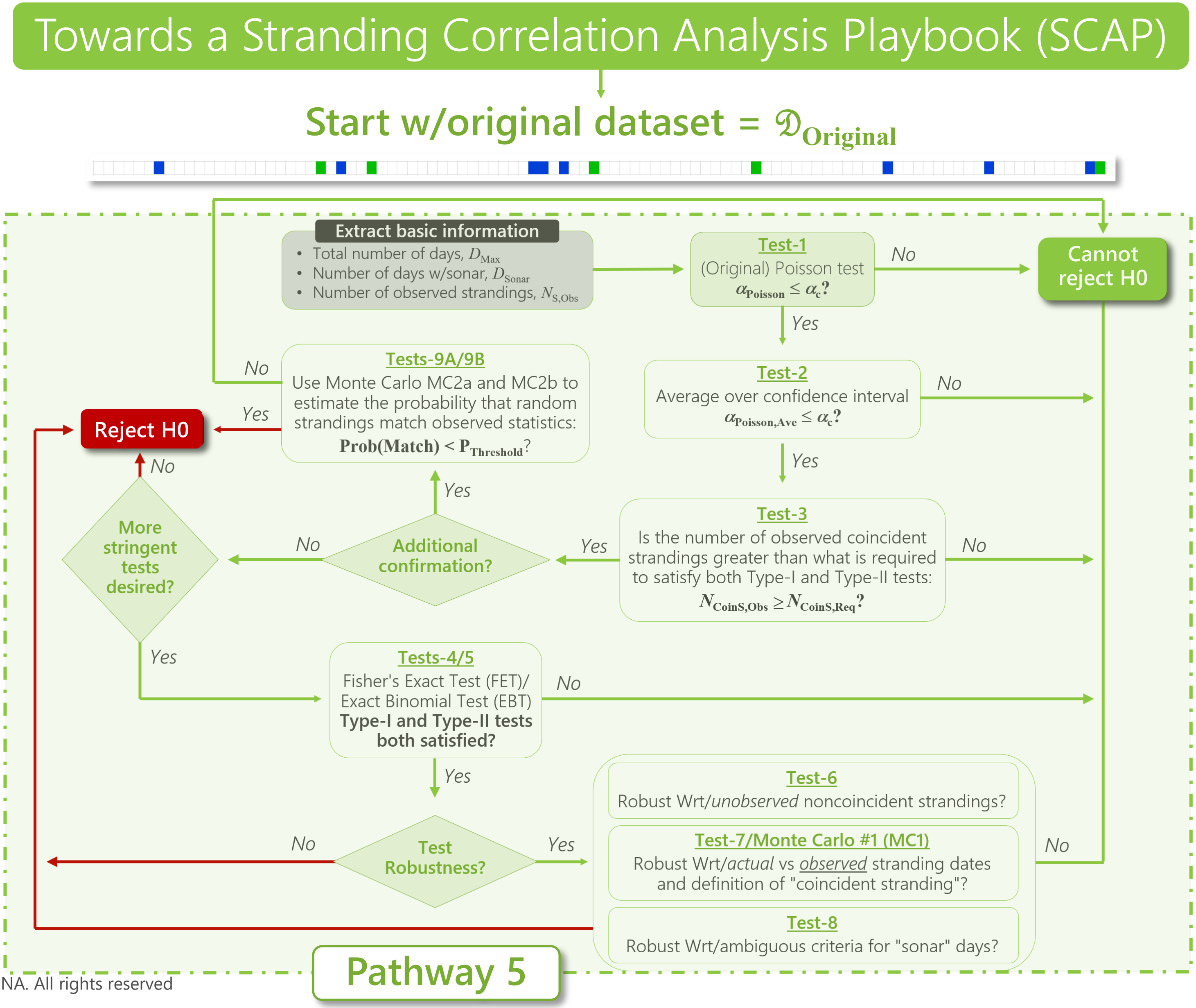
Test Robustness

No

No

Pathway 4

SCAP: *pathway 5*



Towards a Stranding Correlation Analysis Playbook (SCAP)

Start w/original dataset = $\mathcal{D}_{\mathbf{Original}}$

Extract basic information

- Total number of days, D_{Max}
- Number of days w/sonar, D_{Sonar}
- Number of observed strandings, $N_{\mathrm{S,Obs}}$

Test-1
(Original) Poisson test
$\alpha_{\mathbf{Poisson}} \leq \alpha_c$?

No → Cannot reject H0

Yes

Test-2
Average over confidence interval
$\alpha_{\mathbf{Poisson,Ave}} \leq \alpha_c$?

No

Yes

Test-3
Is the number of observed coincident strandings greater than what is required to satisfy both Type-I and Type-II tests:
$N_{\mathbf{CoinS,Obs}} \geq N_{\mathbf{CoinS,Req}}$?

No

Yes

Additional confirmation?

Yes

Tests-9A/9B
Use Monte Carlo MC2a and MC2b to estimate the probability that random strandings match observed statistics:
$\mathbf{Prob(Match)} < \mathbf{P}_{\mathbf{Threshold}}$?

Yes → Reject H0

No

No

More stringent tests desired?

No → Reject H0

Yes

Tests-4/5
Fisher's Exact Test (FET)/
Exact Binomial Test (EBT)
Type-I and Type-II tests both satisfied?

No

Yes

Test Robustness?

No

Yes

Test-6
Robust Wrt/*unobserved* noncoincident strandings?

Test-7/Monte Carlo #1 (MC1)
Robust Wrt/*actual* vs *observed* stranding dates and definition of "coincident stranding"?

Test-8
Robust Wrt/ambiguous criteria for "sonar" days?

No

Pathway 5

Recommendations

Strike a balance between methodological minutiae and expediency

1. Given the limitations of statistical analyses of time-series in general (and the inherent ambiguities and uncertainties of sonar-stranding datasets in particular), use only the *strictest significance tests to reject the null hypothesis*
   - Use α_c = 0.03 or α_c = 0.01 rather than α_c = 0.05
2. **Do not rely on significance tests alone → add tests for *power***
   - Reject null hypothesis if the number of *observed* coincident strandings is greater than the *minimum* number of coincident strandings required to satisfy both significance and power
3. Use Monte Carlo simulation methods to determine how robust "single test" inferences (even those that use both α and π) are with respect to underlying uncertainties in data
4. Follow the general guidelines as implemented in the SCAP flowchart
   - Multiple inferential pathways are possible, subject to the requirements of individual analysts, decision-makers, and other stakeholders

Next steps

- Automate deployment of the SCAP
  - Develop stand-alone interactive decision-aid tailored to individual stakeholders (and other users with varying levels of mathematical and simulation expertise)
- Develop more robust dataset preparation methods for statistical analysis
  - Minimize loss of information due to pigeonholing three-dimensional information (two-dimensional space plus time) into a one-dimensional time-series
- Develop a stranding reconstruction toolkit to complement the use of SCAP
  - Apply traditional reconstruction analysis and visualization methodology
- Explore methods to mitigate uncertainty caused by heretofore unexplored confounding factors and other potential biases
  - Such as seasonality, seismic events, and presence of fringing reefs

References (1/4)

Statistical Tests

[1] Chen, Oliver Y. et al. Dec. 2023. “The Roles, Challenges, and Merits of the P Value.” *Patterns* 4 (12). https://www.sciencedirect.com/science/article/pii/S2666389923002702.

[2] Cheng, Philip E. et al. Apr. 2008. "Information Identities and Testing Hypotheses: Power Analysis for Contingency Tables." *Statistica Sinica* 18 (2). https://arthur.stat.sinica.edu.tw/_media/cv/2008-philip-statistica-sinica.pdf.

[3] Donges, J. F. et al. 2016. "Event Coincidence Analysis for Quantifying Statistical Interrelationships Between Event Time Series." *European Physical Journal Special Topics* 225. https://link.springer.com/article/10.1140/epjst/e2015-50233-y.

[4] Dureh, Nurin, C. Choonpradub, and P. Tongkumchum. 2015. "Comparing Tests for Association in Two-by-Two Tables with Zero Cell Counts." *Chiang Mai Journal of Science* 42 (4). https://www.thaiscience.info/Journals/Article/CMJS/10976684.pdf.

[5] Fagerland, Morten W., S. Lydersen, and P. Laake. 2017. *Statistical Analysis of Contingency Tables*. CRC Press. https://www.routledge.com/Statistical-Analysis-of-Contingency-Tables/Fagerland-Lydersen-Laake/p/book/9780367495268.

[6] Freeman, Jenny and M. Campbell. June 2007. "The Analysis of Categorical Data: Fisher’s Exact Test." *Scope*. https://www.researchgate.net/profile/Michael-Campbell-2/publication/237336173_The_analysis_of_categorical_data_Fisher's_exact_test/links/53d123560cf2a7fbb2e62513/The-analysis-of-categorical-data-Fishers-exact-test.pdf.

[7] Krishnamoorthy, K. and J. Thomson. Jan. 2004. "A More Powerful Test for Comparing Two Poisson Means." *Journal of Statistical Planning and Inference* 119 (1). https://www.sciencedirect.com/science/article/abs/pii/S0378375802004081.

References (2/4)

Statistical Tests – *Continued*

[8] Mathews, Paul. 2010. *Sample Size Calculations*. Mathews Malnar and Bailey, Inc. https://www.mmbstatistical.com/SampleSize.html.

[9] Maxwell, E. A. Feb. 2011. “Chi-Square Intervals for a Poisson Parameter - Bayes, Classical and Structural.” arXiv:1102.0822v1 [math.ST], https://arxiv.org/abs/1102.0822.

[10] Serdar, C. et al. 2021. "Sample Size, Power and Effect Size Revisited: Simplified and Practical Approaches in Pre-clinical, Clinical and Laboratory Studies." *Biochemia Medica* 31 (1). https://www.academia.edu/download/107266239/366825.pdf.

Stranding Analysis

[11] D'Amico, Angela et al. 2009. “Beaked Whale Strandings and Naval Exercises.” *Aquatic Mammals* 35 (4). https://research-portal.st-andrews.ac.uk/en/publications/beaked-whale-strandings-and-naval-exercises.

[12] Domabyl, Karen and Patricia Reslock. July 1986. *Ship Employment Histories and Their Use*. CNA. Research Memorandum 86-178.

[13] Frantzis, A. Mar. 2003. “The First Mass Stranding that Was Associated with the Use of Active Sonar (Kyparissiakos Gulf, Greece, 1996).” *Proceedings of the Workshop on Active Sonar and Cetaceans*. https://www.researchgate.net/publication/237310130_The_first_mass_stranding_that_was_associated_with_the_use_of_active_sonar_Kyparissiakos_Gulf_Greece_1996.

[14] Filadelfo, R. et al. Nov. 2005. *Sonar Use and Beaked-Whale Strandings*. CNA. D0012756.A3/Final.

References (3/4)

Stranding Analysis – *Continued*

[15] Filadelfo, R. Apr. 2006. “Reconstruction of Halalei Bay Whale Incident.” CNA. CME D0013984.A1.

[16] Filadelfo, R. et al. 2009. “Correlating Military Sonar Use with Beaked Whale Mass Strandings: What Do the Historical Data Show?” *Aquatic Mammals* 35 (4). https://research-portal.st-andrews.ac.uk/en/publications/correlating-military-sonar-use-with-beaked-whale-mass-strandings-.

[17] Filadelfo, R. et al. Apr. 2008. *Correlating Whale Strandings with Navy Exercises in Southern California*. CNA. CTRM D0017507.A4.

[18] Filadelfo R. et al. 2009. “Correlating Whale Strandings with Navy Exercises off Southern California.” *Aquatic Mammals* 35 (4). https://www.aquaticmammalsjournal.org/article/vol-35-iss-4-filadelfo-pinelis-et-al/.

[19] Foord, C. S. et al. 2019. “Cetacean Biodiversity, Spatial and Temporal Trends Based On Stranding Records (1920-2016).” *PLoS ONE* 14 (10). https:// doi.org/10.1371/journal.pone.0223712.

[20] *Marine Mammal Strandings Associated with US Navy Sonar Activities*. June 2017. Space and Naval Warfare Systems Center Pacific, San Diego.

[21] Mazzuca, L. et al. 1999. “Cetacean Mass Strandings in the Hawaiian Archipelago, 1957–1998.” *Aquatic Mammals* 25 (2). https://www.aquaticmammalsjournal.org/wp-content/uploads/2009/12/25-02_Mazzuca.pdf.

[23] Parsons, E. C. M. Sept. 2017. “Impacts of Navy Sonar on Whales and Dolphins: Now Beyond a Smoking Gun?” *Frontiers of Marine Science* 4. https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2017.00295/full.

References (4/4)

Stranding Analysis – *Continued*

[24] Prado, J. H. F. et al. 2016. “Long-Term Seasonal and Interannual Patterns of Marine Mammal Strandings in Subtropical Western South Atlantic.” *PLoS ONE* 11 (1). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146339.

[25] de Quiro, Y. Bernaldo et al. Jan. 2019. “Advances in Research on the Impacts of Anti-submarine Sonar on Beaked Whales.” *Proceedings of the Royal Society B* 286 (1895). https://royalsocietypublishing.org/doi/10.1098/rspb.2018.2533.

[26] Savage, Katharine N. et al. Mar. 2021. “Stejneger’s Beaked Whale Strandings in Alaska, 1995–2020.” *Marine Mammal Science* 37. https://www.researchgate.net/publication/349919645_Stejneger's_beaked_whale_strandings_in_Alaska_1995-2020.

[27] Simonis, Anne E. et al. Feb. 2020. “Co-occurrence of Beaked Whale Strandings and Naval Sonar in the Mariana Islands, Western Pacific.” *Proceedings of the Royal Society B* 287 (1921). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7062028/.

[28] West, K. L., C. W. Clifton, N. Hofmann, and I. Silva-Krott. N.D. *Historic Odontocete Stranding Events in the Hawaiian and Mariana Islands (1848–2023) and How Strandings Correlate with Environmental Parameters over an 18-Year Timespan*. College of Tropical Agriculture and Human Services, University of Hawai‘i.

[29] “Strandings.” University of Rhode Island. https://dosits.org/animals/effects-of-sound/potential-effects-of-sound-on-marine-mammals/strandings/.

