Current Topics in Human–Computer Interaction

Quantitative Analysis (Part 2) • Qualitative Analysis

Prof. Dr. Jan Borchers
Media Computing Group
RWTH Aachen University

Summer Semester ’26

https://hci.rwth-aachen.de/cthci

RWTH AACHEN UNIVERSITY

CHAPTER 9

Quantitative Analyses (Part 2)

Recap: Significance Testing

1. Formulate hypothesis
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. Recruit subjects
4. Run experiment (to collect data which you will analyze)
5. **Perform statistical analysis on the collected data to accept or reject hypothesis**

- Significance testing:
  - Assume H_0 to be true (i.e., no difference at the population level)
  - p-value: “The chances of obtaining the experimental data we’ve collected assuming the null hypothesis is true”
- *p*-value gives us confidence in accepting or rejecting the null-hypothesis (i.e., no difference between distributions)

In-class Exercise: p-Value

- Hypothesis (H_1): There is a significant difference in TV viewing duration between school students and college students.
- Method: Gathered survey data from 50 school students and 50 college students
- Result: On average, school students watch 3.5 hours per day, and college students watch 3.0 hours per day, ***p* = 0.03.**

⇒ Which of the following statements are correct?

A. There is a 3% probability that school students watch more TV than college students.

B. There is a 3% probability that school students watch a different amount of TV than college students.

C. Assuming that school students watch a different amount of TV than college students, there is a 3% probability that this result occurs.

D. Assuming that school students and college students watch the same amount of TV, there is a 3% probability that this result occurs.

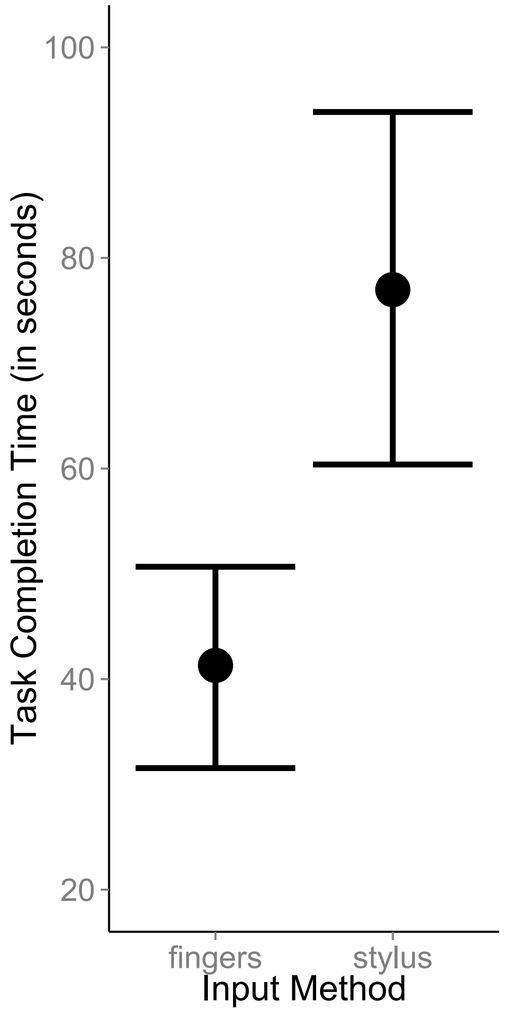
Selecting the Correct Statistical Test

How many levels does the variable have?

	2		> 2	
Between-Groups	Unpaired t-Test	Wilcoxon Rank Sum Test	One-Way ANOVA	Kruskal-Wallis Test
Within-Groups	Paired t-Test	Wilcoxon Signed-Rank Test	One-Way Repeated-Measures ANOVA	Friedman Analysis
	Yes	No	Yes	No

Is the variable of the **interval** type *and* **normally distributed** *and* has **equal variances***?*

Significance Testing and p-Values In Our Paper



"The input method (*Finger*, *Stylus*) had a significant effect on the task completion time, $t(20) = 4.03$, $p < .001$.

