NATIONAL CENTRE FOR MARINE CONSERVATION AND RESOURCE SUSTAINABILITY

(Re)thinking Aquatic Non-indigenous Species Impact Assessment

Alisha Dahlstrom

Prof Chad Hewitt

Assoc Prof Marnie Campbell

(Re)thinking Aquatic Non-indigenous Species Impact Assessment

- Risk assessment and uncertainty
- Null hypothesis significance testing (NHST)
- Criticisms of NHST
- Implications for risk management and precaution
- Methods for (re)thinking impact assessment

(Re)thinking Aquatic Non-indigenous Species Impact Assessment

- **Risk assessment and uncertainty**
- Null hypothesis significance testing (NHST)
- Criticisms of NHST
- Implications for risk management and precaution
- Methods for (re)thinking impact assessment

Risk assessment

- Risk assessment: to reduce the risk of aquatic nonindigenous species (ANS) entering, establishing, spreading, and imposing consequences on core values (environmental, economic, social, cultural and human health)
- Risk estimate= Likelihood x Consequence
  - Impact assessment feeds into consequence assessment
- Risk assessment as important tool:
  1. Clearly defines the components of the hazard, e.g., accounts for all potential consequences to core values
  2. Allows for prioritization of resource use
  3. Required by WTO Sanitary and Phytosanitary (SPS) Agreement to justify national or regional biosecurity policies

Uncertainty within risk assessment

- "The existence of multiple possible future scenarios, given the current knowledge"
- Post-publication
  - Scarcity
  - Temporal and spatial variation
- Pre-publication
  - **Statistical and biological significance**

(Re)thinking Aquatic Non-indigenous Species Impact Assessment

- Risk assessment and uncertainty
- **Null hypothesis significance testing (NHST)**
- Criticisms of NHST
- Implications for risk management and precaution
- Methods for (re)thinking impact assessment

Definitions

- α – the acceptable probability of incorrectly rejecting the H_0 (or, the rate of false positives or Type I errors)
- β – the acceptable probability of incorrectly failing to reject the H_0 (or the rate of false negatives or Type II errors)
- Power – the probability that the test will correctly reject a false H_0; $1 - \beta$)
  - $\uparrow$ power $\propto \uparrow \alpha$, sample size (N), effect size (ES)
  - $\uparrow$ power $\propto \downarrow \beta$, population SD (σ)

Statistical significance

- Null Hypothesis Significance Testing (NHST)
- Decision to reject or fail to reject the H_0 based on *p*-value
- *p*-value is compared to a chosen significance level (conventionally for natural science, $\alpha=0.05$)
- At $p \geq 0.05$ one fails to reject H_0
- At $p<0.05$, one rejects H_0

(Re)thinking Aquatic Non-indigenous Species Impact Assessment

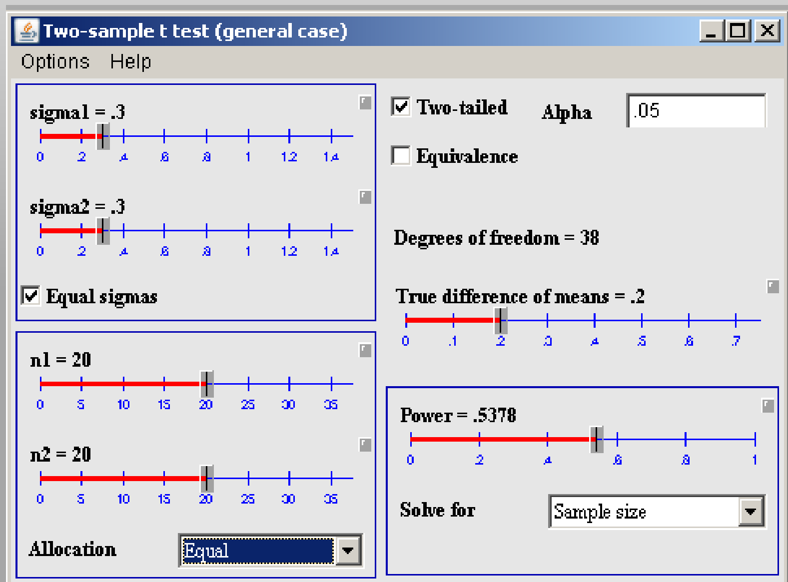
- Risk assessment and uncertainty
- Null hypothesis significance testing (NHST)
- **Criticisms of NHST**
- Implications for risk management and precaution
- Methods for (re)thinking impact assessment