Appendices

- **Appendix A:** Main statistical tests
- **Appendix B:** Satisfying both Type I and Type II errors
- **Appendix C:** Monte Carlo #1 output data fields
- **Appendix D:** Monte Carlo #1 notional examples
- **Appendix E:** Mitigating ambiguous/inconsistent dataset preparation
- **Appendix F:** Necropsy-dependent stranding decay functions
- **Appendix G:** Real-world datasets—case studies 3 and 4
- **Appendix H:** Mathematica functions
- **Appendix I:** Sample Mathematica analysis session

Appendix A: *Main statistical tests*

Poisson

$$\alpha_{\text{Poisson}} = \text{Probablity}\left[N_{\text{CoinS}} \geq \overbrace{N_{\text{CoinS,Exp}}\left(\lambda_0\right)}^{\lambda_0 \cdot D_{\text{Sonar Effec}} = \text{Expected \# of coincident standings assuming the } \textit{null} \text{ stranding rate}}\right] \approx \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \text{Poisson}\left[n; \mu = N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right]$$

$$\approx \sum_{n=N_{\text{CoinS,Obs}}}^{\infty} \frac{e^{-N_{\text{CoinS,Exp}}\left(\lambda_0\right)} \cdot \left[N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right]^{n}}{n!} = 1 - \sum_{n=N_{\text{CoinS,Obs}}}^{N_{\text{CoinS,Obs}}-1} \frac{e^{-N_{\text{CoinS,Exp}}\left(\lambda_0\right)} \cdot \left[N_{\text{CoinS,Exp}}\left(\lambda_0\right)\right]^{n}}{n!}$$

Exact Binomial Test

$$\alpha_{\text{Binomial}} = \sum_{n=0}^{N_{\text{NullS}}} \binom{N_{\text{NullS}} + N_{\text{CoinS}}}{n} \cdot \left(\frac{D_{\text{No Sonar}}}{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}}\right)^{n} \cdot \left(1 - \frac{D_{\text{No Sonar}}}{D_{\text{No Sonar}} + D_{\text{Sonar Effec}}}\right)^{N_{\text{NullS}} + N_{\text{CoinS}} - n}$$

$$\pi_{\text{Binomial}} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \delta\left(\alpha_{\text{Binomial}} \leq \alpha_{\text{c}}\right) \cdot \textit{Poisson}\left(n, \lambda_0 \cdot D_{\text{No Sonar}}\right) \cdot \textit{Poisson}\left(m, \lambda_{\text{CS}} \cdot D_{\text{Sonar Effec}}\right)$$

$$\pi_{\text{Binomial}} \underbrace{\approx}_{\text{Large Mean Approximation}} \Phi\left(-\infty < z < z_{\beta}\right), \text{ where } \Phi\left(-\infty < x < b\right) \equiv \int_{-\infty}^{b} \textit{Normal}\left(x; \mu = 0, \sigma = 1\right)$$

$$\text{and } z_{\beta} = \frac{\left|p_1 - p_2\right|}{\sqrt{\frac{p_1}{D_{\text{No Sonar}}} + \frac{p_2}{D_{\text{Sonar Effec}}}}} - \underbrace{z_{\alpha}}_{\textit{The value of z that corresponds to the area under standard normal distribution} = \alpha}, \; p_1 \equiv \frac{N_{\text{NullS,Obs}}}{D_{\text{No Sonar}}}, \; p_2 \equiv \frac{N_{\text{CoinS,Obs}}}{D_{\text{Sonar Effec}}}$$

Appendix A: *Main statistical tests*

Fisher's Exact Test

$$\alpha_{\text{Fisher}} = \sum_{n=0}^{N_{\text{NullS}}} \overbrace{h}^{h[x] = \textit{Hypergeometric probability distribution}}\left[n; D_{\text{No Sonar}}, D_{\text{Sonar Effec}}, N_{\text{NullS}} + N_{\text{CoinS}}\right] = \sum_{n=0}^{N_{\text{NullS}}} \frac{\begin{pmatrix} D_{\text{No Sonar}} \\ n \end{pmatrix} \cdot \begin{pmatrix} D_{\text{Sonar Effec}} \\ N_{\text{NullS}} + N_{\text{CoinS}} - n \end{pmatrix}}{\begin{pmatrix} D_{\text{No Sonar}} + D_{\text{Sonar Effec}} \\ N_{\text{NullS}} + N_{\text{CoinS}} \end{pmatrix}}$$

$$\pi_{\text{Fisher}} = \sum_{n=0}^{D_{\text{No Sonar}}} \sum_{m=0}^{D_{\text{Sonar Effec}}} \overbrace{\delta\left(\alpha_{\text{Fisher}} \le \alpha_{\text{c}}\right)}^{\delta(x) = \begin{cases} 1 \text{ if } x \text{ is } \textit{True} \\ 0 \text{ else} \end{cases}} \cdot \underbrace{f\left(n; D_{\text{No Sonar}}, \frac{N_{\text{NullS}}}{D_{\text{No Sonar}}}\right)}_{f(x;a,b) = \begin{pmatrix} a \\ x \end{pmatrix} \cdot b^{x} \cdot (1-b)^{a-x}} \cdot f\left(m; D_{\text{Sonar Effec}}, \frac{N_{\text{CoinS}}}{D_{\text{Sonar Effec}}}\right)$$

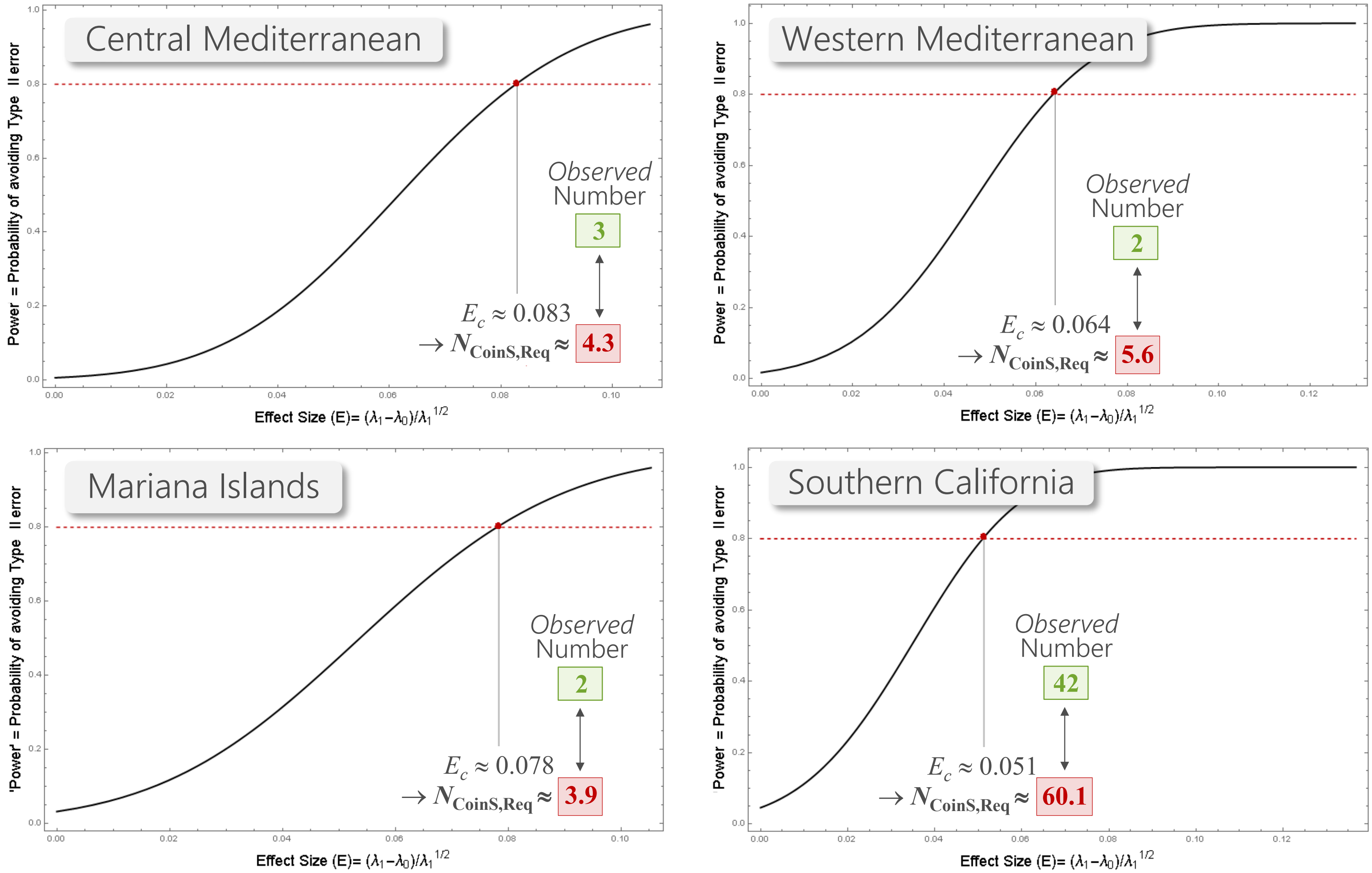
$$\Big\downarrow$$

$$\pi_{\text{Fisher}} \underbrace{\qquad\approx\qquad}_{\text{Large Mean Approximation}} \Phi\left(-\infty < z < z_{\beta}\right), \text{ where } \Phi\left(-\infty < x < b\right) \equiv \int_{-\infty}^{b} \textit{Normal}\left(x; \mu = 0, \sigma = 1\right)$$

$$\text{and } z_{\beta} = \frac{|p_1 - p_2|}{\sqrt{2\hat{p}(1-\hat{p}) / D_{\text{Max}}}} - z_{\alpha}, \ \hat{p} \equiv \frac{1}{2} \cdot (p_1 + p_2), \ p_1 \equiv \frac{N_{\text{NullS,Obs}}}{D_{\text{No Sonar}}}, p_2 \equiv \frac{N_{\text{CoinS,Obs}}}{D_{\text{Sonar Effec}}}$$

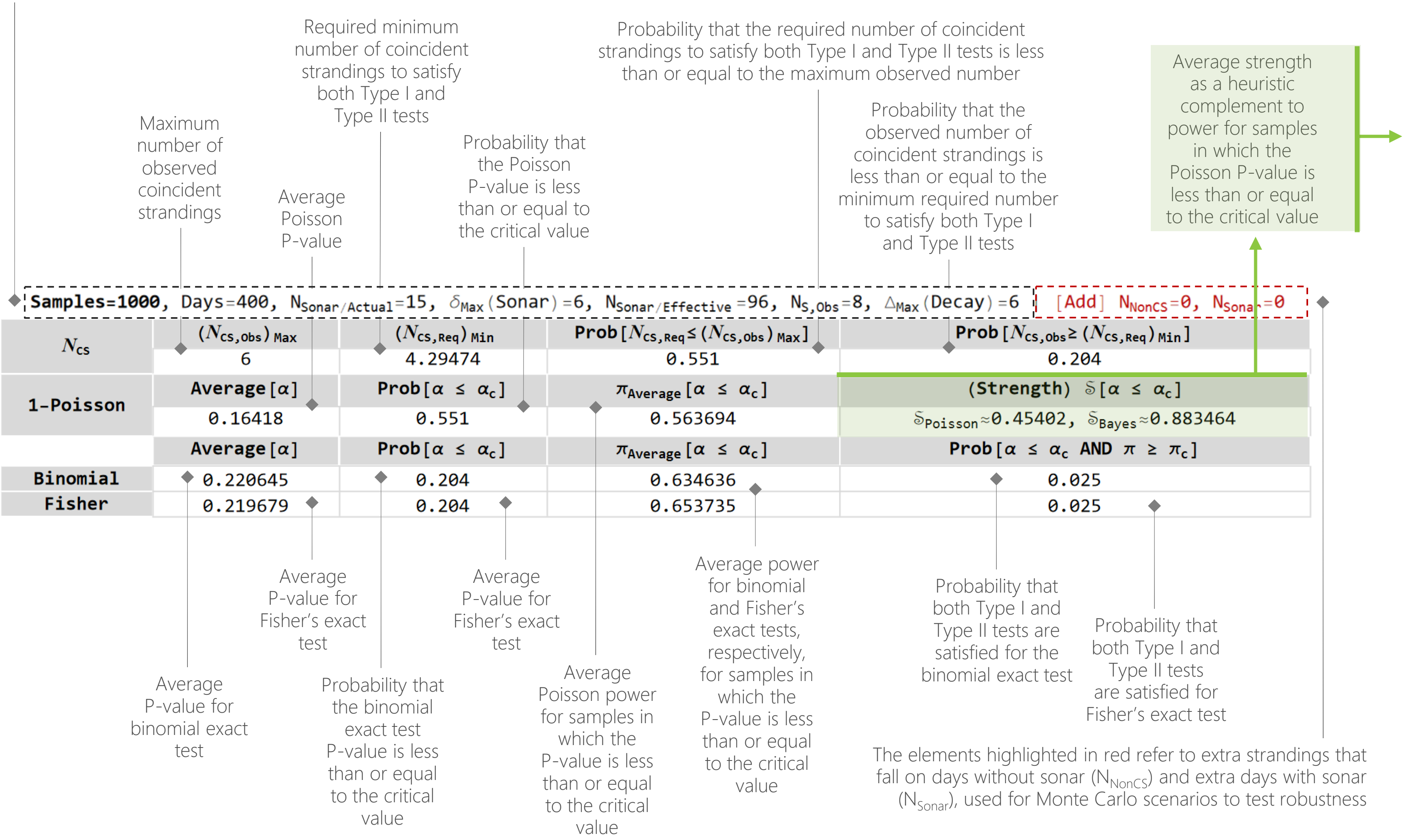
Appendix B: *Satisfying both Type I and Type II errors*

Minimum required number of coincident strandings to satisfy both Type I and II errors



Appendix C: *Monte Carlo #1 output data fields*

Basic statistics summarizing the input dataset, including the number of Monte Carlo samples, total number of days, actual ($N_{Sonar/Actual}$) and effective ($N_{sonar/Effective}$) number of days with sonar (the latter is a function of the maximum sonar decay range (δ_{Max}), and total number of observed strandings ($N_{S,Obs}$)



Maximum number of observed coincident strandings

Average Poisson P-value

Required minimum number of coincident strandings to satisfy both Type I and Type II tests

Probability that the Poisson P-value is less than or equal to the critical value

Probability that the required number of coincident strandings to satisfy both Type I and Type II tests is less than or equal to the maximum observed number

Probability that the observed number of coincident strandings is less than or equal to the minimum required number to satisfy both Type I and Type II tests

Average strength as a heuristic complement to power for samples in which the Poisson P-value is less than or equal to the critical value

Samples=1000, Days=400, $N_{Sonar/Actual}$=15, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=96, $N_{S,Obs}$=8, Δ_{Max}(Decay)=6 [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Req})_{Min}$	**Prob**$[N_{CS,Req} \le (N_{CS,Obs})_{Max}]$	**Prob**$[N_{CS,Obs} \ge (N_{CS,Req})_{Min}]$
	6	4.29474	0.551	0.204
1-Poisson	**Average**$[\alpha]$	**Prob**$[\alpha \le \alpha_c]$	$\pi_{Average}[\alpha \le \alpha_c]$	**(Strength)** $\mathcal{S}[\alpha \le \alpha_c]$
	0.16418	0.551	0.563694	$\mathcal{S}_{Poisson} \approx 0.45402$, $\mathcal{S}_{Bayes} \approx 0.883464$
	Average$[\alpha]$	**Prob**$[\alpha \le \alpha_c]$	$\pi_{Average}[\alpha \le \alpha_c]$	**Prob**$[\alpha \le \alpha_c$ **AND** $\pi \ge \pi_c]$
Binomial	0.220645	0.204	0.634636	0.025
Fisher	0.219679	0.204	0.653735	0.025

Average P-value for binomial exact test

Average P-value for Fisher's exact test

Probability that the binomial exact test P-value is less than or equal to the critical value

Average P-value for Fisher's exact test

Average Poisson power for samples in which the P-value is less than or equal to the critical value

Average power for binomial and Fisher's exact tests, respectively, for samples in which the P-value is less than or equal to the critical value

