

Issues In Single-Subject Research

K. P. Kearns

ASHA CPRI

7.10.09



“Data speak, not men..”

- “Designs have inherent rigor but not all studies using a design are rigorous”
 - (Randy; yesterday)
- “Illusion of strong evidence...”
 - (McPeck & Mosteller, 1978)

Effects of Interpretation Bias on Research Evidence (Kaptchuk, 2003; BMJ)

- “Good science inevitably embodies a tension between the empiricism of concrete data and the rationalism of deeply held convictions.
- ...a view that science is totally objective is mythical and ignores the human element..”

Single-subject designs:

- Single subject experimental designs are among the most prevalent used in SLP treatment research
 - (Kearns & Thompson, 1991; Thompson, 2005; Schlosser et al ,2004).
- Well designed SS studies are now commonly published in our journals as well as in interdisciplinary specialty journals
 - (Psychology, Neuropsychology, Education , PT, OT...).
- Agencies, including NIH, NIDDR etc., commonly fund conceptually salient and well designed SS treatment programs (Aphasia, AAC, autism..).
- Meta-analyses have been employed to examine the overall impact of SS studies on the efficacy and efficiency of interventions
 - (Robey, 1999; ..)

Single-subject designs:

- Quality indicators for SS designs appear to be less well understood than for group designs
 - (Kratichwill & Stoiber, 2002; APA Div. 12; Horner, Carr, Halle, et al, 2005):
- Common threats to internal and external validity persist in our despite readily available solutions. (Schlosser, 2004; Thomson, 2005)

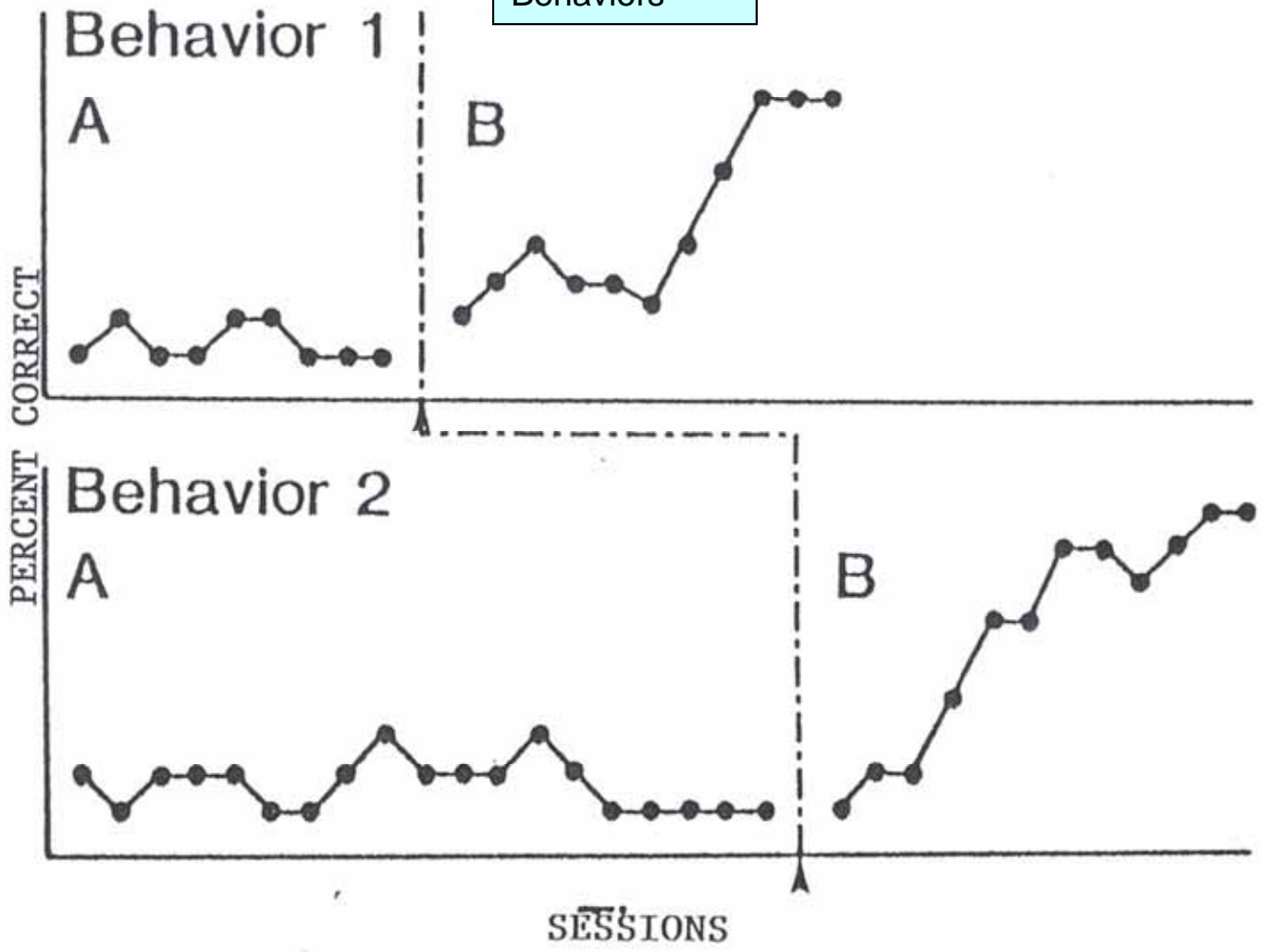
Purpose:

