

Sample Size Determination for Number of Patient Interviews When Developing a PRO Using Qualitative Research Based on the 2009 FDA PRO Guidance

Professor Chris Barker, Ph.D.

Consultant

www.barkerstats.com

2018 ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop
Washington D.C.
September 12-14.

Excerpt from FDA PRO Guidance on Sample Size (# of Interviews)

- .. We cannot provide recommendations for the number or size of the individual patient interviews or focus groups for establishing content validity.
- The sample size depends on the completeness of the information obtained from analysis of the transcripts. Generally, the number of patients is not as critical as interview quality and patient diversity included in the sample in relation to intended clinical trial population characteristics.
- Today: terminology “sample size” I mean:
Number of interviews,

Today –

- Define Type I error in interviews for PRO development
- Statistical /probability based definition of “saturation”
- A worked example with published data
- Sample size considerations
 1. Classical statistical Methods
 2. Capture- Recapture (CRC) framework
 3. On-going- Bayesian CRC
- Roadblocks – availability of “interview datasets” in public domain

Today: Miscellaneous jargon

- interviews for a PRO result in “codes” or “concepts”
- An instrument such as the SF-36 to arise from at least “36 codes” or “36 concepts”

The Background

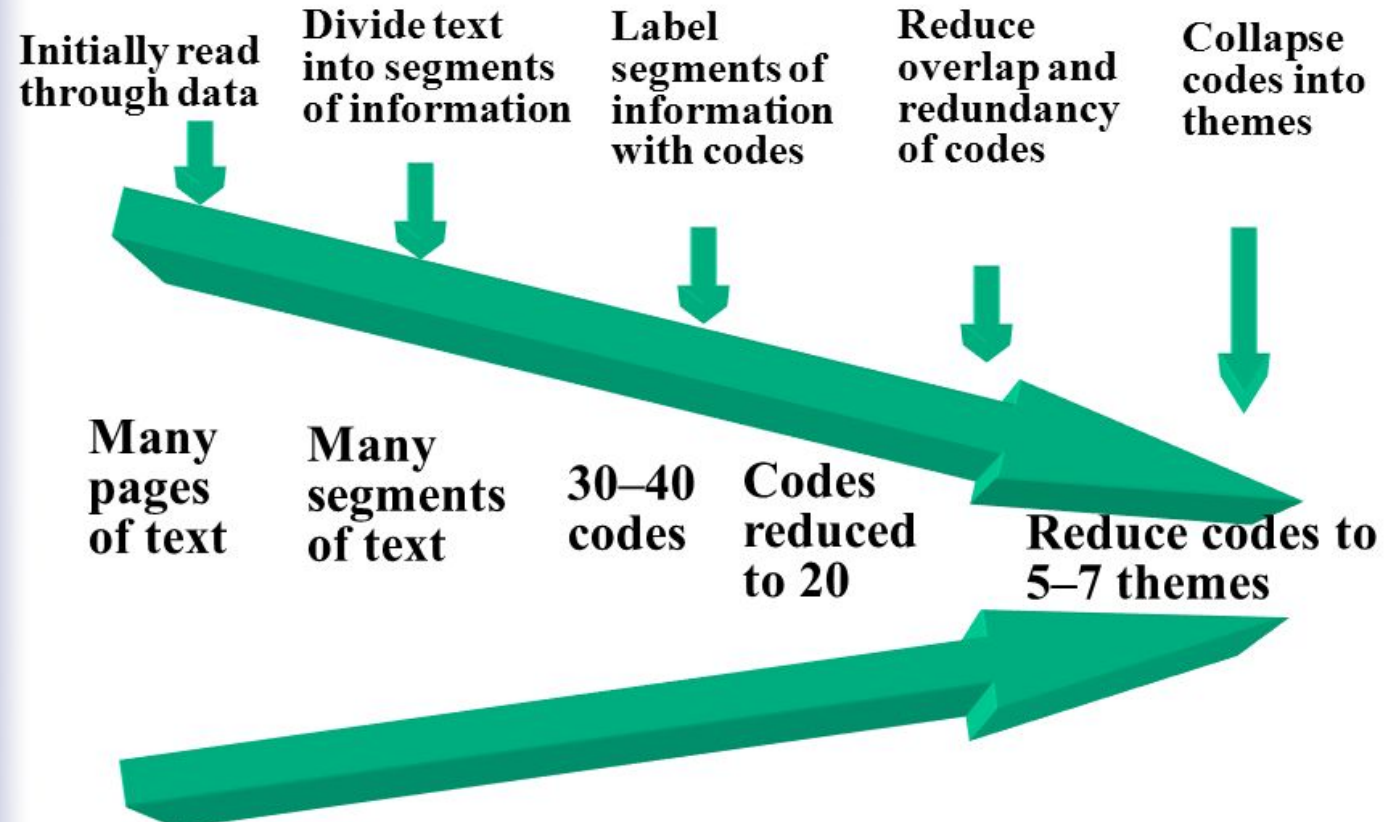
- Teleconference CEO, CMO, Clinops ,a (never-to-be-named) PRO vendor "V:", and CB (me!)
- CEO: "How many more interviews?"
- Vendor: "...as soon as we achieve saturation..."
- CEO: "What is saturation?"
- V: "zero new codes"
- CB: "do interviews stop at first occurrence of 0 or maybe a couple more for possible statistical noise?"
- V: "No"
- CB: "and how many codes expected?"
- V: "We Don't Know"
- CB: "how do we know its not a Type I error?"
- Prominent PRO expert explains "agency frequently asks for more interviews"