Probability that both Type I and Type II tests are satisfied for the binomial exact test

Probability that both Type I and Type II tests are satisfied for Fisher's exact test

The elements highlighted in red refer to extra strandings that fall on days without sonar (N_{NonCS}) and extra days with sonar (N_{Sonar}), used for Monte Carlo scenarios to test robustness

Appendix C: *Monte Carlo #1 output data fields*

- What if $\alpha \le \alpha_c$ but the observed number of coincident strandings, $N_{\text{CoinS,Obs}}$, is ***less*** than the required minimum, $N_{\text{CoinS,Req}}$?
  - We cannot immediately reject the null hypothesis
    - At least, not if the goal is to simultaneously satisfy both Type I and Type II errors
- However, we can still estimate the strength, $\mathbb{S}$, of rejecting the null hypothesis
- Use either the Poisson distribution or Bayesian inference to estimate the probability that the true mean of the Poisson distribution (assumed to describe the distribution of coincident strandings), μ_{True}, is greater than the required minimum, $\mu_{\text{True}} \ge N_{\text{CoinS,Req}}$

1. **Poisson:** $\mathbb{S}(\alpha \le \alpha_c \,|\, N_{\text{CoinS,Obs}}, N_{\text{CoinS,Req}}) \approx 1 - \sum_{i=0}^{N_{\text{CoinS,Req}}-1} e^{-N_{\text{CoinS,Obs}}} \cdot N_{\text{CoinS,Obs}}^{i} \,/\, i!$
2. **Bayes:** $\text{Prob}(\mu_{\text{True}} \,|\, x = N_{\text{CoinS,Obs}}) = \text{Prob}(x = N_{\text{CoinS,Obs}} \,|\, \mu_{\text{True}}) \times \text{Prob}(\mu_{\text{True}}) \,/\, \text{Prob}(x = N_{\text{CoinS,Obs}})$
   - Observed data — $x = N_{\text{CoinS,Obs}}$ (single observation)
   - Parameter of interest — μ_{True} (true mean of the distribution)
   - Likelihood function — $\text{Prob}(x = N_{\text{CoinS,Obs}} \,|\, \mu_{\text{True}}) = \mathit{Poisson}(x, \mu_{\text{True}})$
   - Marginal likelihood — $\text{Prob}(x = N_{\text{CoinS,Obs}})$
   - Prior distribution — $\text{Prob}(\mu_{\text{True}}) \approx \mathit{Gamma}(a, b)$, with $a = b = 1$
   - Posterior distribution —

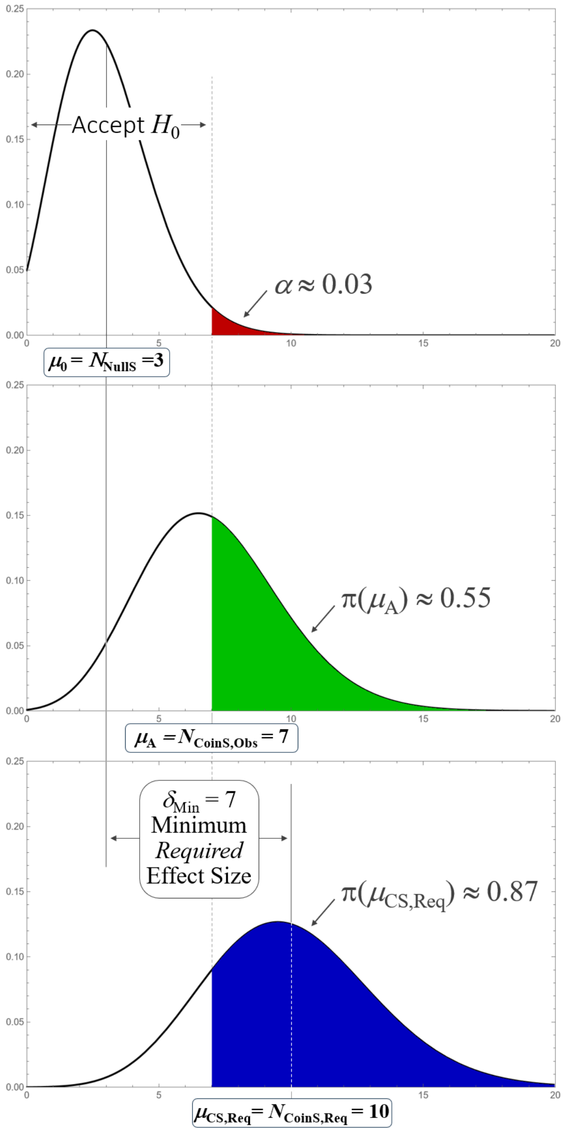
$$\text{Prob}(\mu_{\text{True}} \,|\, x = N_{\text{CoinS,Obs}}) = \text{Prob}(x = N_{\text{CoinS,Obs}} \,|\, \mu_{\text{True}}) \times \text{Prob}(\mu_{\text{True}}) \,/\, \text{Prob}(x = N_{\text{CoinS,Obs}})$$

   - For the Poisson-Gamma conjugate pair, the posterior distribution is also a *Gamma* distribution, $\text{Prob}(\mu_{\text{True}} \,|\, x = N_{\text{CoinS,Obs}}) \sim \mathit{Gamma}(a + N_{\text{CoinS,Obs}}, b + 1)$

$$\rightarrow \mathbb{S}(\alpha \le \alpha_c \,|\, N_{\text{CoinS,Obs}}, N_{\text{CoinS,Req}}) = \text{Prob}(\mu_{\text{True}} \ge N_{\text{CoinS,Req}} \,|\, x = N_{\text{CoinS,Obs}}) = \int_{\mu = N_{\text{CoinS,Req}}}^{\infty} \text{Prob}(\mu_{\text{True}} \,|\, x = N_{\text{CoinS,Obs}})\, \mathrm{d}\mu$$

$$\boxed{\mathbb{S}(\alpha \le \alpha_c \,|\, N_{\text{CoinS,Obs}}, N_{\text{CoinS,Req}}) = 1 - \text{CDF}[\mathit{Gamma}(N_{\text{CoinS,Req}};\, a + N_{\text{CoinS,Obs}}, b + 1)]}$$

Appendix C: *Monte Carlo #1 output data fields*



- *Expected* # coincident strandings $= N_{\mathrm{NullS}} = \mu_0 = 3$
  - $\alpha \approx 0.03 \rightarrow$ *satisfies* Type I test

- *Actual* (observed) number of coincident strandings, $N_{\mathrm{CoinS,Obs}} = \mu_A = 7 \rightarrow \pi(\alpha) \approx 0.55$ does *not* satisfy Type II test

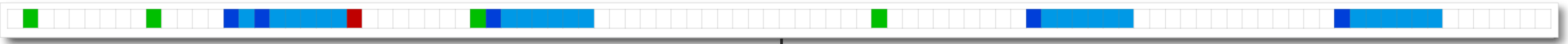
The probability that the true mean of the coincident strandings distribution is at least as large as the minimum number of coincident strandings required to satisfy both Type I and Type II statistics tests, $N_{\mathrm{CoinS,Req}}$

- $\mathbb{S}_{Poisson}(\alpha \leq \alpha_c) \approx 0.17,\ \mathbb{S}_{Bayes}(\alpha \leq \alpha_c) \approx 0.87$

Appendix D: *Monte Carlo #1 notional examples*

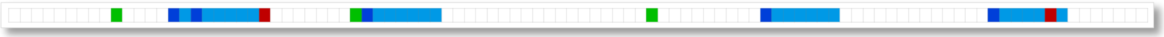
Monte Carlo Algorithm #1: Estimate probability distribution of *coincident* strandings

Start w/ $\mathfrak{D}_{\text{Original}}$



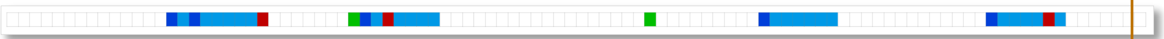
$\delta_{\text{Max}} = 6, \Delta_{\text{Max}} = 6$

N_{CoinS}	N_{NullS}
2	3
3	2
4	1



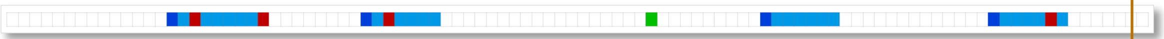
Samples=1000, Days=100, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$ =30, $N_{\text{S,Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})$Max	$(N_{\text{CS,Req}})$Min	Prob[$N_{\text{CS,Req}} \leq (N_{\text{CS,Obs}})$Max]	Prob[$N_{\text{CS,Obs}} \geq (N_{\text{CS,Req}})$Min]
	2	5.65714	0.	0.
1–Poisson	Average[α]	Prob[$\alpha \leq \alpha_c$]	$\pi_{\text{Average}}[\alpha \leq \alpha_c]$	(Strength) $\mathbb{S}[\alpha \leq \alpha_c]$
	0.388438	0.	0	$\mathbb{S}_{\text{Poisson}}\approx 0$, $\mathbb{S}_{\text{Bayes}}\approx 0$
	Average[α]	Prob[$\alpha \leq \alpha_c$]	$\pi_{\text{Average}}[\alpha \leq \alpha_c]$	Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]
Binomial	0.487987	0.	0	0.
Fisher	0.490326	0.	0	0.



Samples=1000, Days=100, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$ =30, $N_{\text{S,Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})$Max	$(N_{\text{CS,Req}})$Min	Prob[$N_{\text{CS,Req}} \leq (N_{\text{CS,Obs}})$Max]	Prob[$N_{\text{CS,Obs}} \geq (N_{\text{CS,Req}})$Min]
	3	5.65714	0.	0.
1–Poisson	Average[α]	Prob[$\alpha \leq \alpha_c$]	$\pi_{\text{Average}}[\alpha \leq \alpha_c]$	(Strength) $\mathbb{S}[\alpha \leq \alpha_c]$
	0.231005	0.	0	$\mathbb{S}_{\text{Poisson}}\approx 0$, $\mathbb{S}_{\text{Bayes}}\approx 0$
	Average[α]	Prob[$\alpha \leq \alpha_c$]	$\pi_{\text{Average}}[\alpha \leq \alpha_c]$	Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]
Binomial	0.3341	0.	0	0.
Fisher	0.332737	0.	0	0.



Samples=1000, Days=100, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$ =30, $N_{\text{S,Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

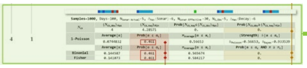
N_{CS}	$(N_{\text{CS,Obs}})$Max	$(N_{\text{CS,Req}})$Min	Prob[$N_{\text{CS,Req}} \leq (N_{\text{CS,Obs}})$Max]	Prob[$N_{\text{CS,Obs}} \geq (N_{\text{CS,Req}})$Min]
	4	4.28571	0.	0.
1–Poisson	Average[α]	Prob[$\alpha \leq \alpha_c$]	$\pi_{\text{Average}}[\alpha \leq \alpha_c]$	(Strength) $\mathbb{S}[\alpha \leq \alpha_c]$
	0.176394	0.112	0.56653	$\mathbb{S}_{\text{Poisson}}\approx 0.56653$, $\mathbb{S}_{\text{Bayes}}\approx 0.933539$
	Average[α]	Prob[$\alpha \leq \alpha_c$]	$\pi_{\text{Average}}[\alpha \leq \alpha_c]$	Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]
Binomial	0.271794	0.112	0.565674	0.
Fisher	0.269515	0.112	0.584217	0.

First scenario with positive **Type I** test

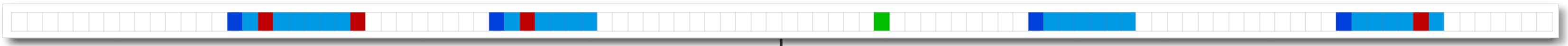
No scenarios satisfy both Type I *and* Type II tests

Appendix D: *Monte Carlo #1 notional examples*

Monte Carlo Algorithm #1: Estimate probability distribution of *coincident* strandings



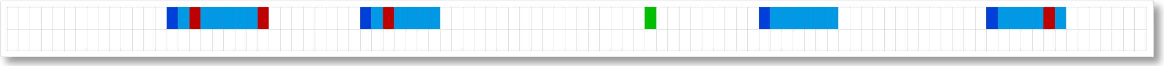
Start w/ $\mathfrak{D}_{\text{Original}} \rightarrow N_{\text{CoinS}} = 4 \quad N_{\text{NullS}} = 1$



$\delta_{\text{Max}} = 6, \Delta_{\text{Max}} = 6$

Days_0 +

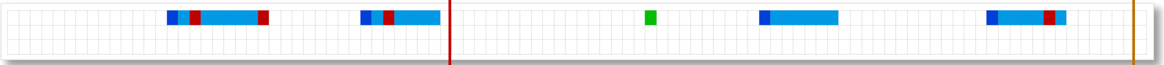
100



Samples=1000, Days=200, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$ =30, $N_{\text{S,Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \leq (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \geq (N_{\text{CS,Req}})_{\text{Min}}$]
	4	3.03529	0.498	0.119
1–Poisson	**Average[α]**	**Prob[$\alpha \leq \alpha_c$]**	**$\pi_{\text{Average}}[\alpha \leq \alpha_c]$**	**(Strength) $\mathbb{S}[\alpha \leq \alpha_c]$**
	0.0489659	0.617	0.574827	$\mathbb{S}_{\text{Poisson}}$≈0.57681, $\mathbb{S}_{\text{Bayes}}$≈0.932123
	Average[α]	**Prob[$\alpha \leq \alpha_c$]**	**$\pi_{\text{Average}}[\alpha \leq \alpha_c]$**	**Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]**
Binomial	0.0844624	0.617	0.615403	0.119
Fisher	0.0829601	0.617	0.631933	0.119

200



Samples=1000, Days=300, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$ =30, $N_{\text{S,Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \leq (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \geq (N_{\text{CS,Req}})_{\text{Min}}$]
	4	3.	0.367	0.62
1–Poisson	**Average[α]**	**Prob[$\alpha \leq \alpha_c$]**	**$\pi_{\text{Average}}[\alpha \leq \alpha_c]$**	**(Strength) $\mathbb{S}[\alpha \leq \alpha_c]$**
	0.0218144	0.987	0.581929	$\mathbb{S}_{\text{Poisson}}$≈0.323324, $\mathbb{S}_{\text{Bayes}}$≈0.808847
	Average[α]	**Prob[$\alpha \leq \alpha_c$]**	**$\pi_{\text{Average}}[\alpha \leq \alpha_c]$**	**Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]**
Binomial	0.0395385	0.62	0.699532	0.122
Fisher	0.0386922	0.62	0.717857	0.122

300

Samples=1000, Days=400, $N_{\text{Sonar/Actual}}$=5, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$ =30, $N_{\text{S,Obs}}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob[$N_{\text{CS,Req}} \leq (N_{\text{CS,Obs}})_{\text{Max}}$]	Prob[$N_{\text{CS,Obs}} \geq (N_{\text{CS,Req}})_{\text{Min}}$]
	4	3.04865	1.	0.127
1–Poisson	**Average[α]**	**Prob[$\alpha \leq \alpha_c$]**	**$\pi_{\text{Average}}[\alpha \leq \alpha_c]$**	**(Strength) $\mathbb{S}[\alpha \leq \alpha_c]$**
	0.0128884	0.987	0.581685	$\mathbb{S}_{\text{Poisson}}$≈0, $\mathbb{S}_{\text{Bayes}}$≈0
	Average[α]	**Prob[$\alpha \leq \alpha_c$]**	**$\pi_{\text{Average}}[\alpha \leq \alpha_c]$**	**Prob[$\alpha \leq \alpha_c$ AND $\pi \geq \pi_c$]**
Binomial	0.0232965	0.987	0.684783	0.127
Fisher	0.0227834	0.987	0.700241	0.127