- Brief introduction to SS designs
- Identify elements of SS designs that contribute to problems with internal validity/ experimental control- reviewer's perspective
- Discuss solutions for some of these issues; ultimately necessary for publication and external funding

Single-subject experimental designs: Obligatory Introduction

- Experimental not observational:
 - Subjects “serve as their own controls”; receive both treatment and no-treat conditions
 - Juxtaposition of Baseline (A) phases with Treatment (B) phases provides mechanism for experimental control (internal validity)
 - Control is based on within and across subject replication

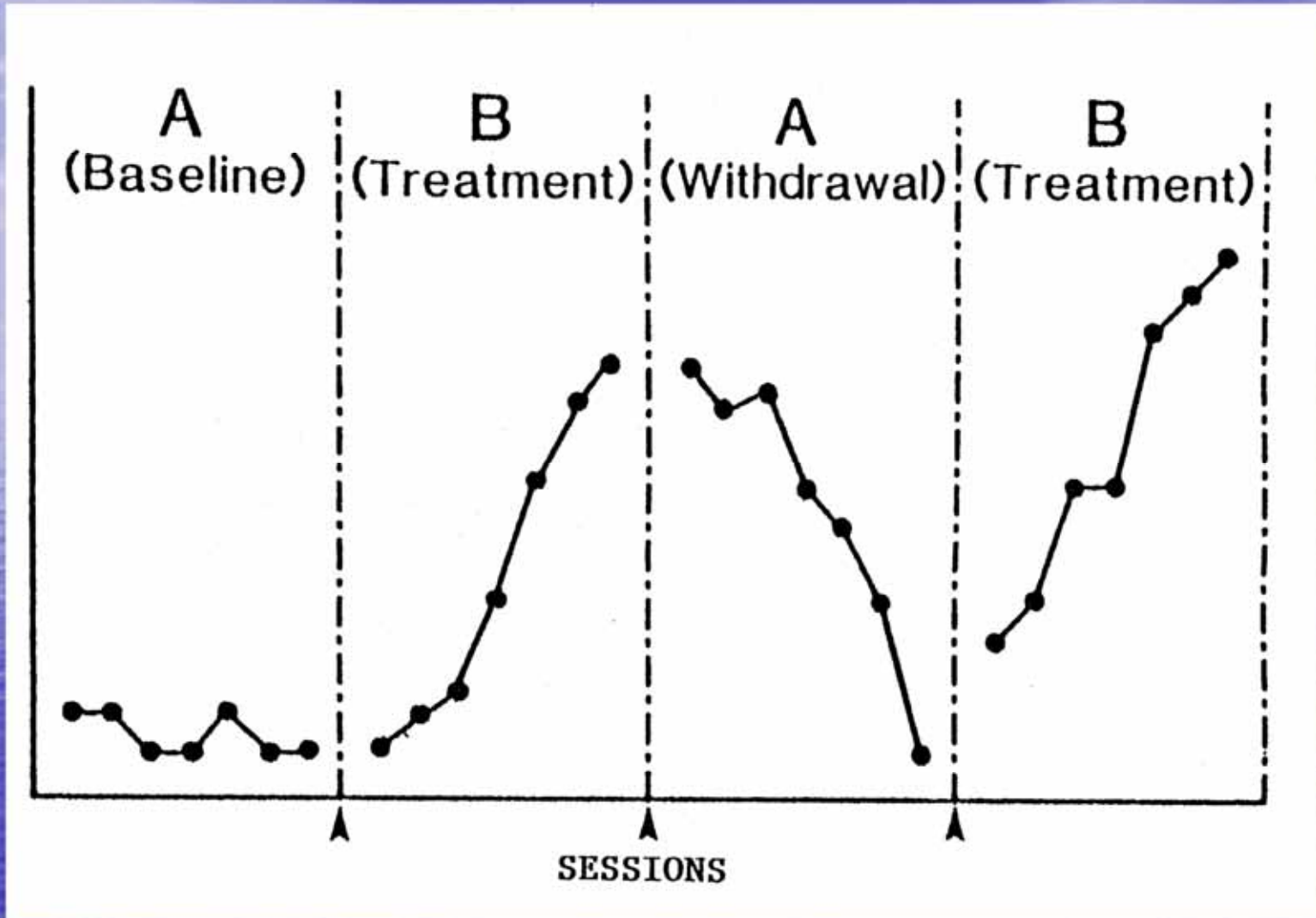
Multiple-Baseline:
Across Behaviors



Common SS Design Strategies

- Treatment vs No-treatment comparisons
 - Examine efficacy of treatment relative to no tx
 - Multiple baselines/ variants; Withdrawal/ reversals
- Component Assessment
 - Relative contribution of treatment components
 - Interaction Designs (variant of *reversals*)
- Successive Level Analysis
 - Examine successive levels of treatment
 - Multiple Probe; Changing Criterion
- Treatment - Treatment Comparisons
 - Alternating Treatments (mixed m b)

