Experimental design in the deep-sea
to answer basic deep-sea mining questions: an initial examination

Jeff Ardron, Daniel Jones, Erik Simon-Lledó

U Southampton, NOC, NERC;
Commonwealth Secretariat

PRZ / IRZ workshop, Berlin, 27-29 Sept. 2017

National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

The Commonwealth

NERC SCIENCE OF THE ENVIRONMENT

Three Questions:
in the Clarion Clipperton Fracture Zone…

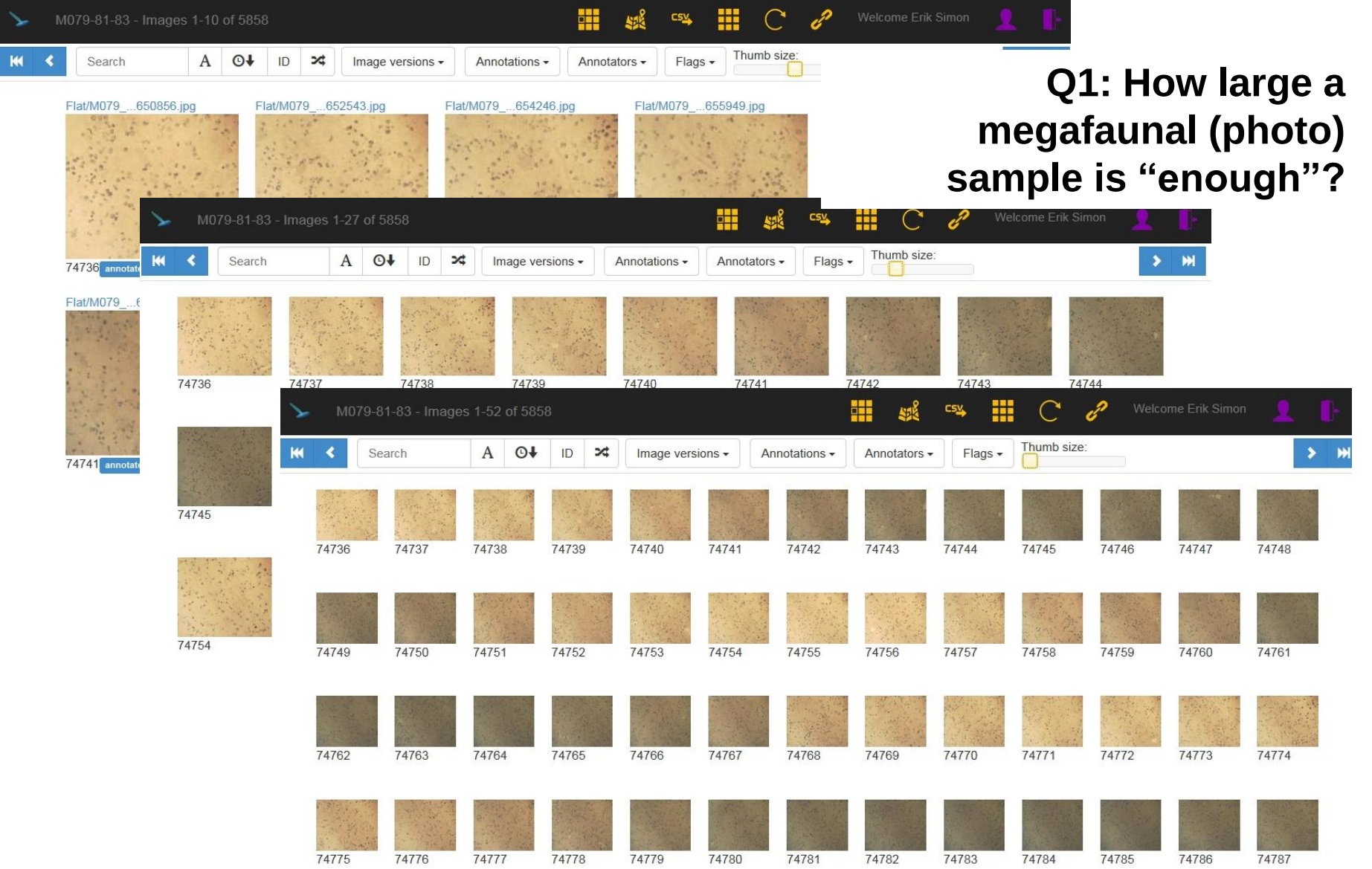
1. How large a sample is “enough”? (= Confidence)

2. How many samples is “enough”? (= Power)

3. What effect size is “enough”? (= Importance)

(PS: note the insignificance of significance testing if these three questions are not addressed…)

Q1: How large a megafaunal (photo) sample is “enough”?