Significant number of samples and scenarios satisfy **Type I** tests

Significant number of samples and scenarios also satisfy **both Type I and Type II** tests

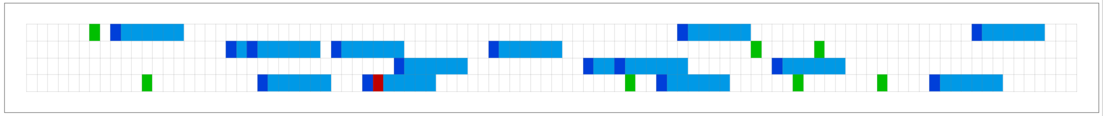
Appendix D: *Monte Carlo #1 notional examples*

N_{CoinS}	N_{NullS}
1	7
4	4
7	1

$\delta_{\text{Max}} = 6, \Delta_{\text{Max}} = 6$

Average value of α *decreases* as the relative number of observed coincident strandings *increases*

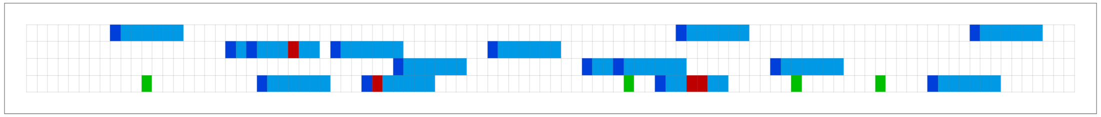
Total days = 400 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non–Coincident Strandings = 7, Coincident Strandings = 1



Samples=1000, Days=400, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effec}}$=96, $N_{\text{S,Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{cs}	$(N_{\text{cs,Obs}})_{\text{Max}}$	$(N_{\text{cs,Req}})_{\text{Min}}$	Prob$[N_{\text{cs,Req}} \le (N_{\text{cs,Obs}})_{\text{Max}}]$	Prob$[N_{\text{cs,Obs}} \ge (N_{\text{cs,Req}})_{\text{Min}}]$
	1	7.95789	0.	0.
1–Poisson	**Average[α]**	**Prob$[\alpha \le \alpha_c]$**	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	**(Strength)** $\mathbb{S}[\alpha \le \alpha_c]$
	0.966559	0.	0	$\mathbb{S}_{\text{Poisson}} \approx 0$, $\mathbb{S}_{\text{Bayes}} \approx 0$
	Average[α]	**Prob$[\alpha \le \alpha_c]$**	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	**Prob$[\alpha \le \alpha_c$ AND $\pi \ge \pi_c]$**
Binomial	0.966052	0.	0	0.
Fisher	0.966805	0.	0	0.

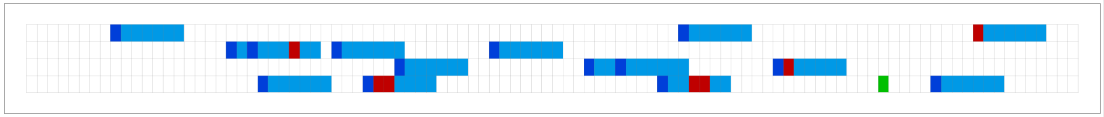
Total days = 400 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non–Coincident Strandings = 4, Coincident Strandings = 4



Samples=1000, Days=400, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=96, $N_{\text{S,Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{cs}	$(N_{\text{cs,obs}})_{\text{Max}}$	$(N_{\text{cs,Req}})_{\text{Min}}$	Prob$[N_{\text{cs,Req}} \le (N_{\text{cs,obs}})_{\text{Max}}]$	Prob$[N_{\text{cs,obs}} \ge (N_{\text{cs,Req}})_{\text{Min}}]$
	4	5.55789	0.	0.
1–Poisson	**Average[α]**	**Prob$[\alpha \le \alpha_c]$**	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	**(Strength)** $\mathbb{S}[\alpha \le \alpha_c]$
	0.316926	0.158	0.56653	$\mathbb{S}_{\text{Poisson}} \approx 0.371163$, $\mathbb{S}_{\text{Bayes}} \approx 0.85094$
	Average[α]	**Prob$[\alpha \le \alpha_c]$**	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	**Prob$[\alpha \le \alpha_c$ AND $\pi \ge \pi_c]$**
Binomial	0.382308	0.	0	0.
Fisher	0.38216	0.	0	0.

Total days = 400 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non–Coincident Strandings = 1, Coincident Strandings = 7



Samples=1000, Days=400, $N_{\text{Sonar/Actual}}$=15, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=96, $N_{\text{S,Obs}}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{cs}	$(N_{\text{cs,obs}})_{\text{Max}}$	$(N_{\text{cs,Req}})_{\text{Min}}$	Prob$[N_{\text{cs,Req}} \le (N_{\text{cs,obs}})_{\text{Max}}]$	Prob$[N_{\text{cs,obs}} \ge (N_{\text{cs,Req}})_{\text{Min}}]$
	6	4.29474	0.551	0.204
1–Poisson	**Average[α]**	**Prob$[\alpha \le \alpha_c]$**	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	**(Strength)** $\mathbb{S}[\alpha \le \alpha_c]$
	0.16418	0.551	0.563694	$\mathbb{S}_{\text{Poisson}} \approx 0.45402$, $\mathbb{S}_{\text{Bayes}} \approx 0.883464$
	Average[α]	**Prob$[\alpha \le \alpha_c]$**	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	**Prob$[\alpha \le \alpha_c$ AND $\pi \ge \pi_c]$**
Binomial	0.220645	0.204	0.634636	0.025
Fisher	0.219679	0.204	0.653735	0.025

Appendix D: *Monte Carlo #1 notional examples*

Days$_0$ +

$\delta_{Max} = 6, \Delta_{Max} = 6$

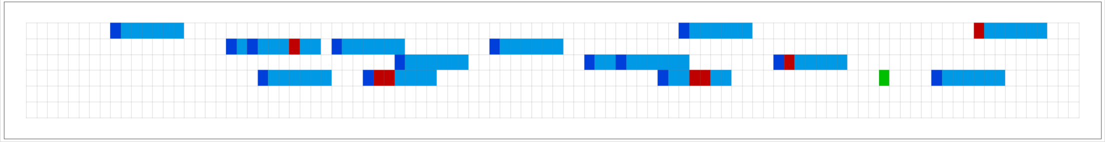
200

400

800

For a fixed number of total strandings, the average value of α *decreases* as the number of days without sonar *increases*

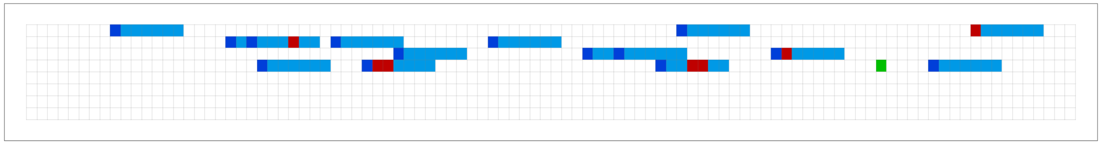
Total days = 600 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non–Coincident Strandings = 1, Coincident Strandings = 7



Samples=1000, Days=600, $N_{Sonar/Actual}$=15, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=96, $N_{S,Obs}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob[$N_{CS,Req} \le (N_{CS,obs})_{Max}$]	Prob[$N_{CS,obs} \ge (N_{CS,Req})_{Min}$]
	6	4.34286	0.971	0.19
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	(Strength) δ[α ≤ α_c]
	0.08004	0.543	0.563805	$\delta_{Poisson}$≈0.619516, δ_{Bayes}≈0.942184
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.110765	0.543	0.639595	0.028
Fisher	0.110074	0.543	0.655838	0.028

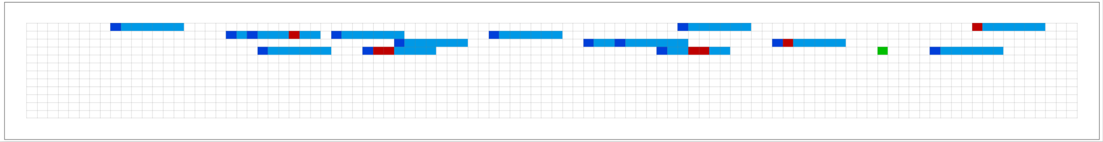
Total days = 800 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non–Coincident Strandings = 1, Coincident Strandings = 7



Samples=1000, Days=800, $N_{Sonar/Actual}$=15, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=96, $N_{S,Obs}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob[$N_{CS,Req} \le (N_{CS,obs})_{Max}$]	Prob[$N_{CS,obs} \ge (N_{CS,Req})_{Min}$]
	6	3.	1.	0.865
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	(Strength) δ[α ≤ α_c]
	0.0458152	0.865	0.568381	$\delta_{Poisson}$≈0.53615, δ_{Bayes}≈0.903145
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.0637817	0.553	0.752937	0.211
Fisher	0.0633373	0.553	0.771199	0.211

Total days = 1200 (100 per row) | Sonar Days = 15 (Total), 96 (Padded, assuming δ_{Max}=6)
Total Strandings = 8 | Non–Coincident Strandings = 1, Coincident Strandings = 7



Samples=1000, Days=1200, $N_{Sonar/Actual}$=15, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=96, $N_{S,Obs}$=8, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,obs})_{Max}$	$(N_{CS,Req})_{Min}$	Prob[$N_{CS,Req} \le (N_{CS,obs})_{Max}$]	Prob[$N_{CS,obs} \ge (N_{CS,Req})_{Min}$]
	6	3.06087	1.	0.549
1-Poisson	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	(Strength) δ[α ≤ α_c]
	0.0229598	0.874	0.568599	$\delta_{Poisson}$≈0.564909, δ_{Bayes}≈0.905086
	Average[α]	Prob[α ≤ α_c]	$\pi_{Average}$[α ≤ α_c]	Prob[α ≤ α_c AND π ≥ π_c]
Binomial	0.0313011	0.874	0.725487	0.199
Fisher	0.0310893	0.874	0.741972	0.199

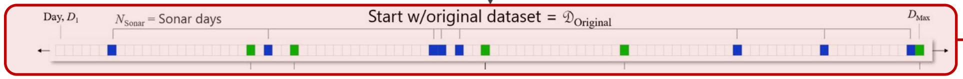
Appendix E: *Uncertainties – additional comments*

Mitigating ambiguous or inconsistent dataset preparation

Recall slide 12

What is the fundamental statistical analysis problem?

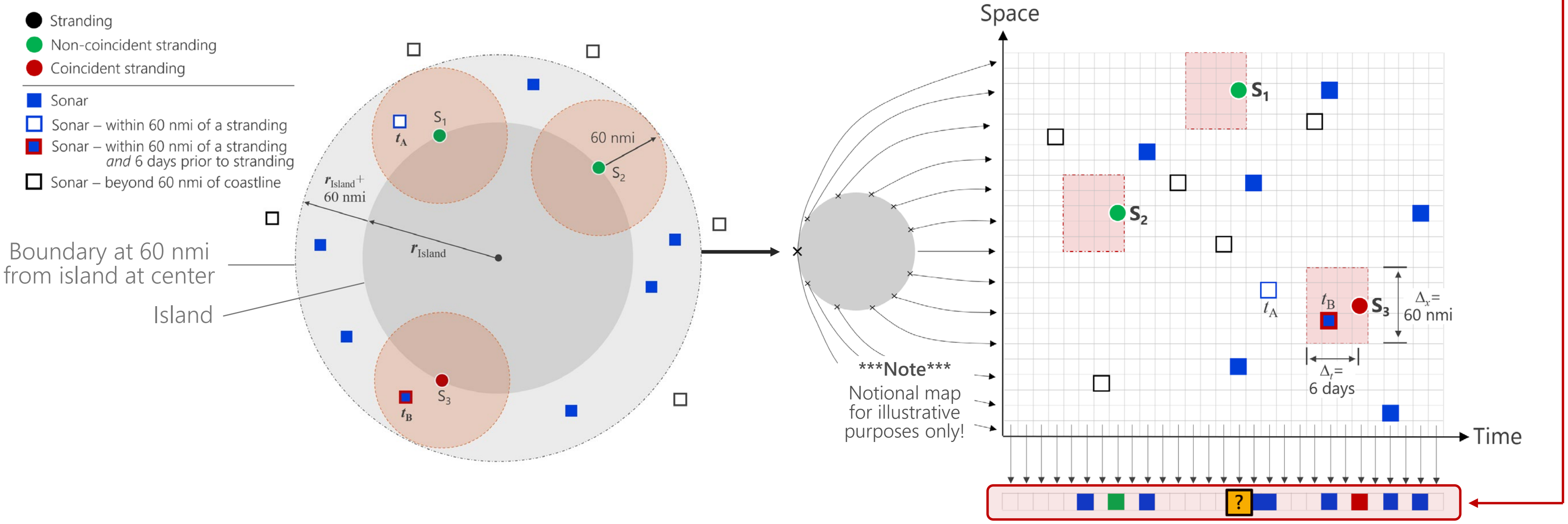
Compare stranding statistics for two typically sparse and partly ambiguously disentangled event datasets



Start w/original dataset = $\mathfrak{D}_{\text{Original}}$

But...how is this dataset *prepared*?

Effectively collapses a three-dimensional space (two spatial coordinates plus time) onto a single dimension (time)



● Stranding
● Non-coincident stranding
● Coincident stranding
■ Sonar
□ Sonar – within 60 nmi of a stranding
■ Sonar – within 60 nmi of a stranding *and* 6 days prior to stranding
□ Sonar – beyond 60 nmi of coastline
S_1
t_A
S_2
60 nmi
r_{Island} + 60 nmi
r_{Island}
S_3
t_B
Space
Time
S_1
S_2
t_A
t_B
S_3
Δ_x = 60 nmi
Δ_t = 6 days
?

Boundary at 60 nmi from island at center

Island

Note Notional map for illustrative purposes only!

The time series allows only one label per day

What label ought to be assigned to this day?

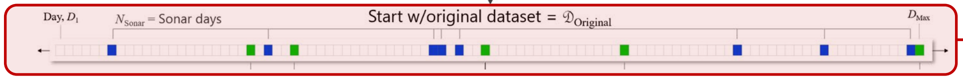
Appendix E: *Uncertainties – additional comments*

Mitigating ambiguous or inconsistent dataset preparation

Recall slide 12

What is the fundamental statistical analysis problem?