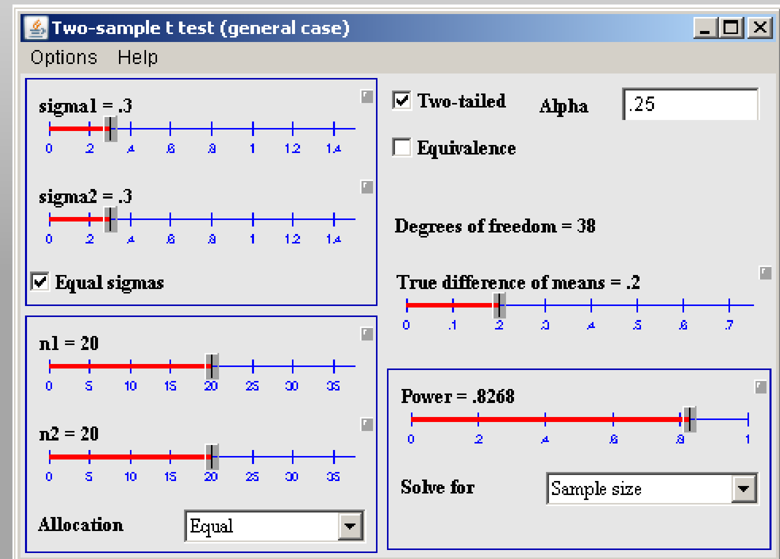
Criticisms of NHST

- Criticisms of conventional NHST include:
  - Statistical significance ≠ biological significance
  - The dichotomous ‘reject/fail to reject’ criteria is arbitrary
  - Traditional significance testing ignores effect size
  - **Focus on α=0.05 ignores or leads to low power**
  - Traditional significance testing leads to misinterpretation of results

α=0.05 ignores or leads to low power

- Focus on α may lead to experimental designs with low power





- Analyses often fail to consider low power and its consequences
  - 92% of articles from *Conservation Biology* and *Biological Conservation* in 2005 did not report statistical power (Fidler et al. 2006)
  - 98% of articles on fisheries and aquatic science with non-significant findings did not report power (Peterman 1990)

Power in ANS impact research

Species	Statistical test	Data set	Mean (ind/m^2; SD)	P-value	Power
Sargassum muticum	Mann-Whitney (non-parametric t-test)	June 1979 N=20	Removal: 0.4 (0.75) Control: 0.2 (0.37)	>0.05	**0.1783**
		Sept 1979 N=20	Removal: 0.1 (0.31) Control: 0 (0)	>0.05	**0.1687**

α=0.05 ignores or leads to low power

- Implications for Type I & II errors (i.e., more Type II errors)
- This represents, perhaps inadvertently, a social policy that protects hazards (or those that create hazards) more than the environment

α=0.05 ignores or leads to low power

- Implications for Type I & II errors (i.e., more Type II errors)
- This represents, perhaps inadvertently, a social policy that protects hazards (or those that create hazards) more than the environment **= not precautionary**

α=0.05 ignores or leads to low power

	Effect sizes and significance levels					
	r= .10		r = .30		r = .50	
N	0.05	0.1	0.05	0.1	0.05	0.1
10	19	9	17	8	13	5
20	19	9	15	6	7	2
30	18	8	13	5	3	1
40	18	8	10	4	2	-
50	18	8	9	3	-	-
100	17	7	3	-	-	-
200	14	6	-	-	-	-
300	12	5	-	-	-	-
400	10	4	-	-	-	-
500	8	3	-	-	-	-
600	6	2	-	-	-	-
700	5	2	-	-	-	-
800	4	1	-	-	-	-
900	3	-	-	-	-	-
1,000	2					

Table 2. Ratios of Type II to Type I error rates for various sample sizes, effect sizes, and significance levels (Rosnow and Rosenthal 1989).