Literature Review: Saturation and Sample Size (number of interviews)

- Define “saturation” as “zero new concepts”
a.k.a. “First interview with saturation”
- Three consecutive interviews with zero new codes
- Expert Judgement “additional interviews would be counter productive”
- SAMPLE SIZE
 - Ethnography/ethnoscience “30 -60 interviews”
 - Grounded Theory “30-50 interviews”
 - Phenomenology “5-25 interviews”
 - All Qualitative research “at least 15 interviews”
 - Funded (\$) research ‘time limited’ interviews ranged from 1 – 95
 - ETC.
 - There do not appear to be formal statistical methods for sample size or saturation

Conceptual overview of “coding” in qualitative research

A Visual Model of the Coding Process in Qualitative Research



ROADBLOCKS to PRO interview Sample size Research

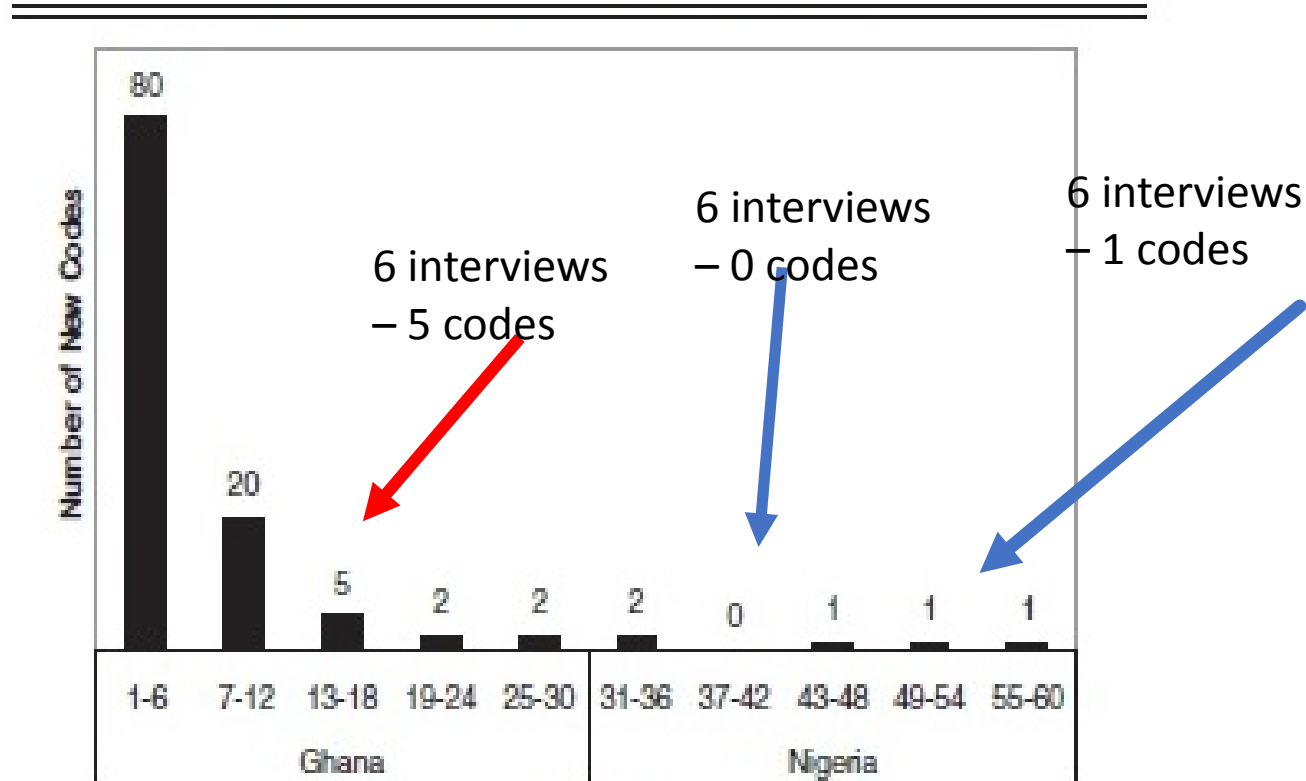
1. “interview” datasets for PRO development not routinely published or in public domain
2. Data set today – from a literature search – Guest , Bunce.
 1. Original interview level data not available per Dr. Guest