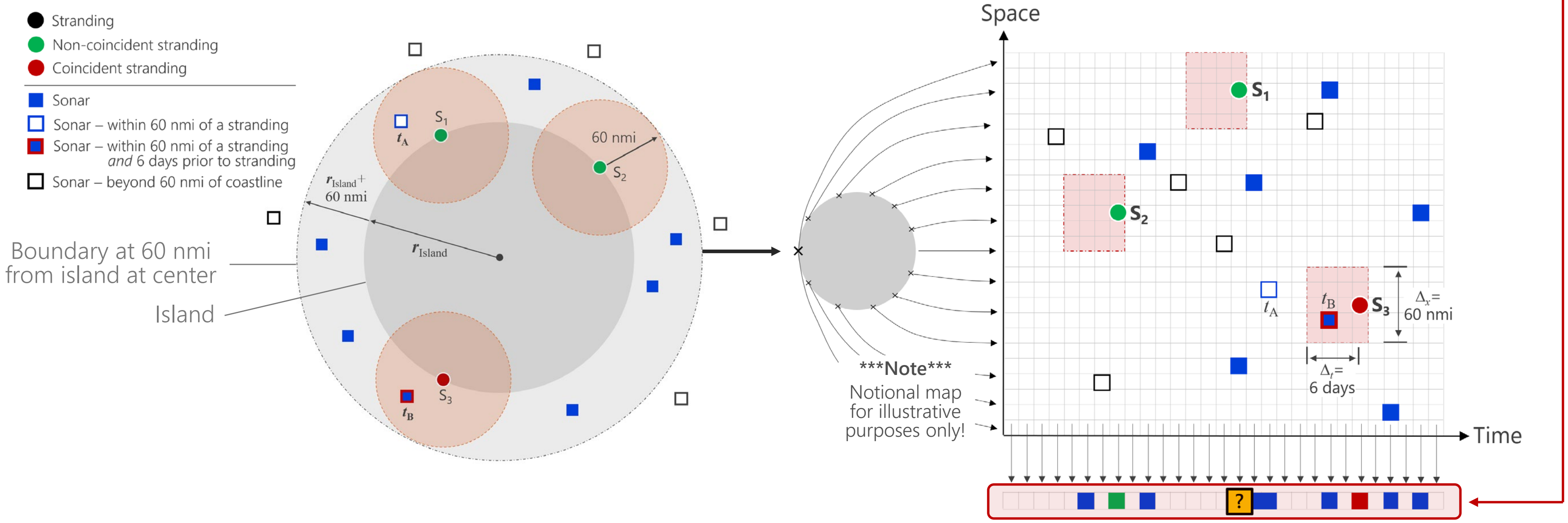
Compare stranding statistics for two typically sparse and partly ambiguously disentangled event datasets



Day, D_1
N_{Sonar} = Sonar days
Start w/original dataset = $\mathcal{D}_{\text{Original}}$
D_{Max}

But…how is this dataset *prepared*?

Effectively collapses a three-dimensional space (two spatial coordinates plus time) onto a single dimension (time)



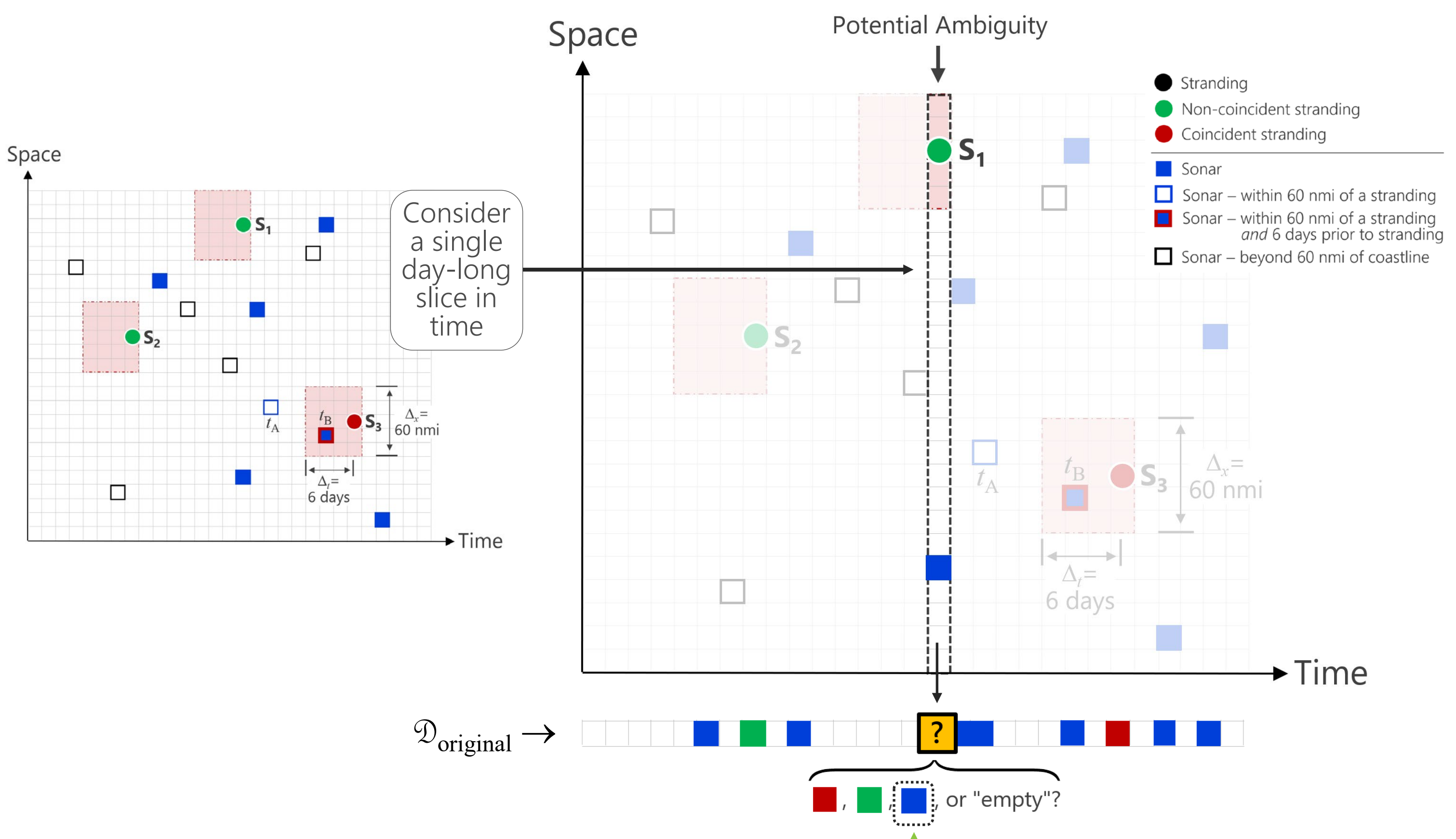
● Stranding
● Non-coincident stranding
● Coincident stranding
■ Sonar
□ Sonar – within 60 nmi of a stranding
■ Sonar – within 60 nmi of a stranding *and* 6 days prior to stranding
□ Sonar – beyond 60 nmi of coastline
Boundary at 60 nmi from island at center
Island
r_{Island} + 60 nmi
r_{Island}
60 nmi
S_1
S_2
S_3
t_A
t_B
Space
Time
Note
Notional map for illustrative purposes only!
Δ_x = 60 nmi
Δ_t = 6 days
?

The time series allows only one label per day

What label ought to be assigned to this day?

Appendix E: *Uncertainties – additional comments*

Effectively collapses a three-dimensional space (two spatial coordinates plus time) onto a single dimension (time)



Consider a single day-long slice in time

Potential Ambiguity

● Stranding
● Non-coincident stranding
● Coincident stranding

■ Sonar
□ Sonar – within 60 nmi of a stranding
■ Sonar – within 60 nmi of a stranding *and* 6 days prior to stranding
□ Sonar – beyond 60 nmi of coastline

$\mathcal{D}_{\text{original}} \rightarrow$

■, ■, ■, or "empty"?

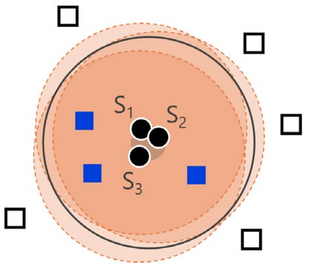
Because we are testing for a statistical difference between null strandings (no sonar) and coincident strandings (sonar) days in a time series, this must be labeled as a ***sonar*** **day**

The stranding, S_1, is—statistically—neither null nor coincident (in time series)

Appendix E: *Uncertainties – additional comments*

Effectively collapses a three-dimensional space (two spatial coordinates plus time) onto a single dimension (time)

Pigeonholing is increasing likely to happen as r_{Island} increases relative to 60 nmi

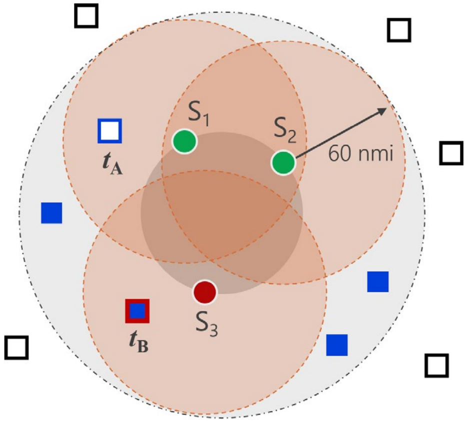


Case A

In the extreme limit when the size of the island ≈ 0, ambiguities do not arise.

All strandings are either coincident or null.

Most datasets implicitly make this assumption

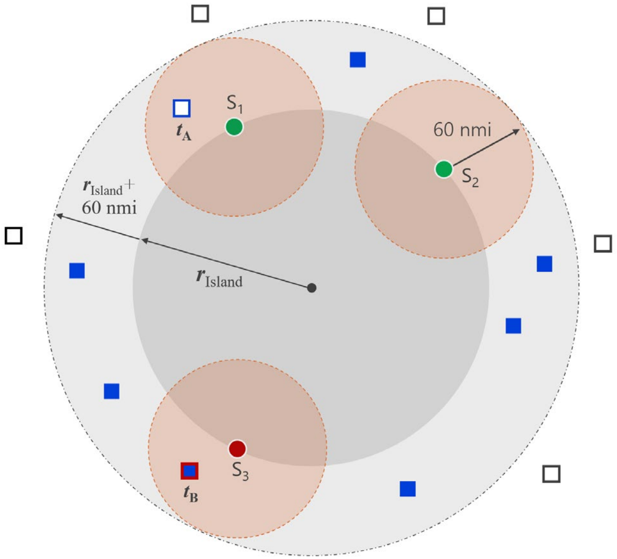


Case B

When island size ≈ 60 nmi, ambiguities may arise.

Non-coincident strandings may be incorrectly labeled coincident because they appear within six days, but they are too far separated in space.

Neither are they null because they do appear on sonar days



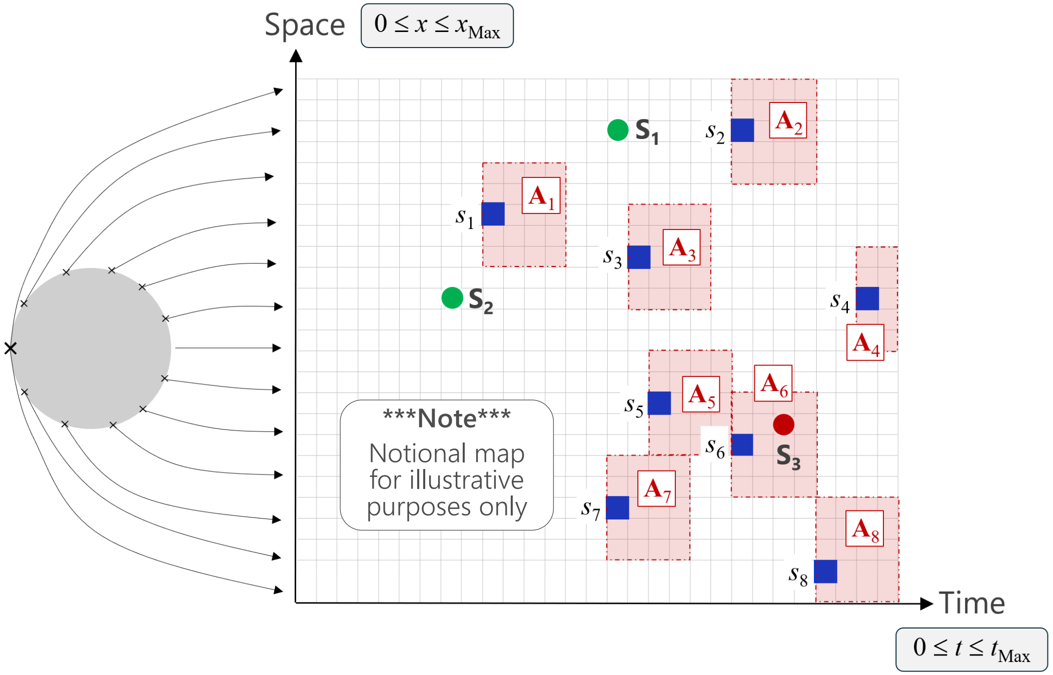
Case C

As island size increases beyond 60 nmi, there is an increasing likelihood that ambiguous labels will arise on any given day in a dataset

Appendix E: *Uncertainties – additional comments*

Toward a possible mitigation

Estimate null and coincident stranding rates using combined *space* plus *time* data

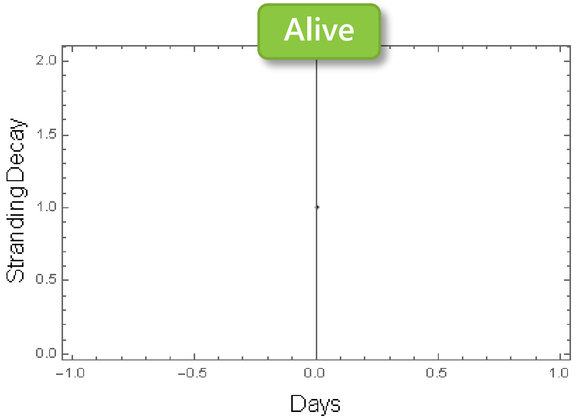


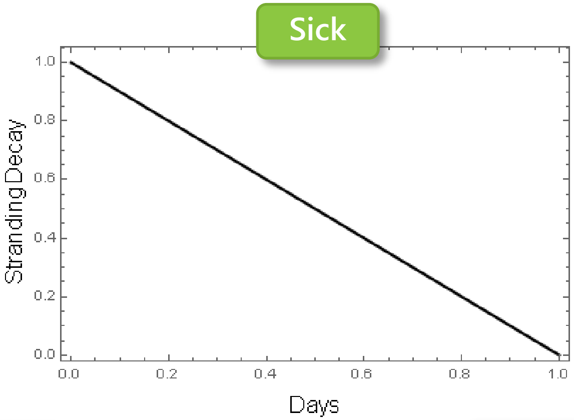
Note
Notional map for illustrative purposes only