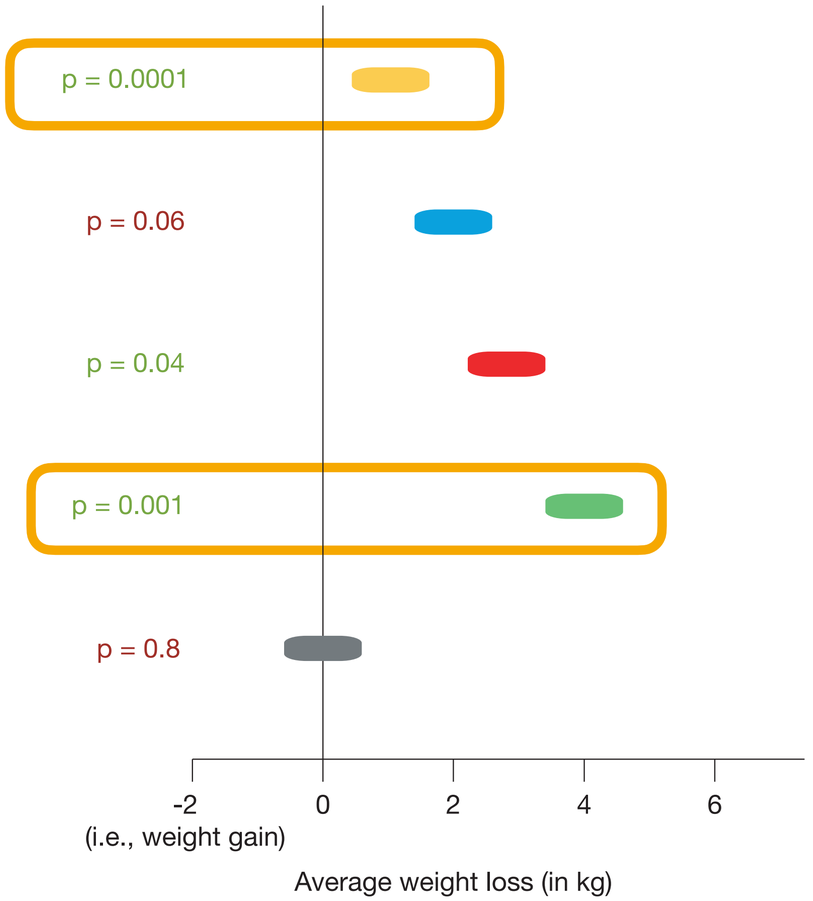
Finger ($M = 42.03$ s; 95% CI [31.78, 52.22]) was faster than *Stylus* ($M = 76.21$ s; 95% CI [59.40, 93.02]).

The difference between the means was 34.18 s."

QUANTITATIVE ANALYSIS

Statistically Significant = Practically Significant?

Scenario: Weight Loss via Pills



Takeaway: p-values are not sufficient for interpretation!

[McCloskey and Ziliak, The Unreasonable Ineffectiveness of Fisherian “Tests” in Biology, and Especially in Medicine, Biological Theory ’09]