Guest, Bunce- Figure 1

Interviews in groups of 6. Interview level data not available

Guest et al. / HOW MANY INTERVIEWS ARE ENOUGH? 67

FIGURE 1
Code Creation over the Course of Data Analysis



Interviews
Chris Barker, Ph.D.

www.barkerstats.com

Key assumptions of my methodology

- Before interviewing a –fixed- and unknown total number of codes
- One interview per subject (patient...)
- Interviews in chronological order
- First interview elicits new codes
- Second and later interviews elicit a mix of new or previously reported codes
- Number of new codes eventually declines to zero
- Interview with 0 new codes may –and must- occur more than once.
- No Hypothesis test/no p-values. •Use all data.

For Today Only – I selected a Kaplan Meier Non-parametric Estimator

- Pragmatic choice
- The methodology can be generalized using other “survival type” distributions (exponential, etc.)
- Generalizing requires an interview/patient level dataset.

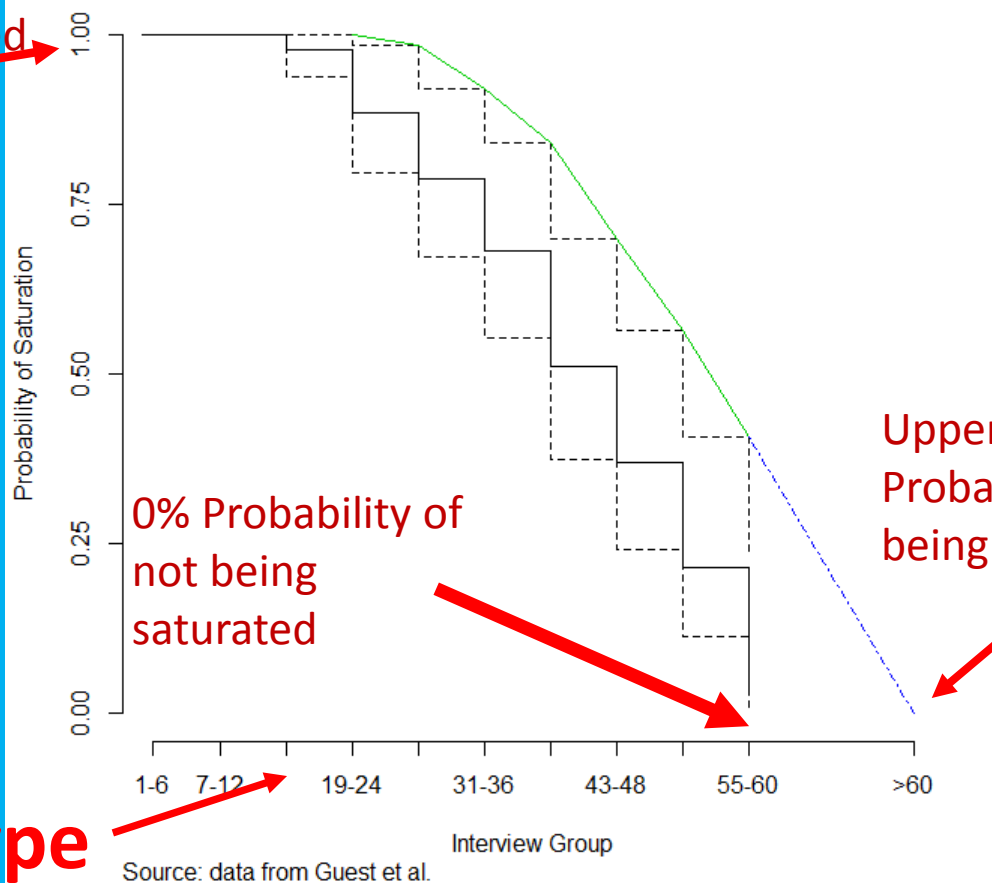
Preparing a dataset from Guest et al. to exhibit a model

- Impute one or more interviews of 6 with 0 new codes when total codes from block of 6 is <6 .
- Prefer to have interview level data

Note: data extracted from Guest Figure I

100% Probability of not being saturated

Kaplan Meier Non-Parametric Estimate of Interviews until Saturation with extrapolated upper 95% KM c.i.