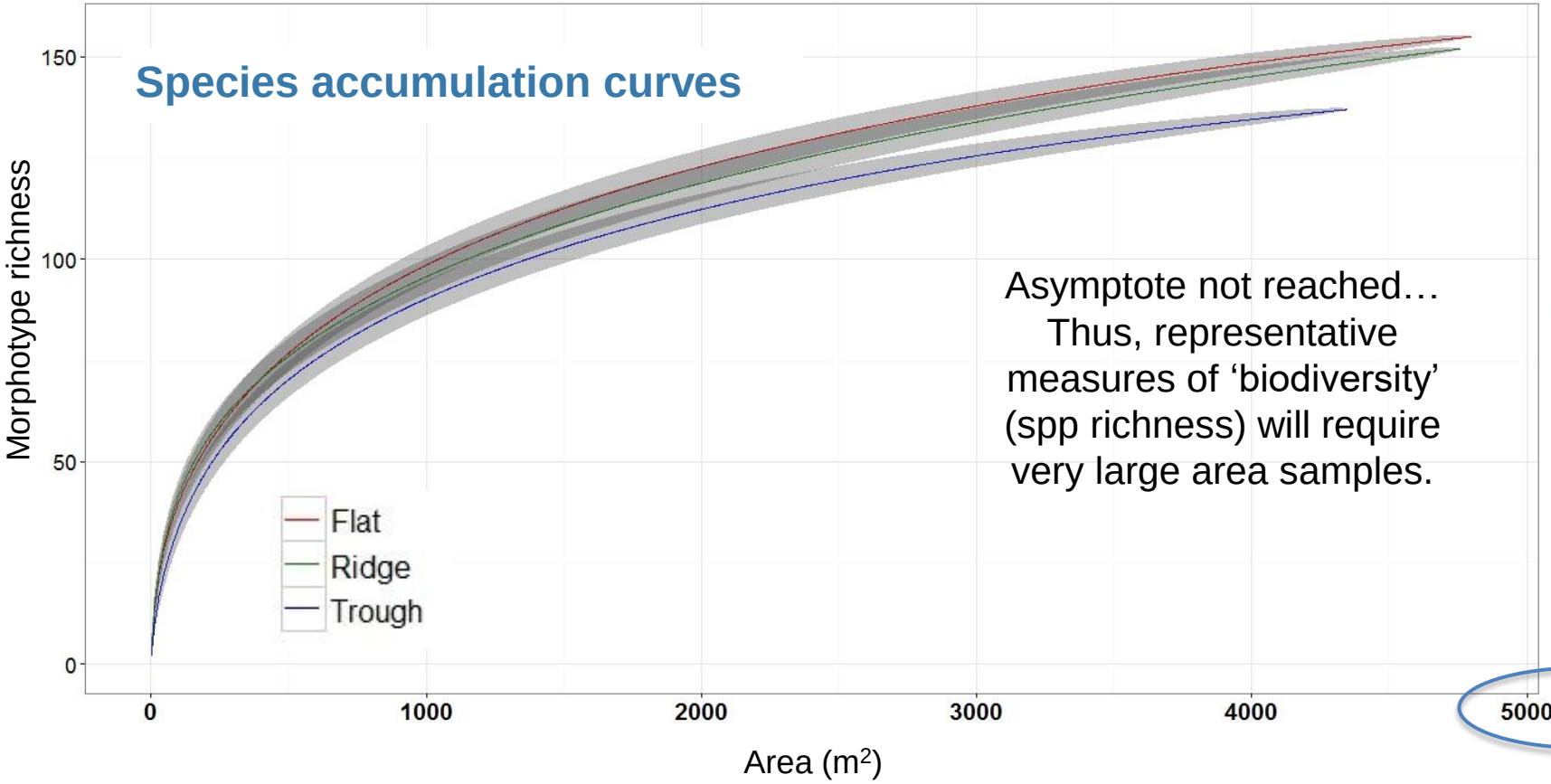


National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

The Commonwealth