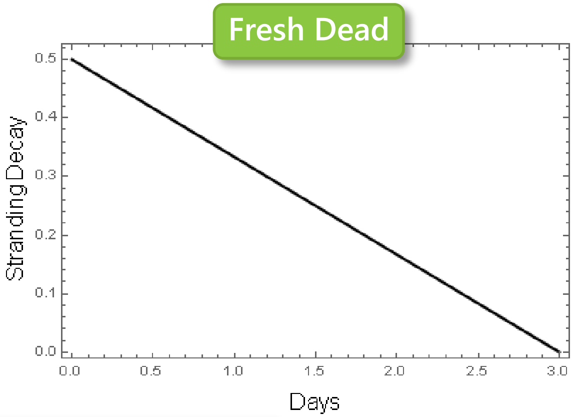
$$\begin{cases} \textbf{Null-stranding rate} = \lambda_0 \equiv \dfrac{\text{Number of strandings } \textit{outside} \text{ of effective sonar zones}}{\text{Total area - Area of effective sonar zones}} = \dfrac{N_{\text{NullS}}}{t_{\text{Max}} \cdot x_{\text{Max}} - \sum_{i=1}^{s_{Max}} A_i} \\ \textbf{Coincident-stranding rate} = \lambda_{\text{CS}} \equiv \dfrac{\text{Number of strandings } \textit{inside} \text{ of effective sonar zones}}{\text{Area of effective sonar zones}} = \dfrac{N_{\text{CoinS}}}{\sum_{i=1}^{s_{Max}} A_i} \end{cases}$$

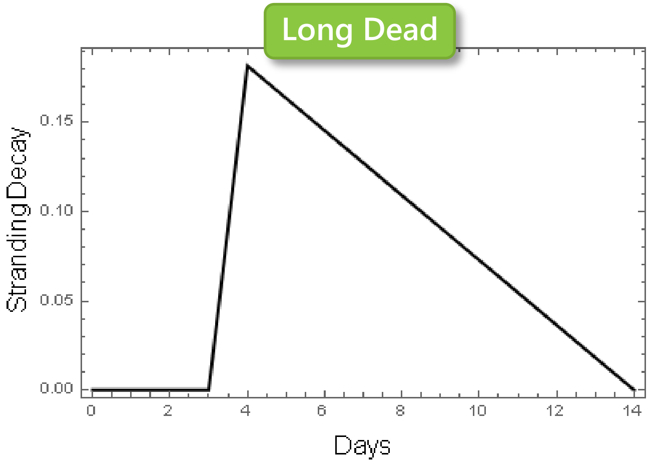
Appendix F: *Necropsy-dependent stranding decay functions*

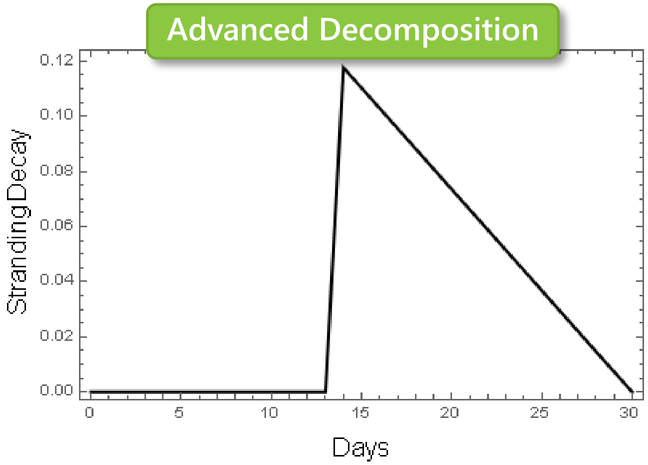
- The Mathematica source code developed for this study (see Appendix F) includes an option that tailors stranding decay (i.e., a probabilistic assignment of an actual stranding date given an observed date) as a function of necropsy state: *alive, sick, fresh dead, long dead*, and *advanced decomposition*
- We show an illustrative set (but many other forms are possible)





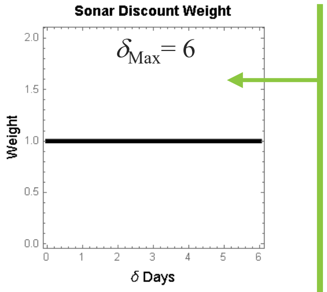


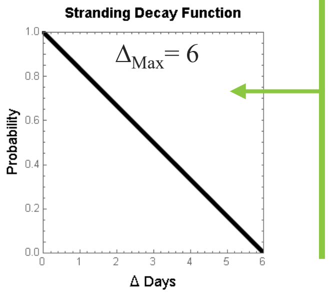


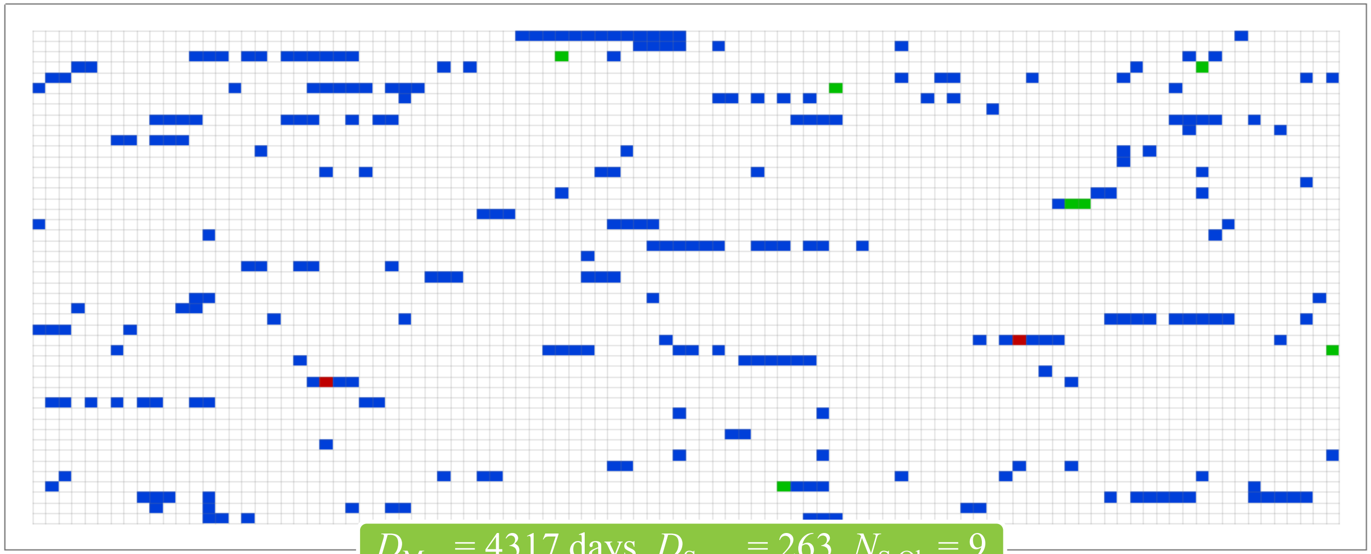


Appendix G: *Mariana Islands case study*

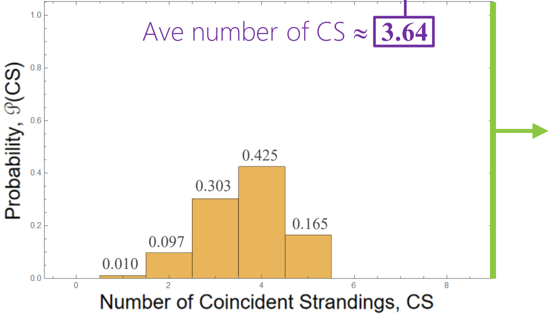
Mariana Islands (Simonis et al. area of study)



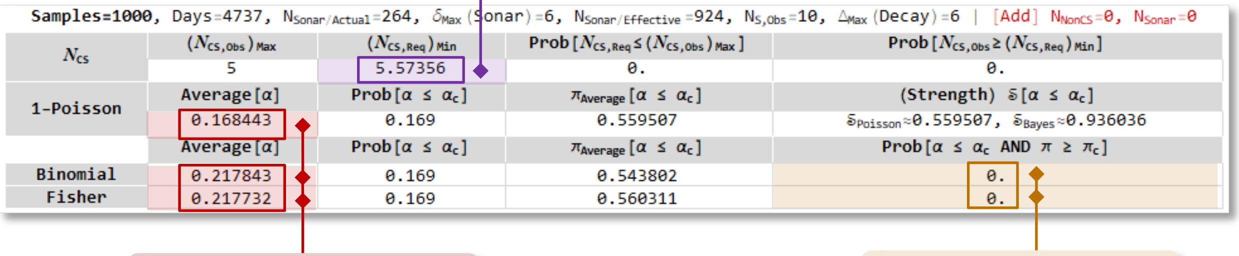




$D_{\text{Max}} = 4317$ days, $D_{\text{Sonar}} = 263$, $N_{\text{S,Obs}} = 9$



Observed CS < Required CS



Samples=1000, Days=4737, $N_{\text{Sonar/Actual}}$=264, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=924, $N_{\text{S,Obs}}$=10, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

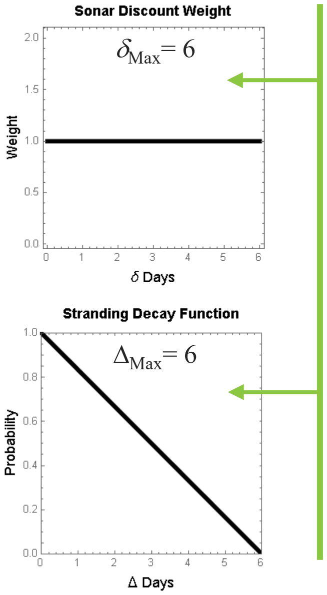
N_{CS}	$(N_{\text{CS,Obs}})_{\text{Max}}$	$(N_{\text{CS,Req}})_{\text{Min}}$	Prob$[N_{\text{CS,Req}} \le (N_{\text{CS,Obs}})_{\text{Max}}]$	Prob$[N_{\text{CS,Obs}} \ge (N_{\text{CS,Req}})_{\text{Min}}]$
	5	5.57356	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\delta[\alpha \le \alpha_c]$
	0.168443	0.169	0.559507	$\delta_{\text{Poisson}} \approx 0.559507$, $\delta_{\text{Bayes}} \approx 0.936036$
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.217843	0.169	0.543802	0.
Fisher	0.217732	0.169	0.560311	0.

Poisson test, BET, and FET all yield $\alpha > \alpha_c = 0.05$

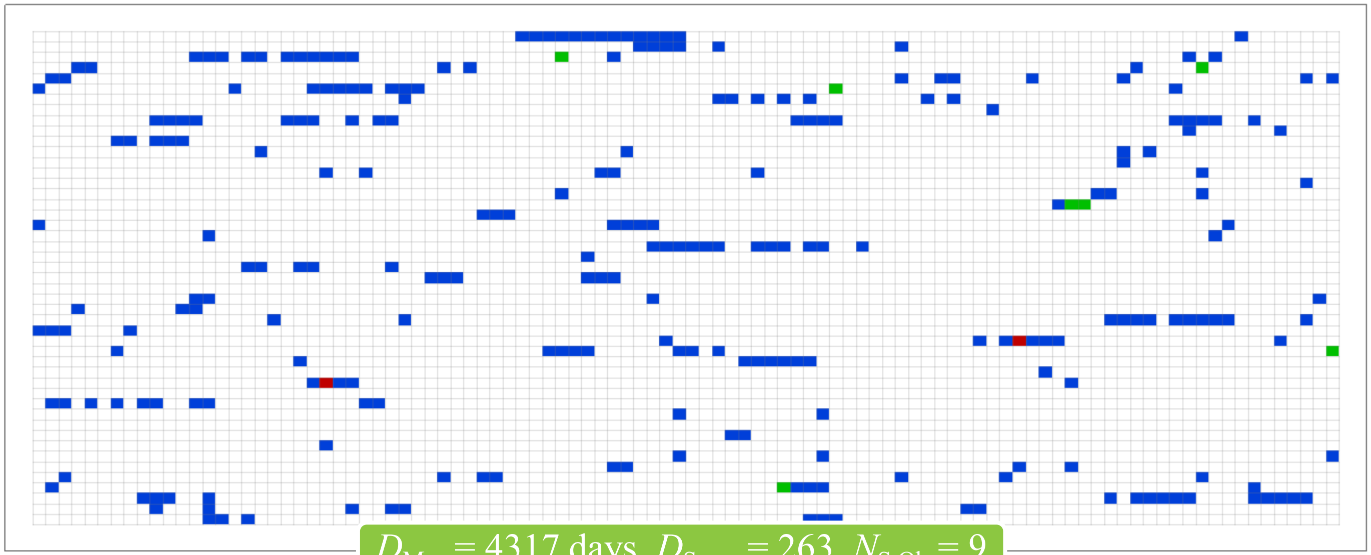
No scenarios satisfy both Type I and Type II tests