0% Probability of not being saturated

Upper 95% C.I. Probability of not being saturation

13-18 Possible Type I error

Saturation Type I Error truth table

	Truth	
Interview Assessment	Not Saturation	Saturation
Saturation	Type I	
Not Saturation		Type II

Direct Sample Size Estimation

- Initial – classical single group “time to event” methods, median, mean, 50th-tile
- “interim” or “indirect” methods using– capture recapture
- Simulation based
- Research underway – Bayesian CRC methods suitable for “small samples”

Indirect sample size estimates using Capture Recapture a.k.a “Lincoln Peterson”

- Capture Recapture (CRC) wide and –diverse- use, for estimating wildlife populations (“# bears in Yellowstone park”), “software bugs”, “human rights violations”, aka.
- For PRO’ development used as interviews are conducted for an INDIRECT method of estimating total number of interviews (sample size).
- I am NOT proposing to use CRC for estimating total number of codes
- Several CRC estimators.

brief review: Capture recapture (CapRec)

- For indirect sample size estimation , NOT for definitive estimate of number of codes.
- Simplest case. Population of bears in Yellowstone
- Randomly sample bears. Tag(mark) bears, release into wild
- Second random sample, count number of marked vs. unmarked bears.
- Estimators use counts of total sample, marked and unmarked to estimate total number of bears with confidence interval
- For PRO's we are sampling 'codes'

Indirect sample size estimates using Capture Recapture (CRC)

- I was privately provided with an anonymized data set of interview information in a matrix format.
- Columns are anonymized “codes” , rows are “interviews in chronological order”,
- I used several CRC estimators, “Lincoln Peterson”, other extensions.
- I estimated total new codes at each interview
- All estimators and their upper confidence limit UNDERESTIMATE the total new codes.
- Underestimation is a –known- problem with CRC estimators

Example of Static PRO interview Matrix at the end of all interviews with 0/1 codes 1=code found 0=not found

Interview Number	Anonymized CODE /concept								
	A	B	C	D	E	F	...	ZZ	
1	1	1	1	1	1	1	1	1	
2	0	0	1	0	0	1	1	0	
3	1	1	1	1	0	0	0	0	
...									
k									

Bayesian Capture Recapture for “small samples” a.k.a “small number of codes”

- The full Capture Recapture matrix evolves by interview

Formal Sample Size Algorithms to appear in publication

- I. Initial Estimate of N of interviews, “ballpark” of number of codes
- II. Revised estimate of number of codes as often as 2nd and later interviews using “capture recapture”
- III. Estimate of probability of saturation
- IV. If needed Extrapolate probability of saturation to estimate total number of interviews with upper confidence interval
- V. Use classical sample size for ‘time to’
- VI. Indirect estimate of sample size - Estimate number of “codes” at each interview > 2 , by capture recapture

Conclusions

- Define Type I error for saturation
- Initial Estimate of number of interviews (sample size)
- Estimate “saturation probability” (0.0%) and C.I.
- indirect assessment of sample size by estimating number of codes (capture recapture)
- Dataset structure required for analysis (add to STDMM / ADaM?)
- Formal Methods intended to be acceptable in a regulatory setting

Questions

References

- Brédart, Anne, et al. "Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience." *Health and quality of life outcomes* 12.1 (2014): 15.
- Francis, Jill J., et al. "What is an adequate sample size? Operationalising data saturation for theory-based interview studies." *Psychology and Health* 25.10 (2010): 1229-1245.
- Fusch, Patricia I., and Lawrence R. Ness. "Are we there yet? Data saturation in qualitative research." *The Qualitative Report* 20.9 (2015): 1408.
- Mason, Mark. "Sample size and saturation in PhD studies using qualitative interviews." *Forum qualitative Sozialforschung/Forum: qualitative social research*. Vol. 11. No. 3. 2010.
- Patrick, Donald L., et al. "Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument." *Value in Health* 14.8 (2011): 967-977.
- Patrick, Donald L., et al. "Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding." *Value in Health* 14.8 (2011): 978-988.
- Wasserstein, Lazar "The ASA's statement on p-values: context, process, and purpose," *The American Statistician*, 2016

Thank you

Chris Barker, Ph.D.
www.barkerstats.com