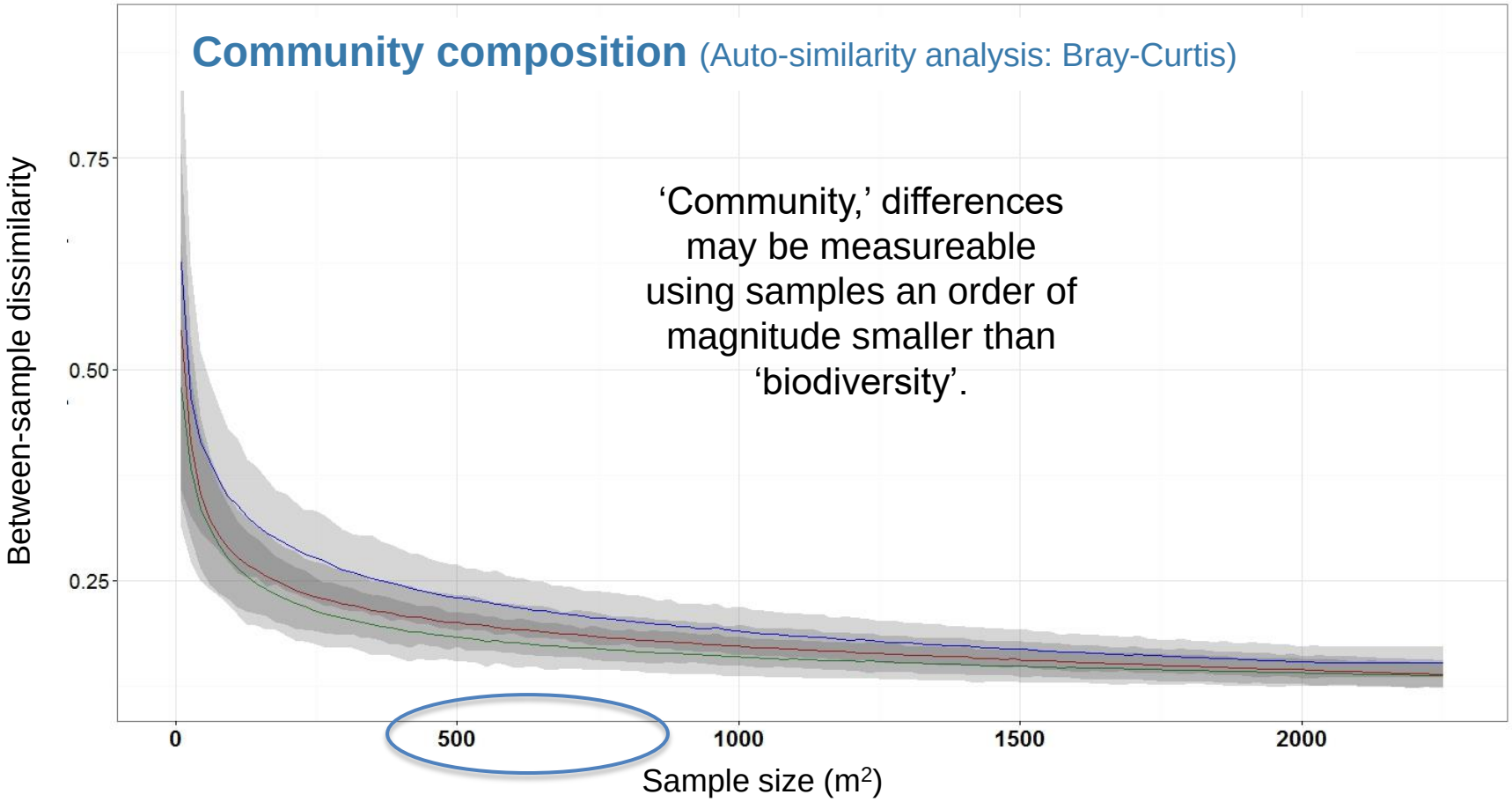
NERC SCIENCE OF THE ENVIRONMENT

A: It depends on what you want to measure…



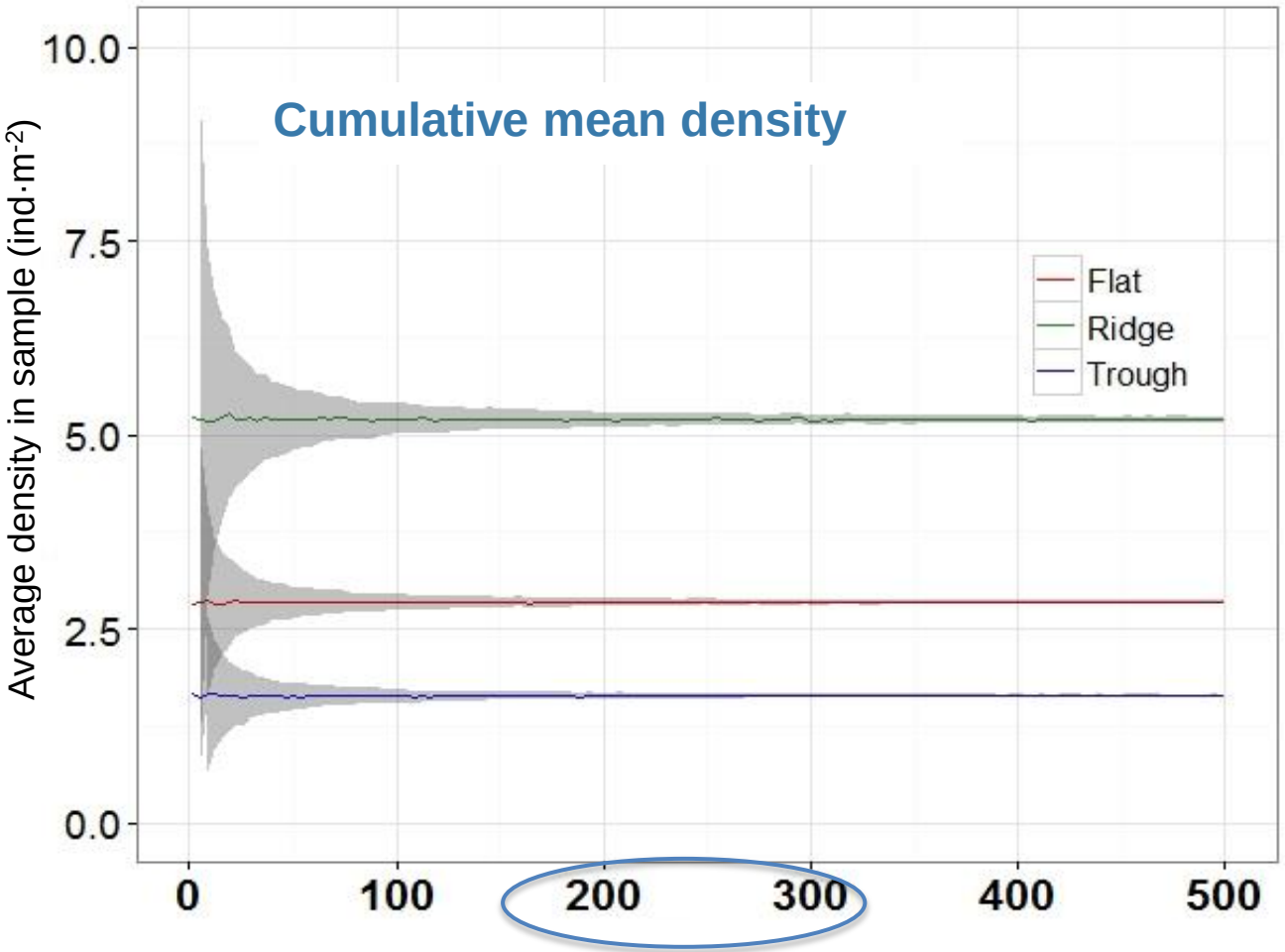
(Unpublished data from the CCZ)

A: It depends on what you want to measure…



(Unpublished data from the CCZ)