Appendix G: *Mariana Islands case study*

Mariana Islands (Simonis et al. area of study) → Cannot reject H0

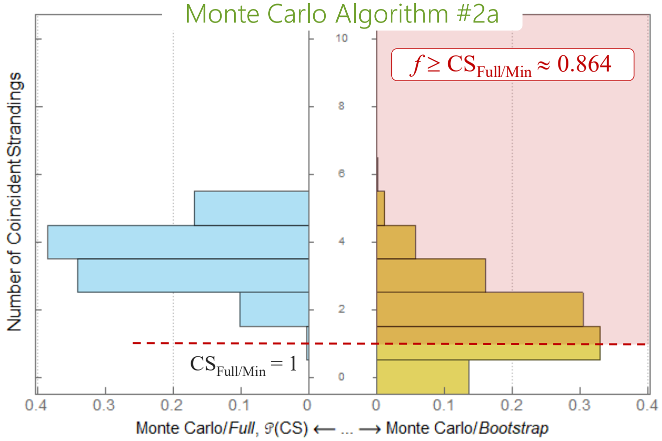

Sonar Discount Weight

Stranding Decay Function

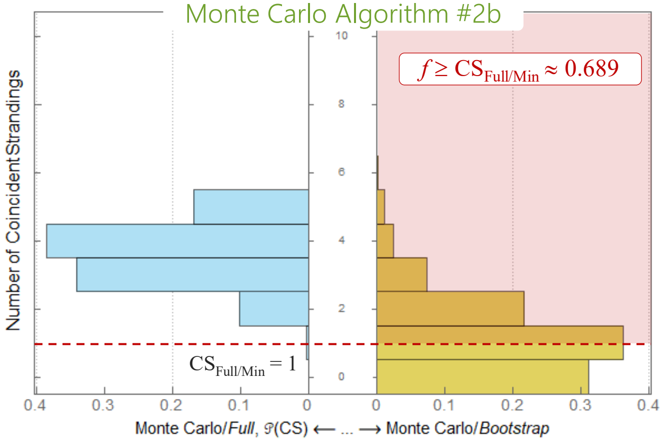


$D_{\text{Max}} = 4317$ days, $D_{\text{Sonar}} = 263$, $N_{\text{S,Obs}} = 9$

Monte Carlo Algorithm #2a

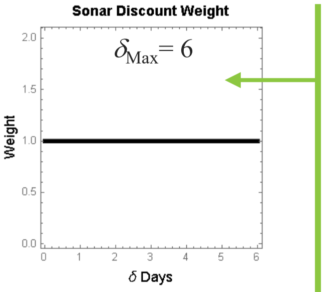


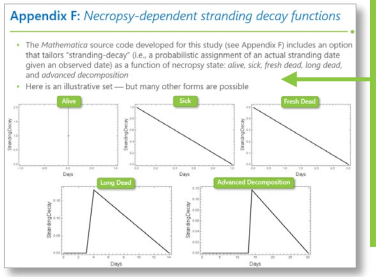
Monte Carlo Algorithm #2b



Appendix G: *SOCAL case study*

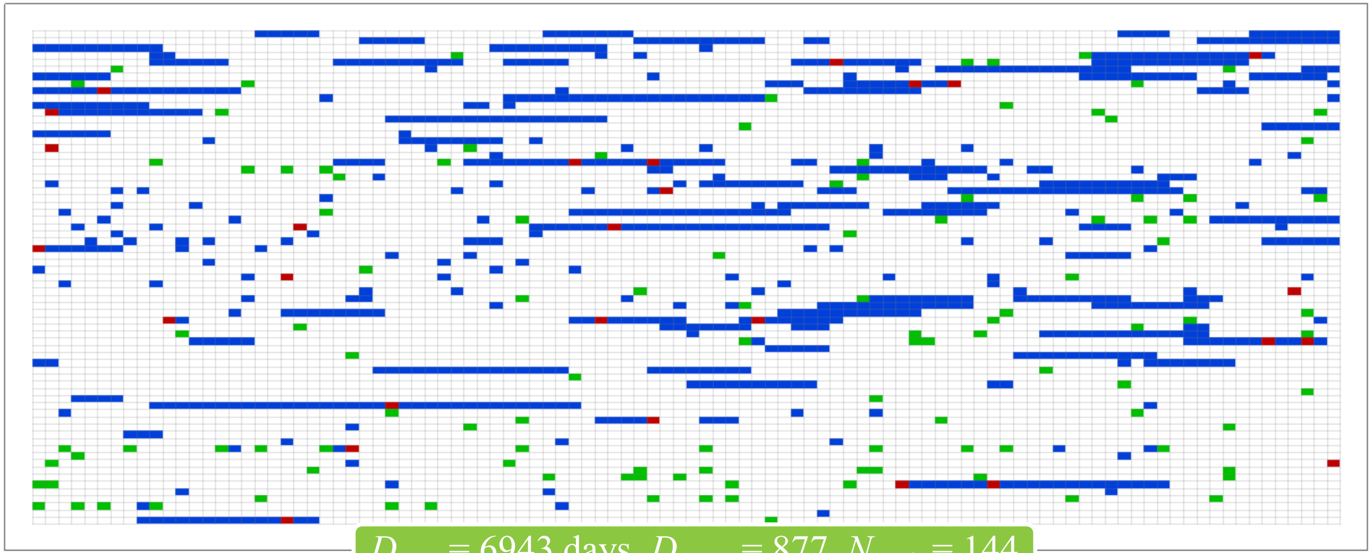
SOCAL



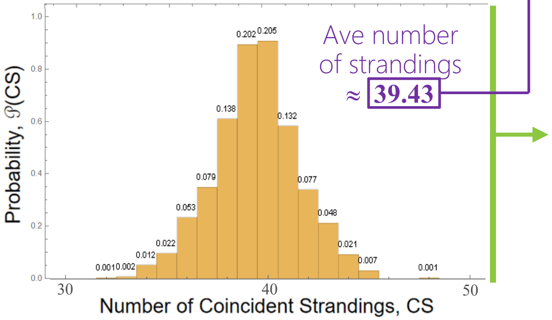


Appendix F: *Necropsy-dependent stranding decay functions*

- The *Mathematica* source code developed for this study (see Appendix F) includes an option that tailors "stranding-decay" (i.e., a probabilistic assignment of an actual stranding date given an observed date) as a function of necropsy state: *alive, sick, fresh dead, long dead,* and *advanced decomposition*
- Here is an illustrative set — but many other forms are possible

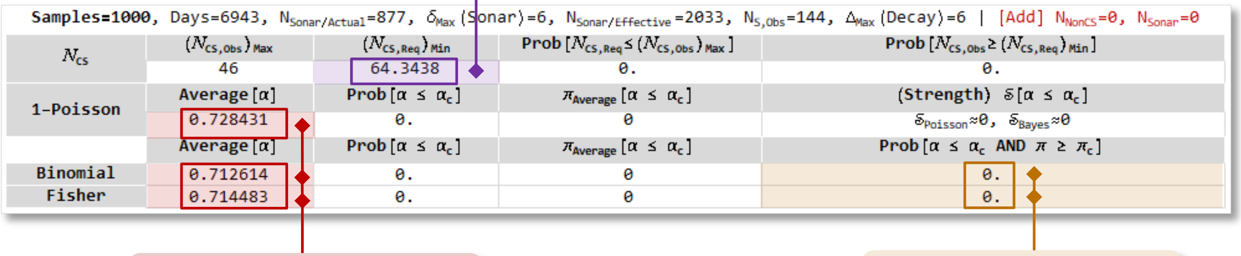


$D_{\text{Max}} = 6943$ days, $D_{\text{Sonar}} = 877$, $N_{\text{S,Obs}} = 144$



Ave number of strandings $\approx \boxed{39.43}$

Observed CS < Required CS



Samples=1000, Days=6943, $N_{\text{Sonar/Actual}}$=877, δ_{Max}(Sonar)=6, $N_{\text{Sonar/Effective}}$=2033, $N_{\text{S,Obs}}$=144, Δ_{Max}(Decay)=6 \| [Add] N_{NonCS}=0, N_{Sonar}=0

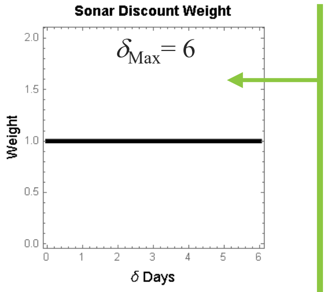
$\mathcal{N}_{\text{CS}}$	$(\mathcal{N}_{\text{CS,Obs}})_{\text{Max}}$	$(\mathcal{N}_{\text{CS,Req}})_{\text{Min}}$	Prob$[\mathcal{N}_{\text{CS,Req}} \le (\mathcal{N}_{\text{CS,Obs}})_{\text{Max}}]$	Prob$[\mathcal{N}_{\text{CS,Obs}} \ge (\mathcal{N}_{\text{CS,Req}})_{\text{Min}}]$
	46	64.3438	0.	0.
1-Poisson	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	(Strength) $\mathcal{S}[\alpha \le \alpha_c]$
	0.728431	0.	0	$\mathcal{S}_{\text{Poisson}} \approx 0$, $\mathcal{S}_{\text{Bayes}} \approx 0$
	Average[α]	Prob[$\alpha \le \alpha_c$]	$\pi_{\text{Average}}[\alpha \le \alpha_c]$	Prob[$\alpha \le \alpha_c$ AND $\pi \ge \pi_c$]
Binomial	0.712614	0.	0	0.
Fisher	0.714483	0.	0	0.

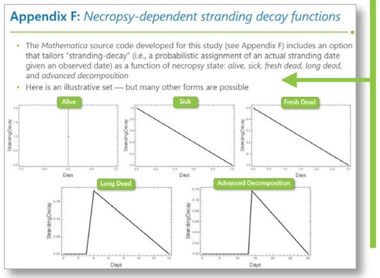
Poisson test, BET, and FET all yield $\alpha > \alpha_c = 0.05$

No scenarios satisfy both Type I and Type II tests

Appendix G: *SOCAL case study*

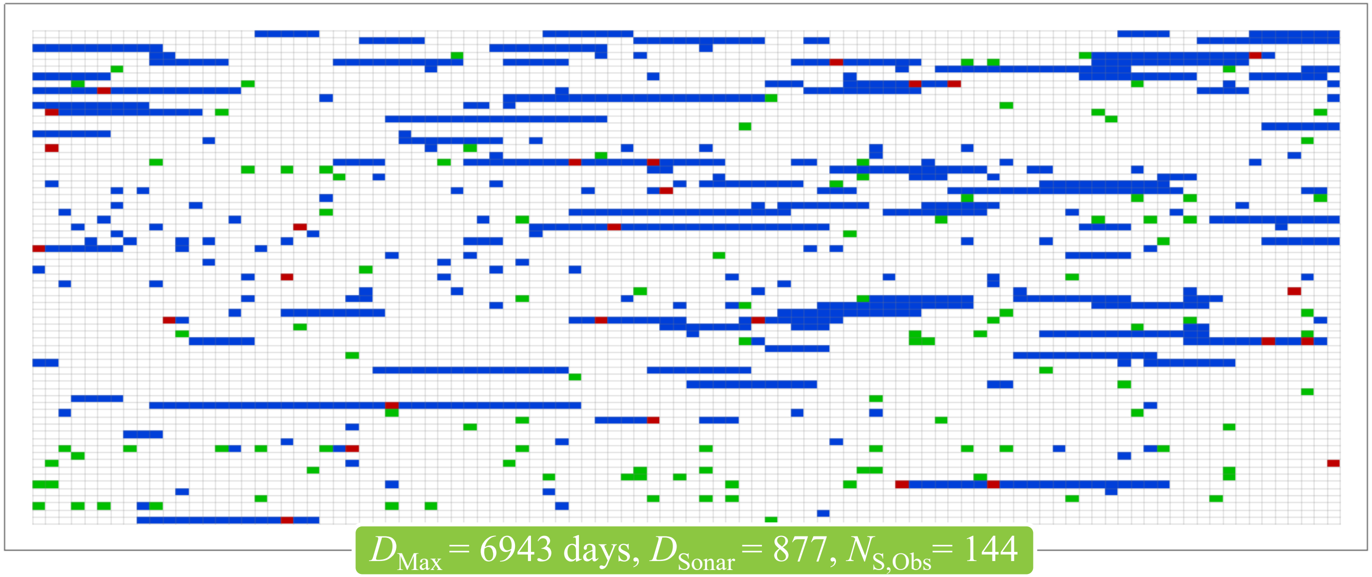
SOCAL → Cannot reject H0



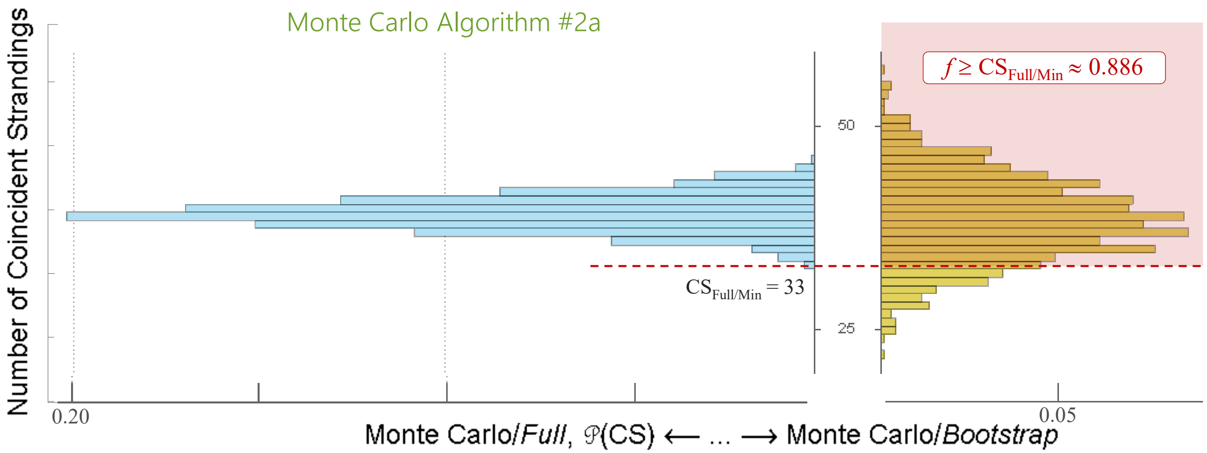


Appendix F: *Necropsy-dependent stranding decay functions*

- The *Mathematica* source code developed for this study (see Appendix F) includes an option that tailors "stranding-decay" (i.e., a probabilistic assignment of an actual stranding date given an observed date) as a function of necropsy state: *alive, sick, fresh dead, long dead,* and *advanced decomposition*
- Here is an illustrative set — but many other forms are possible



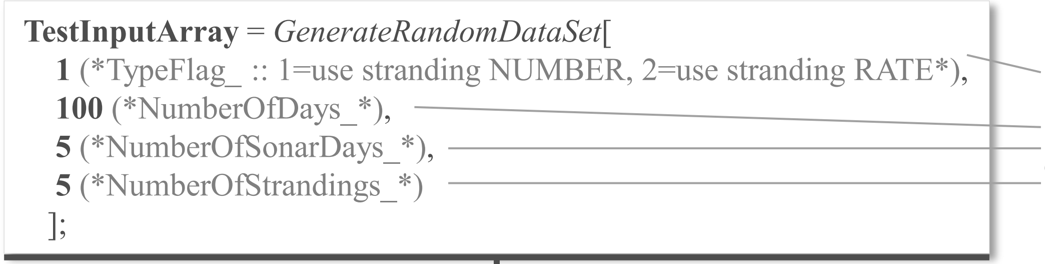
$D_{\mathrm{Max}} = 6943$ days, $D_{\mathrm{Sonar}} = 877$, $N_{\mathrm{S,Obs}} = 144$



Appendix H: *Mathematica functions*

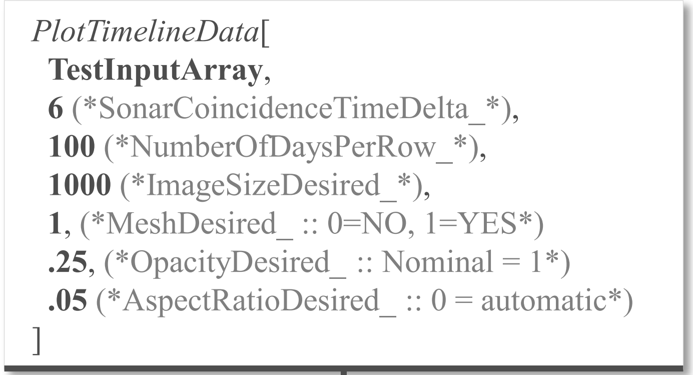
- 2,000+ lines of source code have been developed for this study
  - Require Wolfram Mathematica versions 12.0 and higher
  - Available upon request
- Main function clusters
  - Data import/information-extract/necropsy functions
  - Modify input data files (for scenario development/experimentation)
    - Generate random dataset, add/subtract strandings, add/delete days, add sonar
  - Visualize timeline
  - Stranding-decay/sonar-discount/fractional coincident stranding functions
  - Statistical tests: significance and power estimates
    - Poisson means test
    - Fisher's exact test
    - Exact binomial test
  - Poisson confidence intervals and mean "Accept/Reject Criteria Chart" (PM-ARCC)
  - Estimate # of coincident strandings required to pass Type I and Type II tests
  - Monte Carlo simulations
    - Monte Carlo algorithms #1/modified-1, #2a, and #2b

Appendix I: *Sample Mathematica session* (1/3)