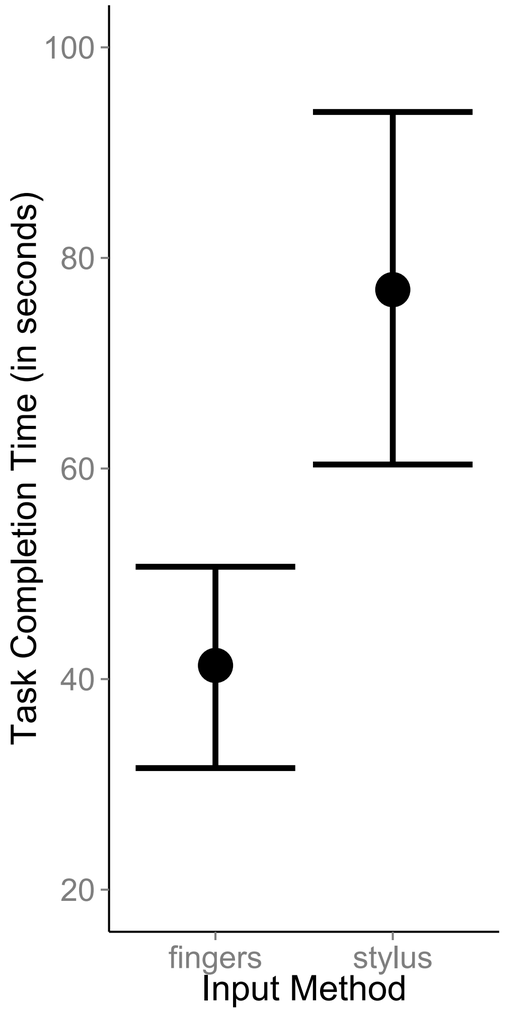
Effect Size

- The *p*-value tells us the chances of the sample distributions coming from the same population distribution
  - But: Statistically significant ($p < 0.05$) != practically significant
- For practical significance, we use **effect sizes**
- Effect size: Measure of how big the differences between distributions are

How to Report Effect Size

- Absolute difference between two means
  - E.g., “*Stylus* is 40 s slower than *Finger*”
  - In original unit, intuitive
  - But has measurement unit (s, points,…) so requires domain knowledge
  - Often used in HCI
- Relative differences (percentages or ratios) of two means
  - E.g., “*Stylus* is twice as slow as *Finger*”
  - Emphasizes relative magnitude of the effect

Effect Sizes In Our Research Paper



"The input method (*Finger*, *Stylus*) had a significant effect on the task completion time, t(20) = 4.03, p < .001.

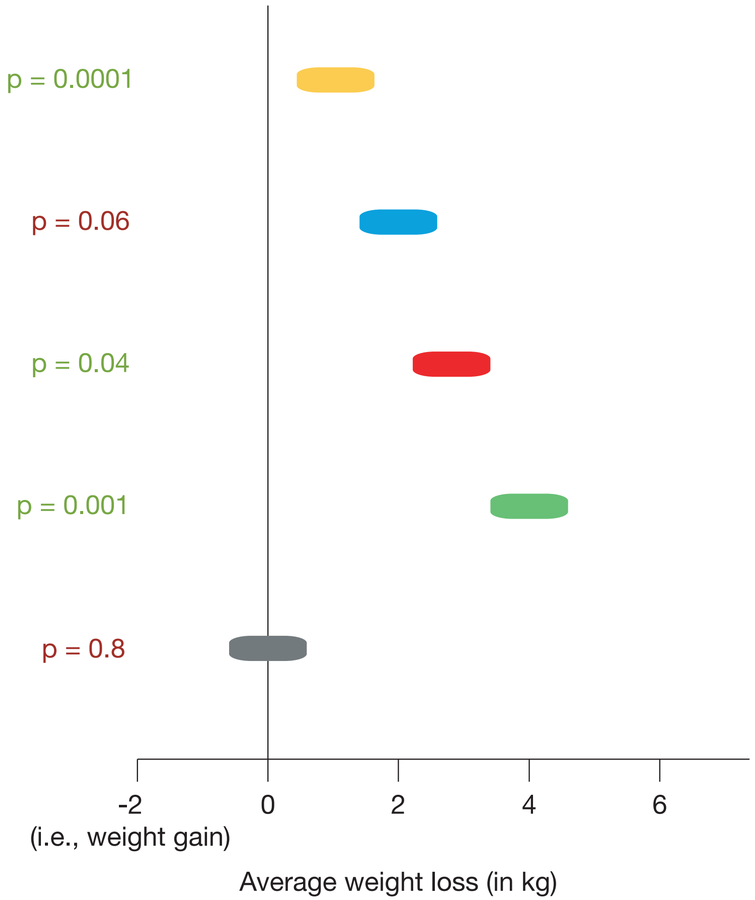
Finger (*M* = 42.03 s; 95% CI [31.78, 52.22]) was faster than *Stylus* (*M* = 76.21 s; 95% CI [59.40, 93.02]).

The difference between the means was 34.18 s."