A: It depends on what you want to measure…

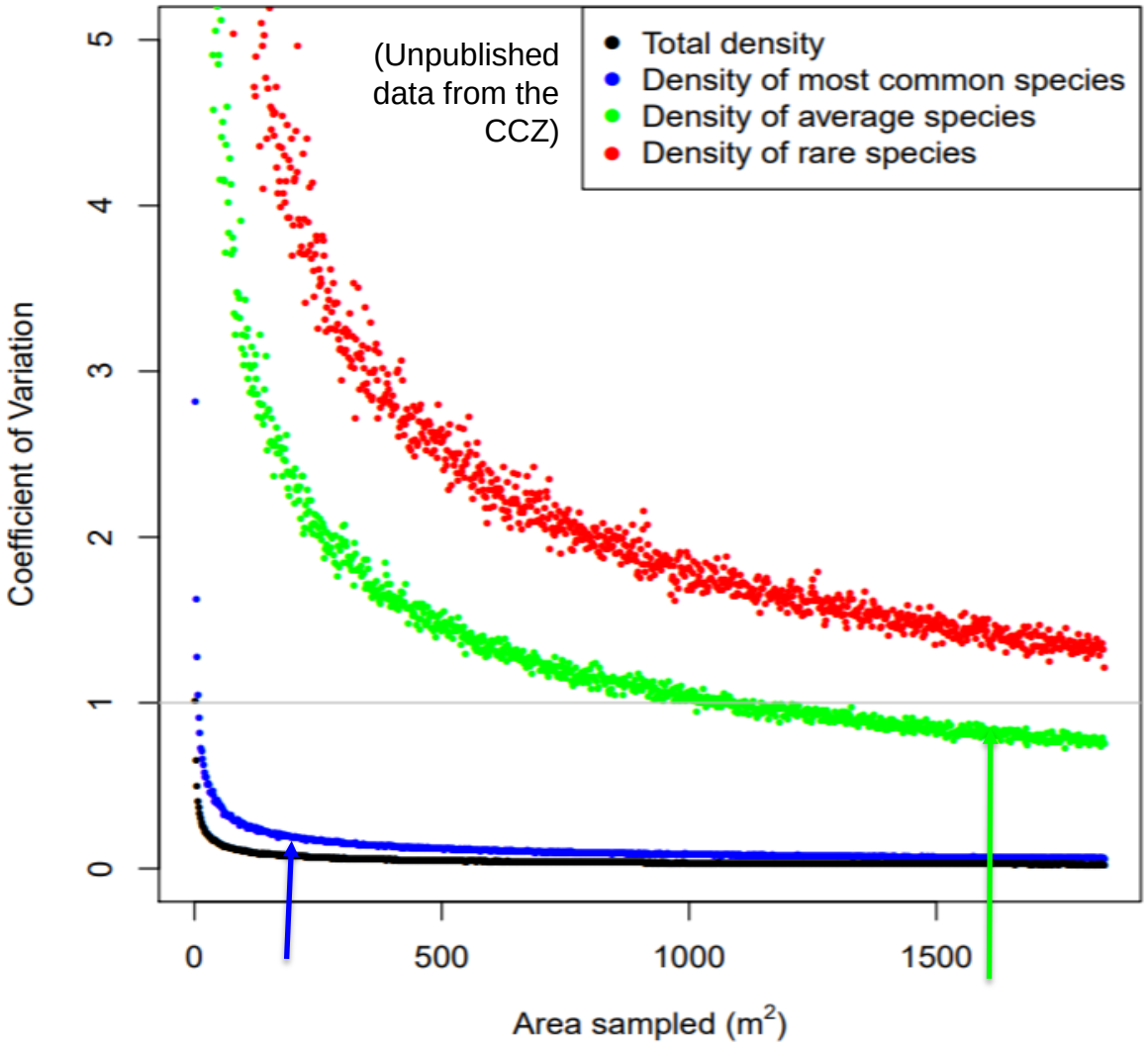


(Unpublished data from the CCZ)

Cumulative megafaunal density appears measureable using samples two or three times smaller than for 'community…'

(And 20-30x smaller than for spp richness…)

But… densities of individual species (red & green dots) require much larger sample areas than cumulative densities



One typical conservation strategy is to monitor & protect less common large species, but it is unlikely in the CCZ, due to sampling requirements (**red: 25th percentile of abundance**).

Monitoring and thus protecting 'average' species (**green: median abundance**) is more tractable.

?

Q2: How many samples is “enough”?

pression, *Journal of Clinical Psychiatry*, 51, 61–69 (1990).