ABAB Withdrawal Design



**ATD- MB
comparison:
Broca's aphasia**

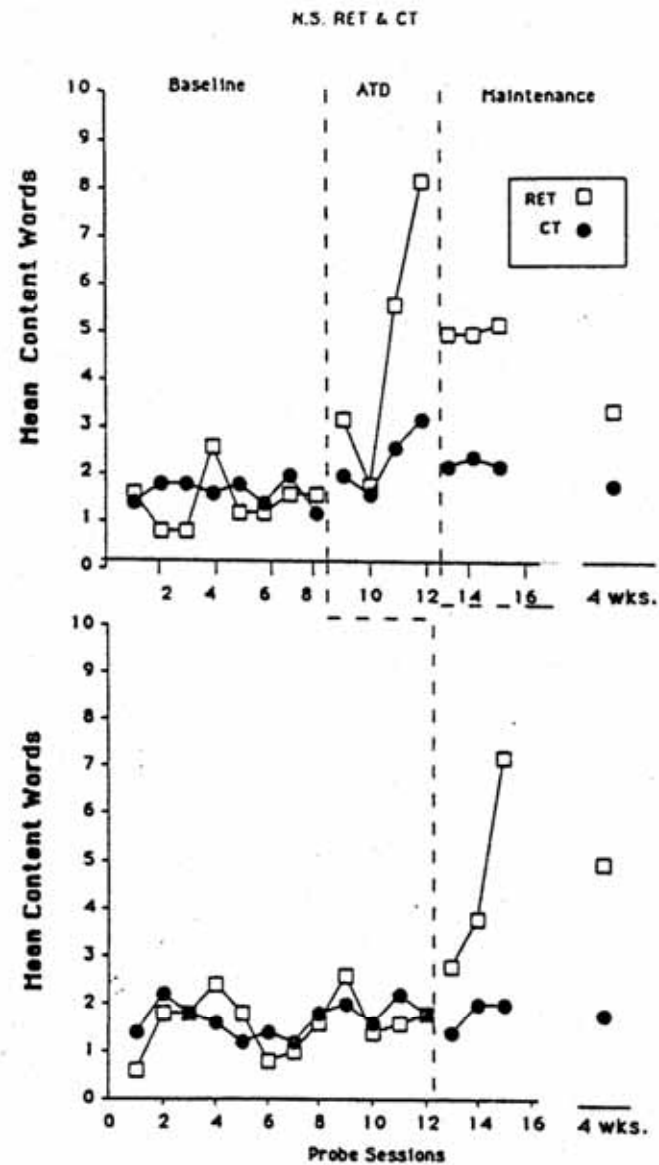


Figure 21.5. Mean number of content words produced by N. S. on response elaboration training (RET) probes (open squares) and convergent treatment (CT) probes (closed circles) during baseline, alternating treatments, and maintenance phases.