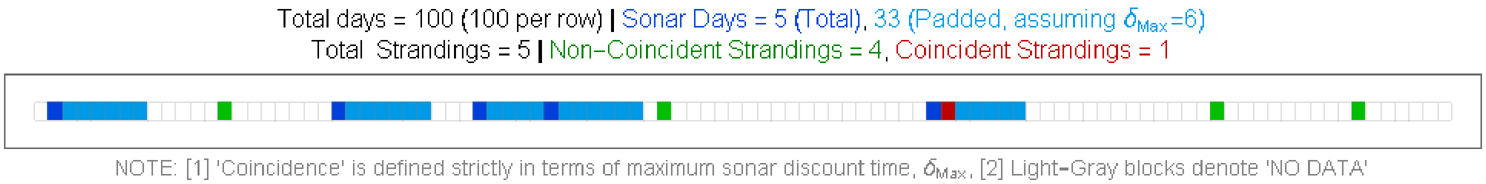


```
TestInputArray = GenerateRandomDataSet[
   1 (*TypeFlag_ :: 1=use stranding NUMBER, 2=use stranding RATE*),
   100 (*NumberOfDays_*),
   5 (*NumberOfSonarDays_*),
   5 (*NumberOfStrandings_*)
  ];
```

Note
Text highlighted in light gray between the parentheses represents comments, not executable source code



```
PlotTimelineData[
  TestInputArray,
  6 (*SonarCoincidenceTimeDelta_*),
  100 (*NumberOfDaysPerRow_*),
  1000 (*ImageSizeDesired_*),
  1, (*MeshDesired_ :: 0=NO, 1=YES*)
  .25, (*OpacityDesired_ :: Nominal = 1*)
  .05 (*AspectRatioDesired_ :: 0 = automatic*)
]
```



Total days = 100 (100 per row) | Sonar Days = 5 (Total), 33 (Padded, assuming δ_{Max}=6)
Total Strandings = 5 | Non-Coincident Strandings = 4, Coincident Strandings = 1

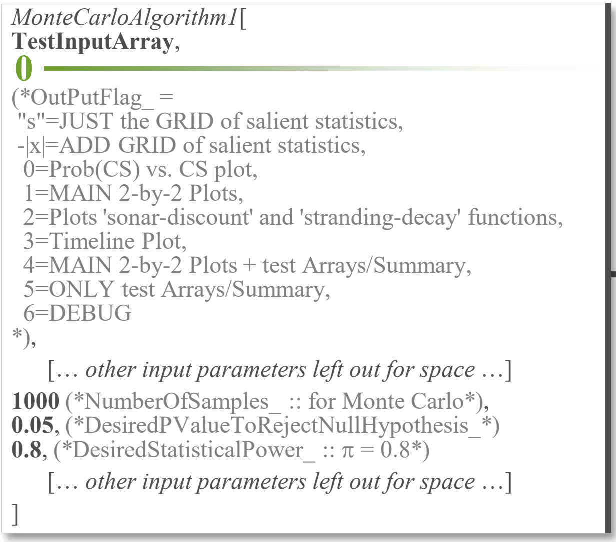
NOTE: [1] 'Coincidence' is defined strictly in terms of maximum sonar discount time, δ_{Max}, [2] Light-Gray blocks denote 'NO DATA'

Appendix I: *Sample Mathematica session (2/3)*

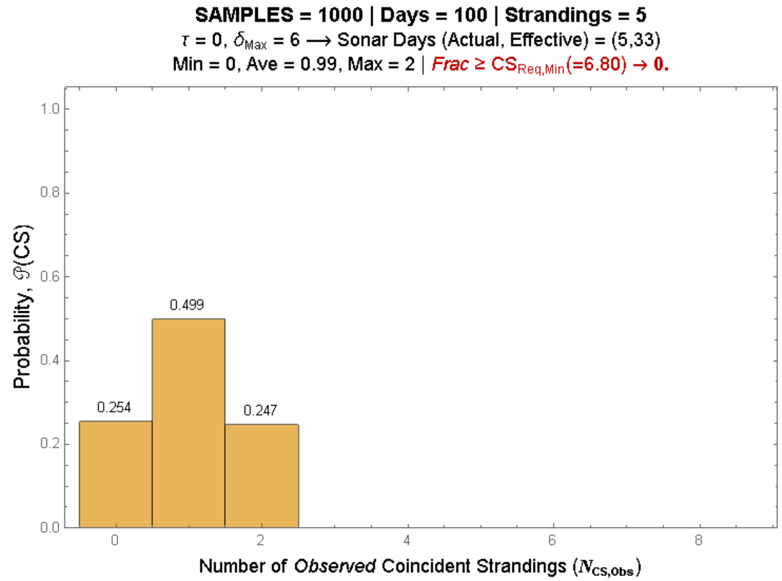
```
MonteCarloAlgorithm1[
TestInputArray,
"s"
(*OutPutFlag_ =
 "s"=JUST the GRID of salient statistics,
 -|x|=ADD GRID of salient statistics,
  0=Prob(CS) vs. CS plot,
  1=MAIN 2-by-2 Plots,
  2=Plots 'sonar-discount' and 'stranding-decay' functions,
  3=Timeline Plot,
  4=MAIN 2-by-2 Plots + test Arrays/Summary,
  5=ONLY test Arrays/Summary,
  6=DEBUG
*),
0, (*AddOneUnobservedNonCoincidentStrandingFlag_ :: 0=NO,1=YES :: addition
ONLY changes value of ProbabilityOfStrandingNull, and therefore,
ExpectedNumberOfStrandings*)
0, (*AddSonarDaysNum_ :: Nominal =0, basic 'robustness' probe = 1, but can use
any positive number :: Monte Carlo sampling includes random insertion of
specified number of additional sonar days*)
100 (*NumberOfDaysPerRowForVisualTimeline_*),
"Test Text", (*DescriptiveTextStringForVisualTimeline__ :: Text to displkay
'between QUOTES' for visual timeline display*)
1200 (*ImageSizePixelsDesired_ :: Nominal for Andy Home PC = 1400*),
1000 (*NumberOfSamples_ :: for Monte Carlo*),
0.05, (*DesiredPValueToRejectNullHypothesis_ :: nominal -> 0.05*)
0.8, (*DesiredStatisticalPower_ :: nominal -> 0.8*)
6, (*LastSonarDayToStrandingCoincidenceIntervalThreshold_ :: nominal=6
days*)
6, (*LastSonarDayToStrandingCoincidenceIntervalThresholdMax_ :: for extended
parse*)
(*------------*)
(*Necropsy Flag*)
(*------------*)
0, (*NecropsyFlag_ :: 0=use DEFAULT values, 1=use NECROPSY-STATE-
SPECIFIC stranding-decay parameters*)
(*-----------------------------------*)
(*Stranding Decay function parameters*)
(*-----------------------------------*)
0 (*DayMin_*), 6 (*DayMax_*), 1 (*FuncMin_*), 0 (*FuncMax_*),
1 (*PowerN_*), 1 (*MinValue"At0ORMaxFlag_*),
```

```
(*-----------------------------------------------------------*)
(*Stranding Decay function parameters :: NECROPSY-STATE-SPECIFIC
... these are used ONLY if NecropsyFlag==0*)
(*...These must all be ARRAYS ::
1=alive,
  2=sick/injured,
  3=fresh dead,
  4=long dead/moderate decomposition,
  5=advanced decomposition*)
(*-----------------------------------------------------------*)
{0,0,0,4,14}, (*StrandingDecayDayMinNecropsyStateSpecific_*)
{0,1,3,14,30}, (*StrandingDecayDayMaxNecropsyStateSpecific_=Subscript[Δ, Max] *)
{1,1,1,1,1}, (*StrandingDecayFuncMinNecropsyStateSpecific_*)
{1,0,0,0,0}, (*StrandingDecayFuncMaxNecropsyStateSpecific_*)
{0,1,1,1,1}, (*StrandingDecayPowerNNecropsyStateSpecific_*)
{1,1,1,1,1}, (*StrandingDecayMinValueAt0ORMaxFlagNecropsyStateSpecific_*)
(*-----------------------------------*)
(*Sonar discount function parameters*)
(*-----------------------------------*)
0 (* SonarDiscountFunctionTypeFlag_ :: 0 = nominal/ramp-style; 1 = sigmoid*),
0 (*SonarDiscountDayMin_*), 0 (*SonarDiscountDayCen_*),
0 (*SonarDiscountImpactDelay_*), 6 (*SonarDiscountDayMax_*),
1 (*SonarDiscountFuncMin_*), 0 (*SonarDiscountFuncMax_*),
0 (*SonarDiscountPowerN_*),
(*-----------------------------------*)
(*Statistical test parameters*)
(*-----------------------------------*)
10 (*Lambda0MultiplicativeFactorMax_*),
50 (*Lambda1Samples_*),
5 (*BinomialExactTestPowerCountDeltaMax_*),
(*-----------------------------------*)
(*Timeline plot display parameters*)
(*-----------------------------------*)
1200 (*ImageSizePixelsDesiredTimeline*),
1, (*MeshDesired_ :: 0=NO, 1=YES*)
.5, (*OpacityDesired_ :: Nominal = 1*)
.1, (*AspectRatioDesired_ :: 0 = automatic*)
1 (*PValueAndPowerPlotMaxFlag_ :: 0=adaptive; 1=use '1' as MAX for all plots*)
]
```

Appendix I: *Sample Mathematica session* (3/3)



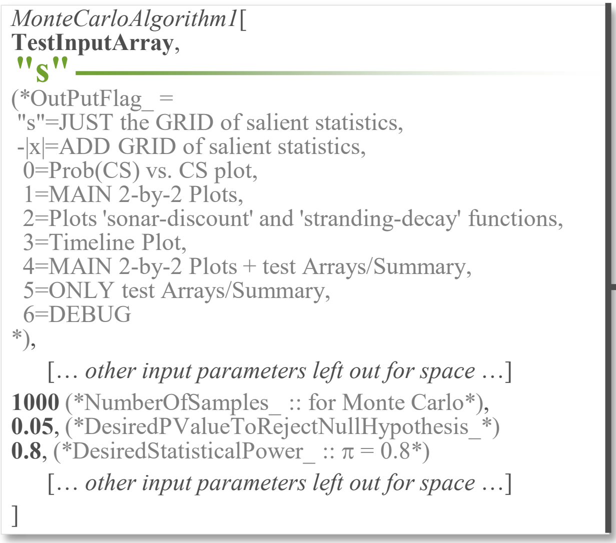
```
MonteCarloAlgorithm1[
TestInputArray,
0
(*OutPutFlag_ =
 "s"=JUST the GRID of salient statistics,
 -|x|=ADD GRID of salient statistics,
  0=Prob(CS) vs. CS plot,
  1=MAIN 2-by-2 Plots,
  2=Plots 'sonar-discount' and 'stranding-decay' functions,
  3=Timeline Plot,
  4=MAIN 2-by-2 Plots + test Arrays/Summary,
  5=ONLY test Arrays/Summary,
  6=DEBUG
*),
    [… other input parameters left out for space …]
1000 (*NumberOfSamples_ :: for Monte Carlo*),
0.05, (*DesiredPValueToRejectNullHypothesis_*)
0.8, (*DesiredStatisticalPower_ :: π = 0.8*)
    [… other input parameters left out for space …]
]
```



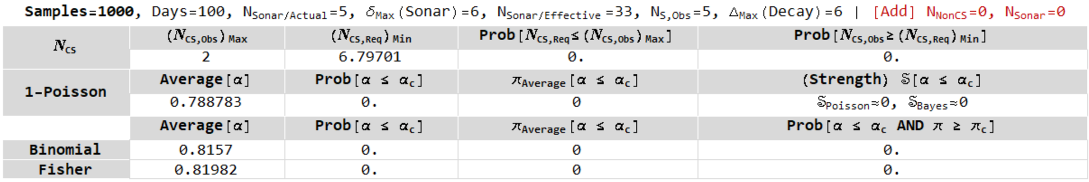
SAMPLES = 1000 | Days = 100 | Strandings = 5

$\tau = 0$, $\delta_{Max} = 6 \longrightarrow$ Sonar Days (Actual, Effective) = (5,33)

Min = 0, Ave = 0.99, Max = 2 | *Frac* ≥ $CS_{Req,Min}$(=6.80) → **0.**



```
MonteCarloAlgorithm1[
TestInputArray,
"s"
(*OutPutFlag_ =
 "s"=JUST the GRID of salient statistics,
 -|x|=ADD GRID of salient statistics,
  0=Prob(CS) vs. CS plot,
  1=MAIN 2-by-2 Plots,
  2=Plots 'sonar-discount' and 'stranding-decay' functions,
  3=Timeline Plot,
  4=MAIN 2-by-2 Plots + test Arrays/Summary,
  5=ONLY test Arrays/Summary,
  6=DEBUG
*),
    [… other input parameters left out for space …]
1000 (*NumberOfSamples_ :: for Monte Carlo*),
0.05, (*DesiredPValueToRejectNullHypothesis_*)
0.8, (*DesiredStatisticalPower_ :: π = 0.8*)
    [… other input parameters left out for space …]
]
```



Samples=1000, Days=100, $N_{Sonar/Actual}$=5, δ_{Max}(Sonar)=6, $N_{Sonar/Effective}$=33, $N_{S,Obs}$=5, Δ_{Max}(Decay)=6 | [Add] N_{NonCS}=0, N_{Sonar}=0

N_{CS}	$(N_{CS,Obs})_{Max}$	$(N_{CS,Req})_{Min}$	**Prob**$[N_{CS,Req} \le (N_{CS,Obs})_{Max}]$	**Prob**$[N_{CS,Obs} \ge (N_{CS,Req})_{Min}]$
	2	6.79701	0.	0.
1-Poisson	**Average**$[\alpha]$	**Prob**$[\alpha \le \alpha_c]$	$\pi_{Average}[\alpha \le \alpha_c]$	**(Strength)** $\mathbb{S}[\alpha \le \alpha_c]$
	0.788783	0.	0	$\mathbb{S}_{Poisson} \approx 0$, $\mathbb{S}_{Bayes} \approx 0$
	Average$[\alpha]$	**Prob**$[\alpha \le \alpha_c]$	$\pi_{Average}[\alpha \le \alpha_c]$	**Prob**$[\alpha \le \alpha_c$ **AND** $\pi \ge \pi_c]$
Binomial	0.8157	0.	0	0.
Fisher	0.81982	0.	0	0.

This page intentionally left blank.

This report was written by CNA's Resources and Force Readiness Division (RFR).

RFR provides analytic support grounded in data to inform resource, process, and policy decisions that affect military and force readiness. RFR's quantitative and qualitative analyses provide insights on a full range of resource allocation and investment decisions, including those pertaining to manning, maintenance, supply, and training. Drawing on years of accumulated individual and unit data, as well as primary data collections, the RFR toolbox includes predictive data analytics, statistical analysis, and simulation to answer optimization and what-if questions, allowing military leaders to make better informed decisions.

This report may contain hyperlinks to websites and servers maintained by third parties. CNA does not control, evaluate, endorse, or guarantee content found in those sites. We do not assume any responsibility or liability for the actions, products, services, and content of those sites or the parties that operate them.

CNA

Dedicated to the Safety and Security of the Nation

DAB-2025-U-040942-2Rev

3003 Washington Boulevard, Arlington, VA 22201
www.cna.org 703-824-2000