11. L.R. Baxter, Jr., J.M. Schwartz, B.H. Guze, J.C. Mazziotta, M.P. Szuba, K. Bergman, A. Alazraki, C.E. Selin, H.K. Freng, P. Munford, and M.E. Phelps, *Obsessive-compulsive disorder vs. Tourette's disorder: Differential function in subdivisions of the neostriatum*, paper presented at the annual meeting of the American Co
chopharmacology, San Jua
(December 1991).

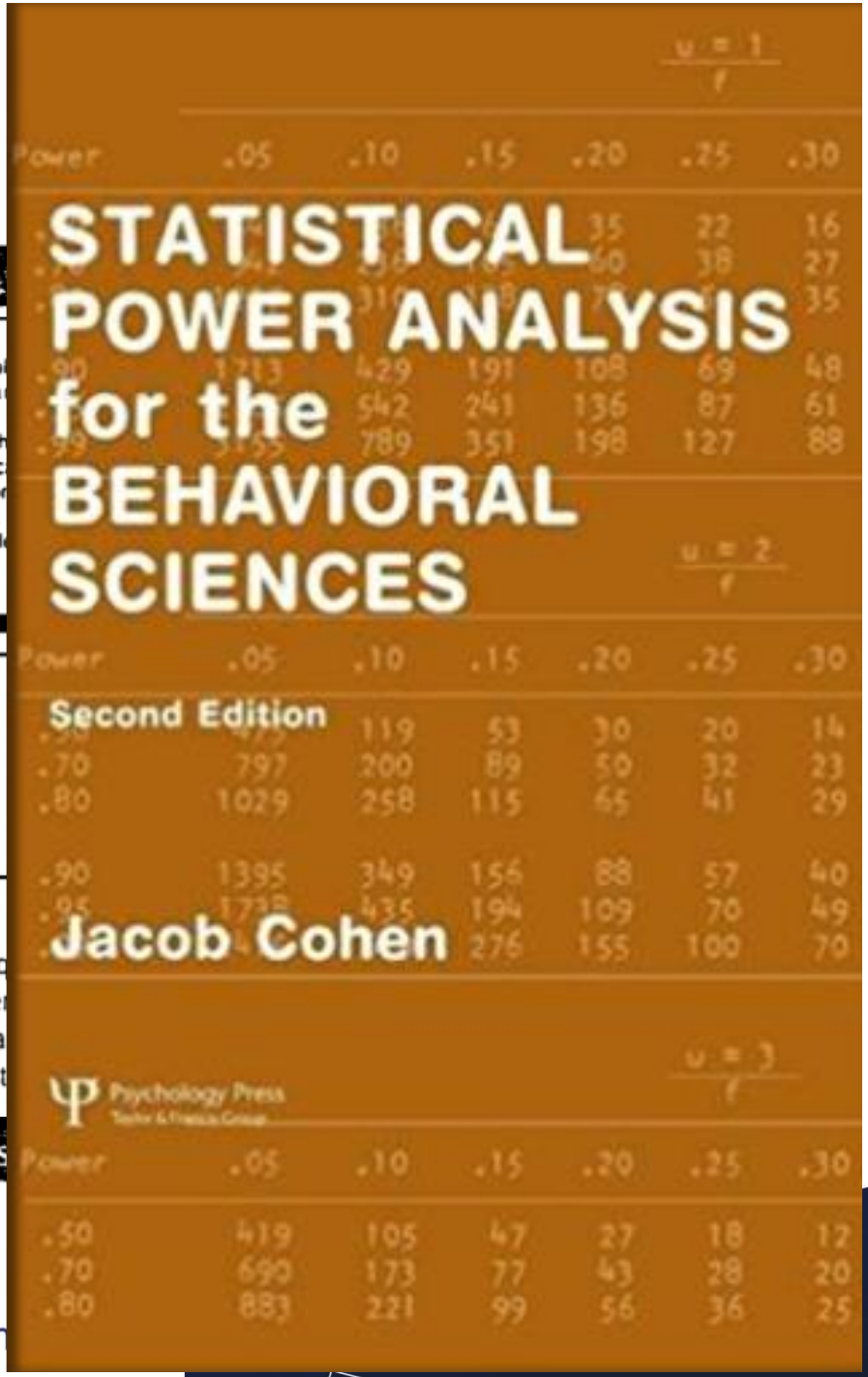
12. E.M. Reiman, M.E. Raich
Herscovitch, and E. Robins, A foc
ity in panic disorder, a severe fo
ture, *310*, 683–685 (1984); E.M.
ichle, E. Robins, F.K. Butler, P. H

Statistical Power Analysis

Jacob Cohen

The power of a statistical test of a null hypothesis (H_0) is the probability that the H_0 will be rejected when it is false, that is, the probability of

which *r* indeed does ec
searchers risk mistake
the H_0 when it is true, a
whose rate (.05) is cont



National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

The Comm

Q2: How many samples is “enough”?