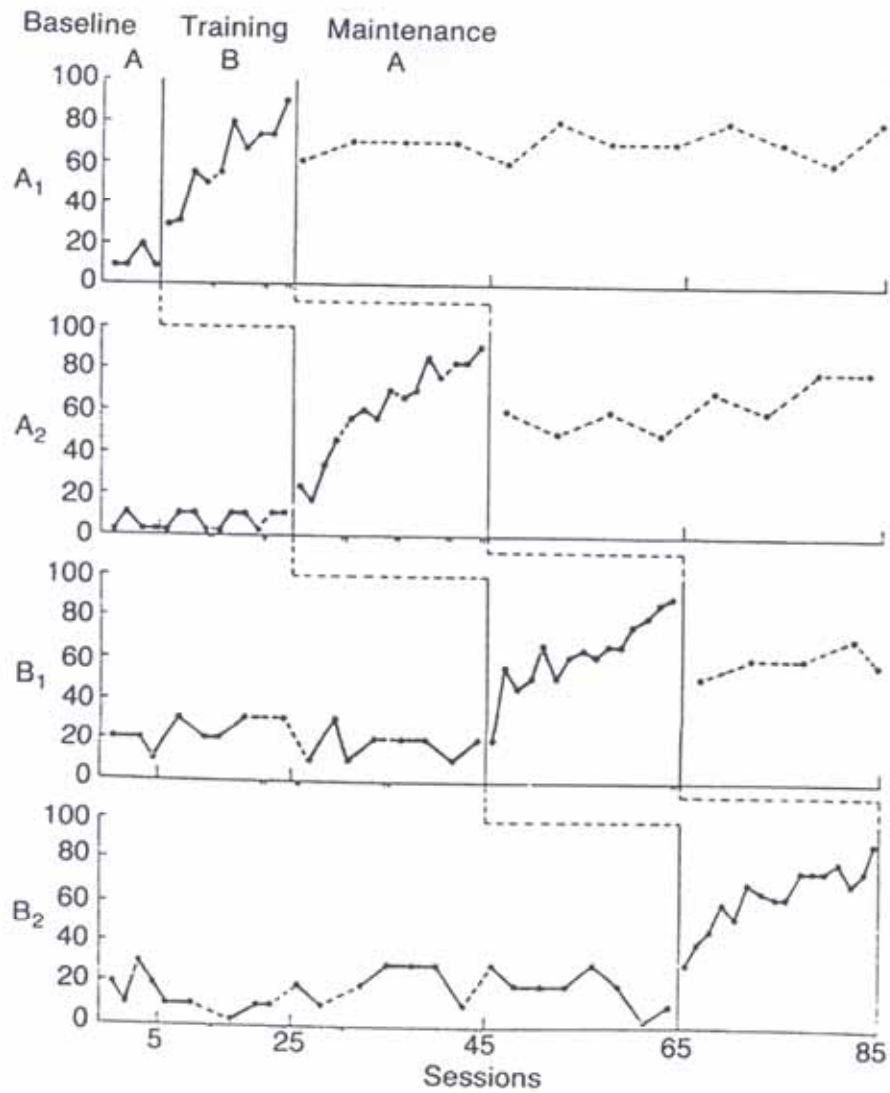
Single-subject experimental designs

- Internal Validity:
 - Operational specificity; reliability of IV, DV; tx integrity; appropriate design..
 - Artifact, Bias
 - Visual analysis of “control”
 - Loss of baseline (unstable; drifting trend..)
 - W/I and across phase changes: L, S, T...
 - Replicated treatment effects
 - three demonstrations of the effect at three points in time

Visual-Graphic Analysis

- Within and across phase analysis of
 - Level (on the ordinate; %..)
 - Slope (stable, increasing, decreasing)
 - Trend over time (variable; changes with phases; overlapping..)
- Overlap, immediacy of effect, similarity of effect for similar phases
- Correlation of change and phase change

Retrospective of An Experimental Analysis



(Thompson, Kearns, Edmonds, 2006)

I. Research on Visual Inspection of S-S Data

(Franklin et al, 1996; Robey et al, 1999)

- Low level of inter-rater agreement
 - De Prospero & Cohen (1979) Reported $R = .61$ among behavioral journal reviewers