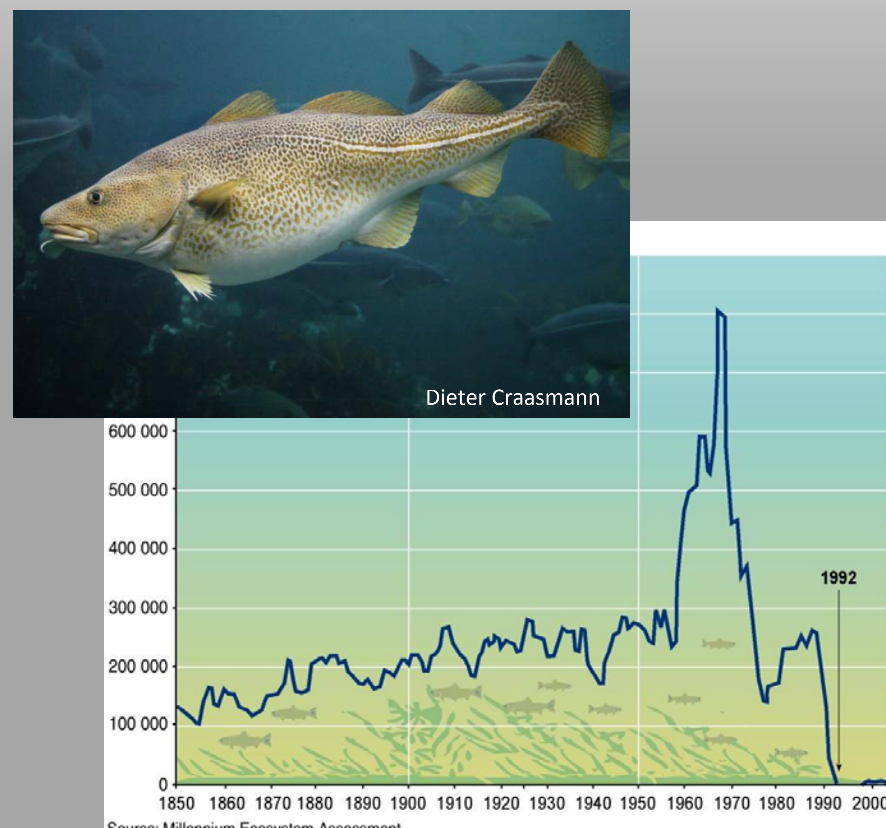
(Re)thinking Aquatic Non-indigenous Species Impact Assessment

- Risk assessment and uncertainty
- Null hypothesis significance testing (NHST)
- Criticisms of NHST
- **Implications for risk management and precaution**
- Methods for (re)thinking impact assessment

Risk management with Type I & II errors

- Consistent use of low α levels creates a bias against environmental protection and management
- For natural resource management, the cost of Type II errors is often greater than cost of Type I errors, particularly over the long-term (Page 1978; Peterman 1990; Fairweather 1991; and Toft and Shea 1983).
- ANS management depends on detecting effects

How to mitigate this bias?