Cohen's *d* [edit]

Cohen's *d* is defined as the difference between two means divided by a standard deviation for the data, *i.e.*

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}.$$

Wikipedia; 20 Sept. 2017

Jacob Cohen defined *s*, the pooled standard deviation, as (for two independent samples):[7]:67

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Nerdy word of caution over Wikipedia…

Actually, Cohen’s d does not subtract 2.

This was a later suggestion by Hedge to compensate for smaller numbers of samples –a variant called ‘Hedge’s g’ – which is probably the case in the deep-sea.

Q2: How many samples is “enough”?

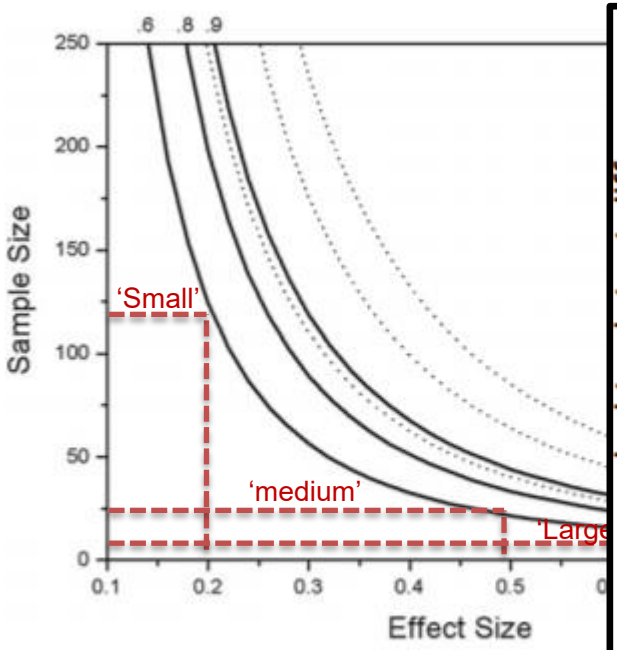


Fig. 2. Sample size as a function of effect size of *t*-tests for diff
(0.6, 0.8, and 0.9). Continuous lines are for *t*-test with paired-sa
for *t*-test with independent samples. α for all curves is 0.05.