QUANTITATIVE ANALYSIS

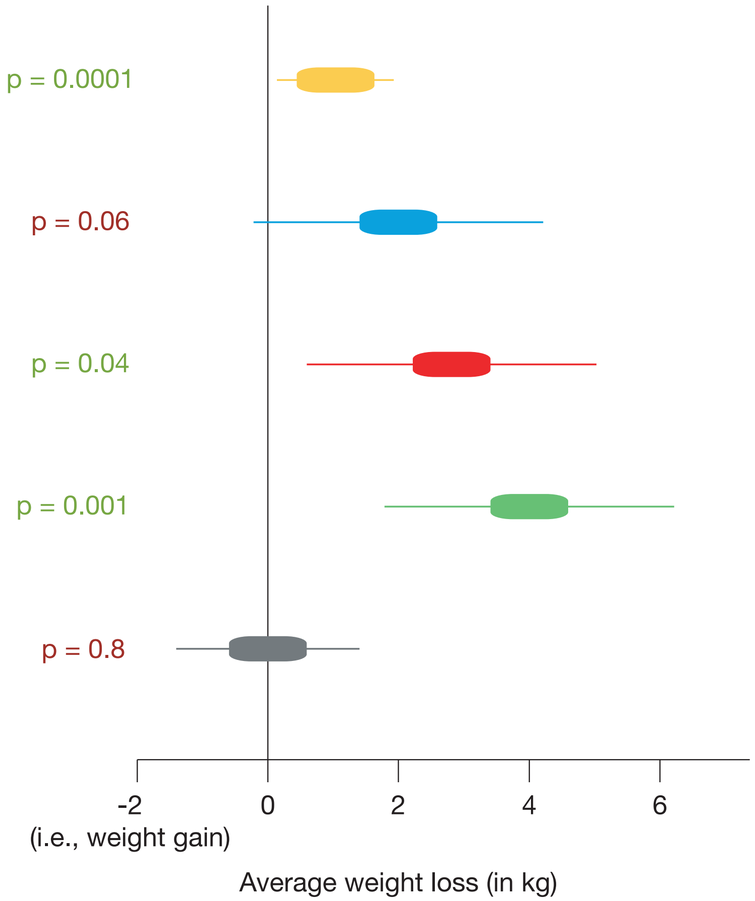
Confidence Intervals

Scenario: Weight Loss via Pills



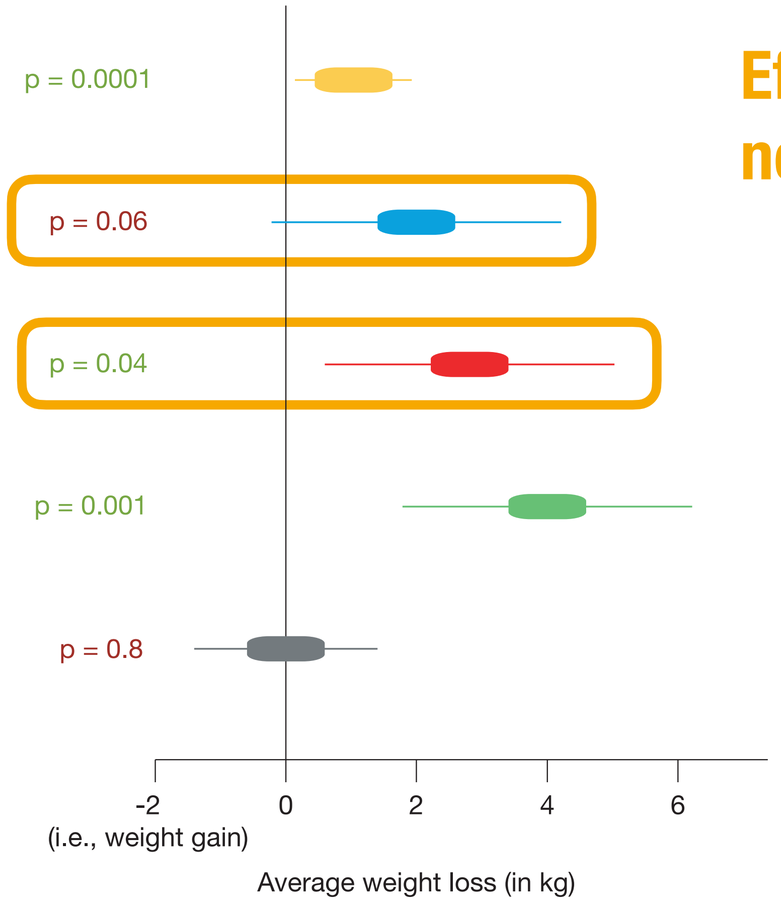
[McCloskey and Ziliak, The Unreasonable Ineffectiveness of Fisherian “Tests” in Biology, and Especially in Medicine, Biological Theory ’09]