- Reliability and validity of visual inspection can be improved with training (Hargopian et al, 1997)
- Visual aids (trend lines) may have produced only modest increase in reliability
- Traditional statistical analyses (eg. Binomial test) are highly affected by serial dependence (Crosbie, 1993)

Serial Dependence/Autocorrelation

- The level of behavior at one point in time is influenced by or correlated with the level of behavior at another point in time
- Autocorrelation biases interpretation and leads to Type I errors (falsely concluding a tx effect exists; positive autocorrelation) and Type II errors (falsely concluding no tx effect; negative autocorrelation)
- Independence assumption

Solutions:

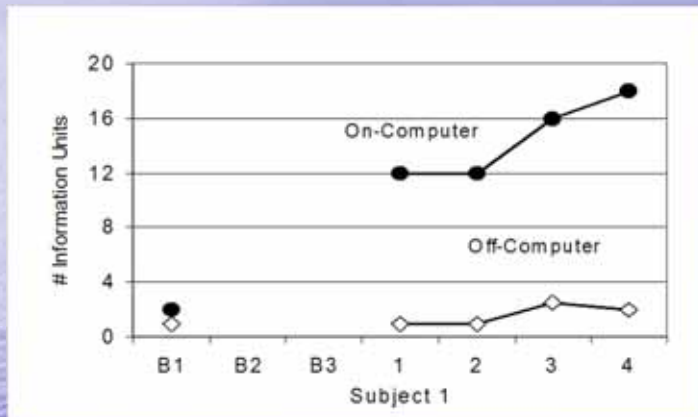
- ITSACORR: A statistical procedure that controls for autocorrelation (Crosbie, 1993)
- Visual Inspection and Structured Criteria (Fisher, Kelley & Lomas, 2003; JABA)
- SMA bootstrapping approach (Borckhardt, et al, 2008; AM Psychologist)
 - <http://clinicalresearcher.org>