Cod landings and Atlantic Cod (inset)

Precaution

- “Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation” (Rio Declaration on Environment and Development 1992)
- Increased use (e.g., Convention on Biological Diversity), particularly for impact assessment
- Use and status is controversial
  - Application criticized as unscientific and ambiguous
  - Few standardized methods for its application

β as a precautionary tool

- Establishment of fixed, low β is policy tool to incorporate the principles of precaution (Underwood and Chapman 2003)
- Type II error rate (β) can be interpreted as inversely proportional to the level of precaution

β as a precautionary tool: advantages

- Aligns the means with the ends in impact assessment (IA):
  - Traditional 'means' of IA include experimental design and analysis that minimize Type I errors
  - The 'ends' of IA include experimental design and analysis that minimize Type II errors to avoid or reduce impacts to core values
  - Focus on β better reflects 'ends' of IA
- Identified the confidence with which a researcher can draw conclusions regarding non-significant results (Fairweather 1991)
- Saves time and money from experiments whose power would have been too low to detect an impact, regardless of its presence (Fairweather 1991)
- Improve the quality of scientific enquiry and discussion (Fairbrother and Bennett 1999)

β as a precautionary tool: advantages

- Establishing the use of β as a method to incorporate precaution would also quell much of the criticism that cites the use of precaution as non-scientific or ambiguous

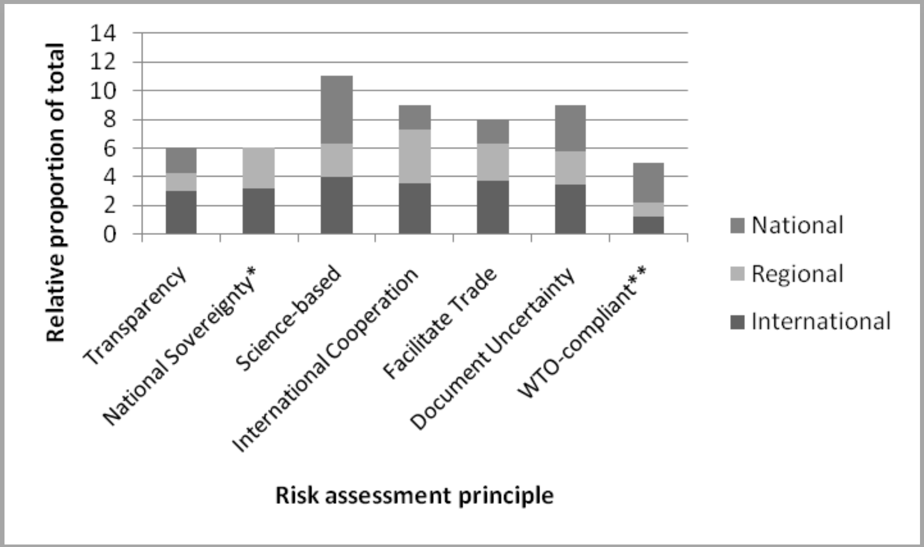


Figure 1. Risk assessment principles from national, regional and international risk assessment framework analysis. *The three national frameworks were excluded to avoid redundancy; ** WTO was excluded to avoid redundancy. (Dahlstrom et al. in prep)

(Re)thinking Aquatic Non-indigenous Species Impact Assessment

- Risk assessment and uncertainty
- Null hypothesis significance testing (NHST)
- Criticisms of NHST
- Implications for risk management and precaution
- **Methods for (re)thinking impact assessment**

(Re)thinking impact assessment methods via α and β

Several authors suggest an alternative to traditional NHST (e.g., Rotenberry and Wiens 1985; Mapstone 1995; di Stefano 2003):

1.Choose a critical effect size. (e.g., based on the acceptable impact to a core value or ALOR)

2.Values of α and β established *a priori* based on the acceptable level and/or costs of Type I (C_I) and Type II (C_{II}) errors (e.g., using a ratio $k=C_{II}/C_I$ to establish relative errors rates: $\alpha=k\beta$)

3.Sample size necessary to realize these values

Establishing critical effect size

- Associating quantitative descriptions of impact with a qualitative level (low, medium, high) allows comparison between values despite differences in the units or standards of measurement.