Contents lists available at ScienceDire

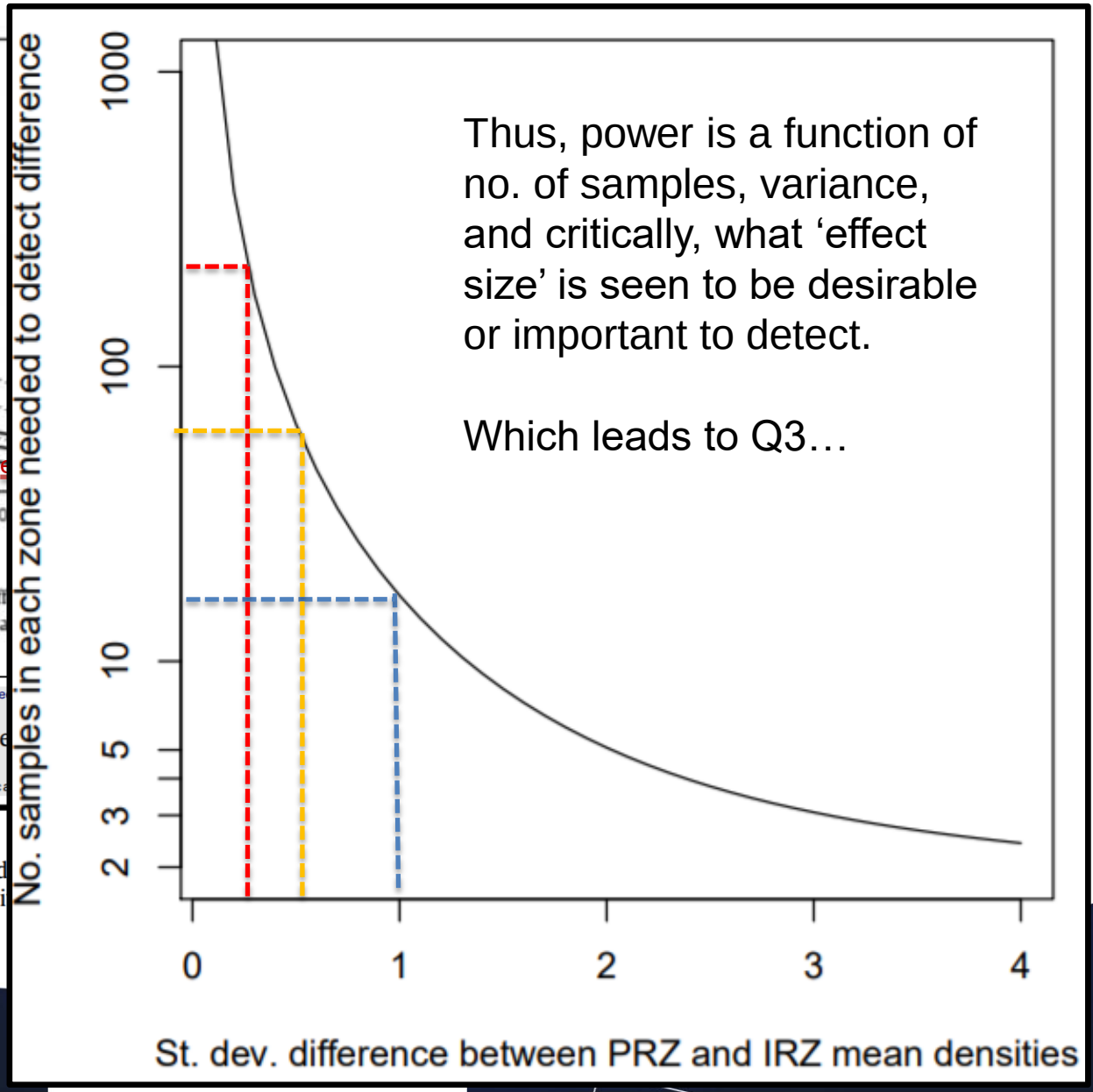
Building and Environme

ELSEVIER

journal homepage: www.elsevier.com/loca

Application of statistical power analysis – How to d
sample size in human health, comfort and producti

Li Lan, Zhiwei Lian*



Thus, power is a function of no. of samples, variance, and critically, what ‘effect size’ is seen to be desirable or important to detect.

Which leads to Q3…

Q3: What effect size is “enough”?

There is no ‘right answer’, but…

1. Cohen’s guidelines were based on psychology and human testing; it is unknown if these are transferable to deep-sea ecology.

2. Effects on deep-sea ecology will be limited to the parameters that can be measured with tractable size and number of samples. Thus, some critical questions (e.g. ‘biodiversity’) may not be directly sampled in a statistically meaningful way (will need modelling and macro-ecological indicators).

3. BACI effects of most interest (e.g. effects of deposition of fines on communities, and signs of natural recovery in abundance) could conceivably be in the realm of 0.5 SD magnitude, and require about 75 samples (of appropriate size).

PS: A non-significant result without enough power tells us nothing; and a significant result arising from just a few samples (i.e. low power) is going to be very obvious anyway…

Three Closing Thoughts

1. Measuring some parameters will require larger sample areas than others.
    - *Selection of parameters will be a balance of cost versus criticality (legal obligations and risk).*

2. Power analyses are necessary to separate out meaningful from statistically 'trivial' or inconclusive significance results.
    - *Power analyses will need to be done beforehand, to determine the appropriate experimental design, esp. number of samples. Power analyses, however, require comprehensive baseline data.*

3. Determining what is a meaningful effect size for a given variable is both a scientific and a policy question. Answers may vary according to the risk of 'serious harm'.
    - *Agreement on effect sizes will be necessary in order to determine the experimental design and management responses. Examining Cohen's recommendations and the discussion since, could be the starting point.*