II. Baseline measures

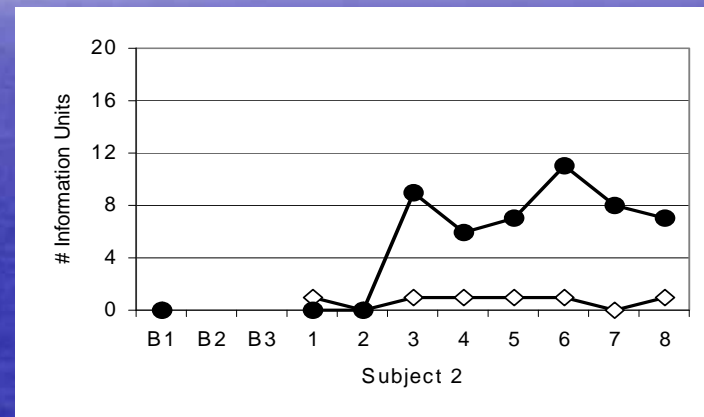
- Randomize order or stimulus sets/ conditions
- “All” treatment stimuli need to be assessed in baseline
- Establish equivalence for subsets of stimuli used as representative
- Avoid false baselines
- Apriori stability decisions greatly reduce bias
- *At least 7 baseline probes may be needed for reliable and valid visual analysis*

Statistical conclusion validity?

S1



S2



- S1 ITSACORR results were ns
- S2 ITSACORR results were sig ($F \leq .05$)
- Too few data points for valid analysis

III. Intervention

- Explicit steps; directions....a Manual
- Control for order effects
- Reliability
- Assess integrity of intervention (see Schlosser, 2004)
- One variable rule
- Is treatment intensity: sufficient; typical?
- Dual criteria for termination of treatment
 - Performance level (e.g. % correct)
 - Maximum allowable length of treatment (but not equal phases)

IV. Dependent Measures

- Use multiple measures
- Try not to collect during treatment sessions
- Probe often (weekly or more)
- Pre-train assistants the scoring code and periodically check for “drift”
- Are definitions specific, observable and replicable?

V. Reliability

- Reliability for both IV and DV
- Obtain for each phase of the study and adequately sample
- Control for sources of bias including drift and expectancy (ABC's)
- Use point to point reliability when possible
- Calculate probability of chance agreement; critical for periods of high or low responding
 - Occurrence and non occurrence reliability

VI. Apriori decisions

- Failure to establish and make explicit criteria for guiding procedural and methodological decisions prior to change is a serious threat to internal validity that is difficult.
 - Participant selection/ exclusion criteria (report attrition)
 - Baseline variability, length...
 - Phase changes
 - Clinical significance
 - Generalization

VII. Consider clinically meaningful change:

- SS and "clinical significance"
- Clinical significance can not be assumed from our perspective alone
 - Change in level of performance on any outcome measure, even when effects are large and visually obvious or significant, is an insufficient metric of the impact of experimental tx on our participants/ patients

Minimal Clinically Important Difference (MCID)

- “the smallest difference in a score that is considered worthwhile or important”

Hayes & Wooley, 2000

Responsiveness of Health Measures Husted et al (2000)

1. Distribution based approaches examine Internal responsiveness
 - using distribution/ variability of initial (baseline) scores to examine differences (e.g. Effect size)
2. Anchor based approaches examine External responsiveness
 - by comparing change detected by a dependent measure with an external criterion. For example, specify a level of change that meets "minimal clinically important difference" (MCID).

Anchor-based Responsiveness

measures (see Beninato, et al Archives of PMR, 2006)

- Use external criterion as “anchor”
 - Compare change score on outcome measure to some other estimate of important change
 - Patient’s/Family estimates
 - Clinician’s estimates
 - Necessary to complete the EBP triangle?

Revisiting Clinically Important Change (Social Validation)

- When the perceived change is important to the patient, clinician, researcher, payor or society (Beaton et al ,2000)
- Requires that we extend our conceptual frame of reference beyond typical outcome measures and distribution based measures of responsiveness

"Time will tell"

(M. Planck, 1950)

"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die."

in Kaptchuk, (2003)



“Here. We promised you a big lab.”