Scenario: Weight Loss via Pills



[McCloskey and Ziliak, The Unreasonable Ineffectiveness of Fisherian “Tests” in Biology, and Especially in Medicine, Biological Theory ’09]

Scenario: Weight Loss via Pills



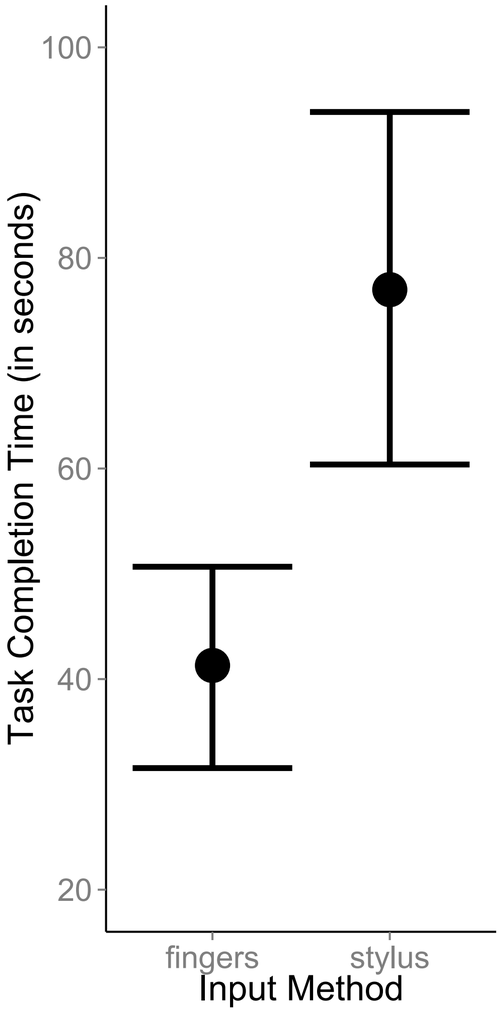
Effect sizes by themselves are not adequate for interpretation!

[McCloskey and Ziliak, The Unreasonable Ineffectiveness of Fisherian “Tests” in Biology, and Especially in Medicine, Biological Theory ’09]

95% Confidence Interval

- An interval estimate (i.e., a range) of the population mean
- In an infinite number of experiments, 95% of the time, the 95% CIs will contain the population mean
- 95% is a convention, might vary across domains
  (e.g., medicine, psychology have different conventions)

CIs In Our Research Paper



"The input method (*Finger*, *Stylus*) had a significant effect on the task completion time, t(20) = 4.03, p < .001.

Finger (*M* = 42.03 s; 95% CI [31.78, 52.22]) was faster than *Stylus* (*M* = 76.21 s; 95% CI [59.40, 93.02]).

The difference between the means was 34.18 s."

Guidelines for Reporting

- Report effect sizes and 95% confidence intervals in addition to descriptive statistics (mean, sd) and results of the significance test

- **Be transparent:** Include the experimental data, analysis script, and your rationale for analysis

Useful Resources

- Koji Yatani, Statistical Methods for HCI Research, 2022
- Jacob O. Wobbrock, Practical Statistics for HCI, 2011
- Geoff Cumming, The New Statistics: Why and How. *Psychological Science,* vol. 25, issue 1, pp. 7–29, 2014.
- John Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan.* 2nd edition, Academic Press, 2014

Summary

- We need statistical analysis to establish causal relationship between our dependent and independent variables
- Significance testing can be used to reject the null hypothesis
- Effect size quantifies the effect of the independent variable on the dependent variable
- 95% confidence intervals help deal with uncertainty in data

CHAPTER 10

Qualitative Analysis

When to Use Qualitative Analysis?