Descriptor	Environmental Impacts from Introduced Species
Negligible	Biodiversity impacted by non-indigenous marine species is small (<10%) compared to total impact by other hazards. Reductions in species richness and composition are not readily detectable (<10% variation).
Low	Biodiversity impacted by non-indigenous marine species is <20% compared to total impact by other hazards Reductions in species richness and composition are <20%
Moderate	...
High	...
Extreme	Biodiversity impacted by non-indigenous marine species is >70% compared to total impact by other hazards Reductions in species richness and composition are >70%

Table 3. Consequence table, with example descriptors for negligible, low and extreme adapted from Hewitt et al (2010). Many consequence matrices provide multiple descriptions of each level of impact; the attribution of an impact level to a species requires the satisfaction of only one of these.

(Re)thinking impact assessment methods via α and β

- Several authors suggest an alternative to traditional NHST (e.g., Rotenberry and Wiens 1985; Mapstone 1995; di Stefano 2003):

1. Critical effect size based on the acceptable impact given spatial and temporal considerations of core values and their respective thresholds (e.g. ALOR)
2. **Establish α and β values *a priori* (e.g., based on the acceptable level and/or costs of Type I (C_I) and Type II (C_{II}) errors, using a ratio $k=C_{II}/C_I$ to establish relative errors rates: $\alpha=k\beta$)**
3. Sample size necessary to realize these values

(Re)thinking impact assessment methods via α and β

- Several authors suggest an alternative to traditional NHST (e.g., Rotenberry and Wiens 1985; Mapstone 1995; di Stefano 2003):

1. Critical effect size based on the acceptable impact given spatial and temporal considerations of core values and their respective thresholds (e.g. ALOR)
2. Values of α and β established *a priori* based on the acceptable level and/or costs of Type I (C_I) and Type II (C_{II}) errors (e.g., using a ratio $k=C_{II}/C_I$ to establish relative errors rates: $\alpha=k\beta$)
3. **Sample size necessary to realize these values**

Challenges

- **May require the use of an alpha level greater than 0.05 (such as α=0.10 or 0.25; supported by Hayes 1987; Fairweather 1991; Strayer 1999; Spitz and Lek 1999; and di Stephano 2001)**
- These authors emphasized that α should not be set as a matter of custom, but as a matter of biological, ecological or socio-economic considerations
- **Common to have differences in "currency" between the two types of risk (e.g., due to the economic implications common to Type I errors and the environmental, social, cultural or human health implications common to Type II errors)**
- In these cases, use k=1. Setting $\alpha=\beta$ gives equal weighting to both error rates

Benefits

- The *a priori* focus on the ES, α, and β ensures:
  - Prioritization of impacts on core values;
  - Discussion and agreement on the associated impact levels that trigger action; and
  - Consideration of the consequences of both types of errors (rather than just Type I).
- Improved experimental design (and more efficient use of scare financial and other resources (Andrew and Mapstone 1987))
- An improved science/management relationship (Figure 4)

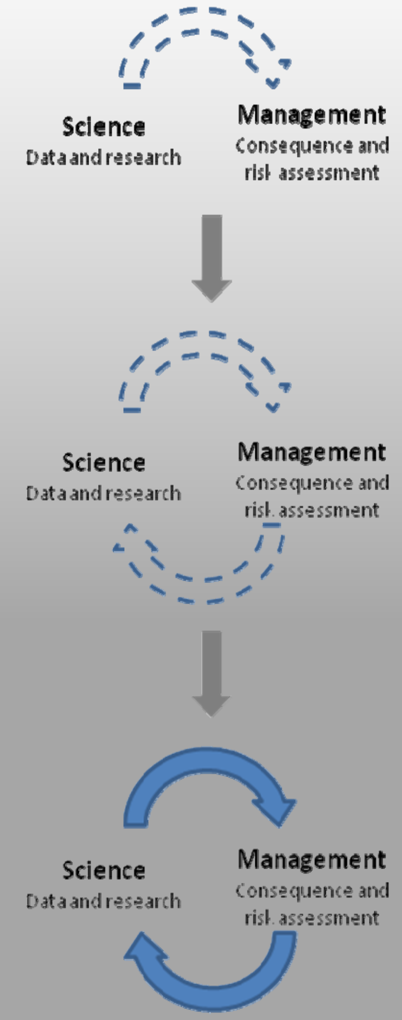


Figure 4. Improvement of information transfer with increased communication of needs.

References

Andrew, N. and B. Mapstone (1987). "Sampling and description of spatial pattern in marine ecology." Oceanography and Marine Biology: An annual review 25: 39-89.

Buhl-Mortensen, L. (1996). "Type-II statistical errors in environmental science and the precautionary principle." Marine Pollution Bulletin 32(7): 528-531.

Hewitt, C.L., M.L. Campbell, A.D.M. Coutts, A. Dahlstrom, D. Shields and J. Valentine. 2010 (in press). Assessment of marine pest risks associated with biofouling. A final report for the National System for the Prevention and Management of Marine Pest Incursions. NCMCRS Research Report. 223pp.di Stefano, J. (2001). "Power analysis and sustainable forest management." Forest Ecology and Management 154(1-2): 141-153.

di Stefano, J. (2003). "How much power is enough? Against the development of an arbitrary convention for statistical power calculations." Funct. Ecol. 17: 707-709.

Fairbrother, A. and R. S. Bennett (1999). "Ecological risk assessment and the precautionary principle." Human and Ecological Risk Assessment 5(5): 943-949.

Fairweather, P. G. (1991). "Statistical power and design requirements for environmental monitoring." Marine and Freshwater Research 42(5): 555-567.

Fidler, F., M. A. Burgman, et al. (2006). "Impact of Criticism of Null-Hypothesis Significance Testing on Statistical Reporting Practices in Conservation Biology." Conservation Biology 20(5): 1539-1544.

Hayes, J. (1987). "The positive approach to negative results in toxicology studies." Ecotoxicology and Environmental Safety 14: 73-77.

Mapstone, B. D. (1995). "Scalable Decision Rules for Environmental Impact Studies: Effect Size, Type I, and Type II Errors." Ecological Applications 5(2): 401-410.

Neyman, J. and E. S. Pearson (1933). "On the Problem of the Most Efficient Tests of Statistical Hypotheses." Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 231: 289-337.

References

Page, T. (1978). "A Generic View of Toxic Chemicals and Similar Risks." Ecology Law Quarterly 7: 207-244.

Peterman, R. M. (1990). "The Importance of Reporting Statistical Power: The Forest Decline and Acidic Deposition Example." Ecology 71(5): 2024-2027.

Peterman, R. M. (1990). Statistical power analysis can improve fisheries research and management. Ottawa, ON, CANADA, National Research Council of Canada.

Rosnow, R. and R. Rosenthal (1989). "Statistical Procedures and the Justification of Knowledge in Psychological Science." American Psychologist 44(10): 1276-1284.

Rotenberry, J. T. and J. A. Wiens (1985). "Statistical Power Analysis and Community-Wide Patterns." The American Naturalist 125(1): 164-168.

Spitz, F. and S. Lek (1999). "Environmental impact prediction using neural network modelling. An example in wildlife damage." Journal of Applied Ecology 36: 317-326.

Strayer, D. L. (1999). "Statistical Power of Presence-Absence Data to Detect Population Declines." Conservation Biology 13(5): 1034-1038.

Toft, C. A. and P. J. Shea (1983). "Detecting Community-Wide Patterns: Estimating Power Strengthens Statistical Inference." The American Naturalist 122(5): 618-625.

Underwood, A. J. and M. G. Chapman (2003). "Power, precaution, Type II error and sampling design in assessment of environmental impacts." Journal of Experimental Marine Biology and Ecology 296(1): 49-70.

United Nations General Assembly (1992). Rio Declaration on Environment and Development. The United Nations Conference on Environment and Development, Rio de Janeiro, United Nations Department of Economic and Social Affairs (DESA).

Weinberg, J., H. Caswell, et al. (1986). "Demographic importance of ecological interactions: how much do statistics tell us?" Marine Biology 93: 305-310.

The significance (meaning) of (statistical) significance is significant (important).

Questions?

I would like to acknowledge the AMC John Bicknell Research Scholarship for funding support.