- To answer “how” and “why” questions, e.g.,
  - How do people use ticket vending machines?
  - Why do elderly people have difficulty using some touchscreens?
- Can be useful to identify problems, a useful first step before building artifacts
- Can also be used to evaluate an artifact
  - Often helps understand the reasons behind quantitative results
  - But: You cannot make generalizable claims about any hypothesis; use quantitative analysis for this

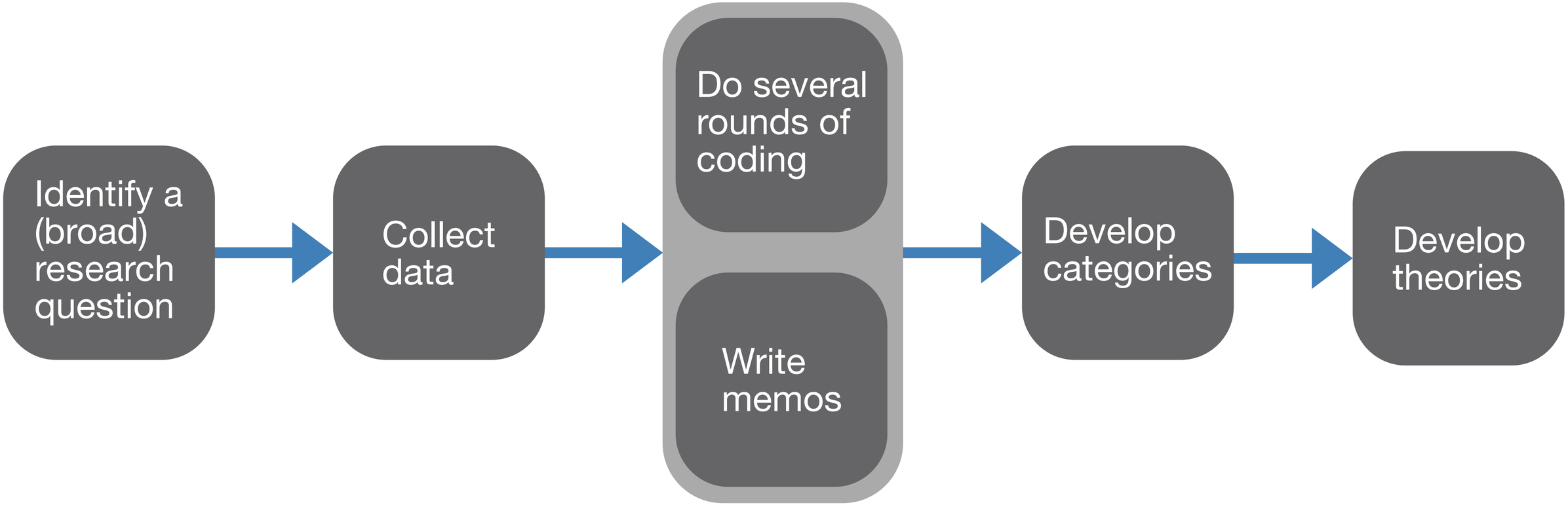
Data Collection Techniques

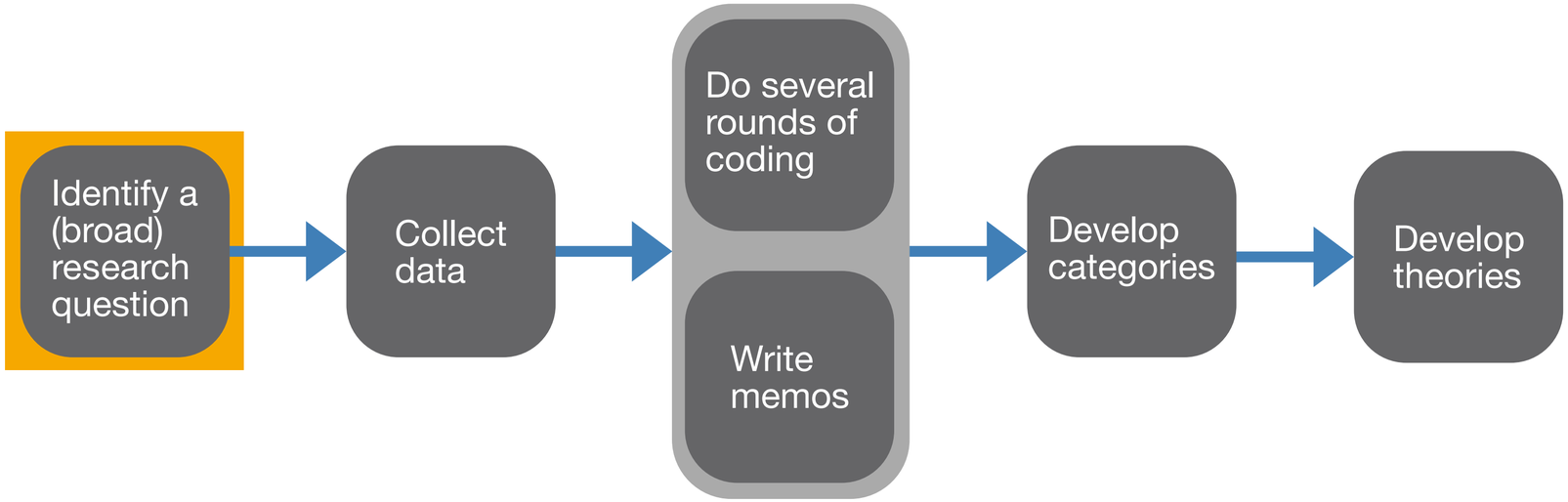
- Interviews
- Diary studies
- Observation
- Focus groups
- …

Evaluation Techniques

- Grounded-theory
- Affinity diagramming/contextual design

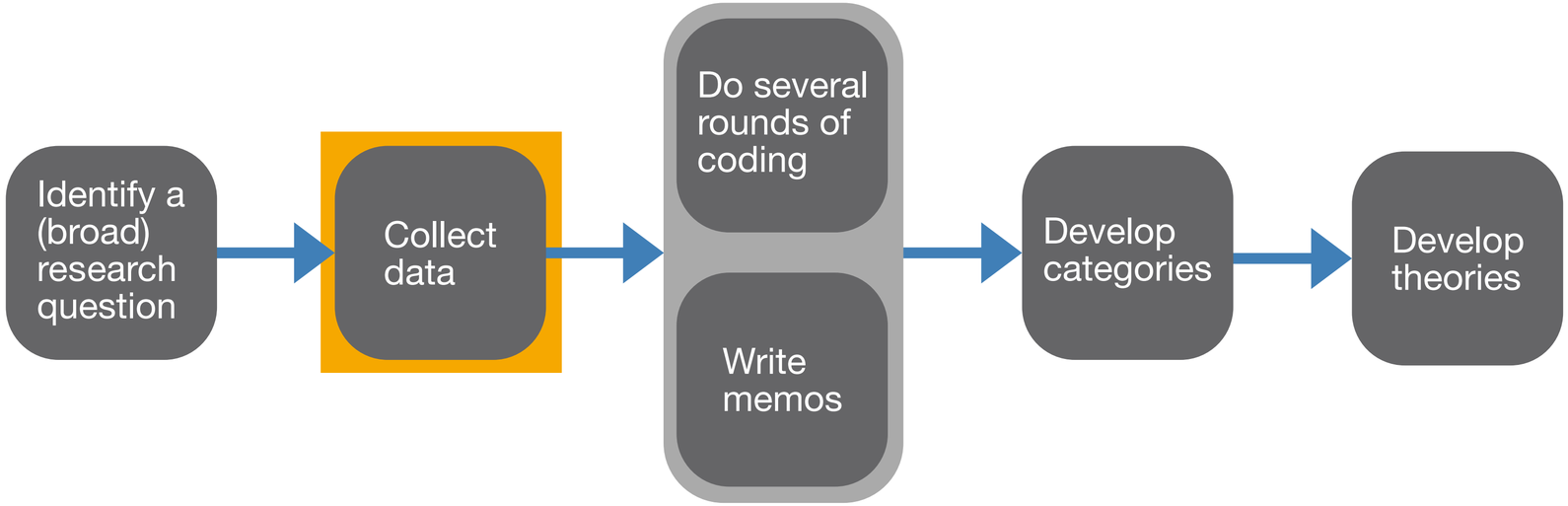
Overview of Grounded-Theory Method



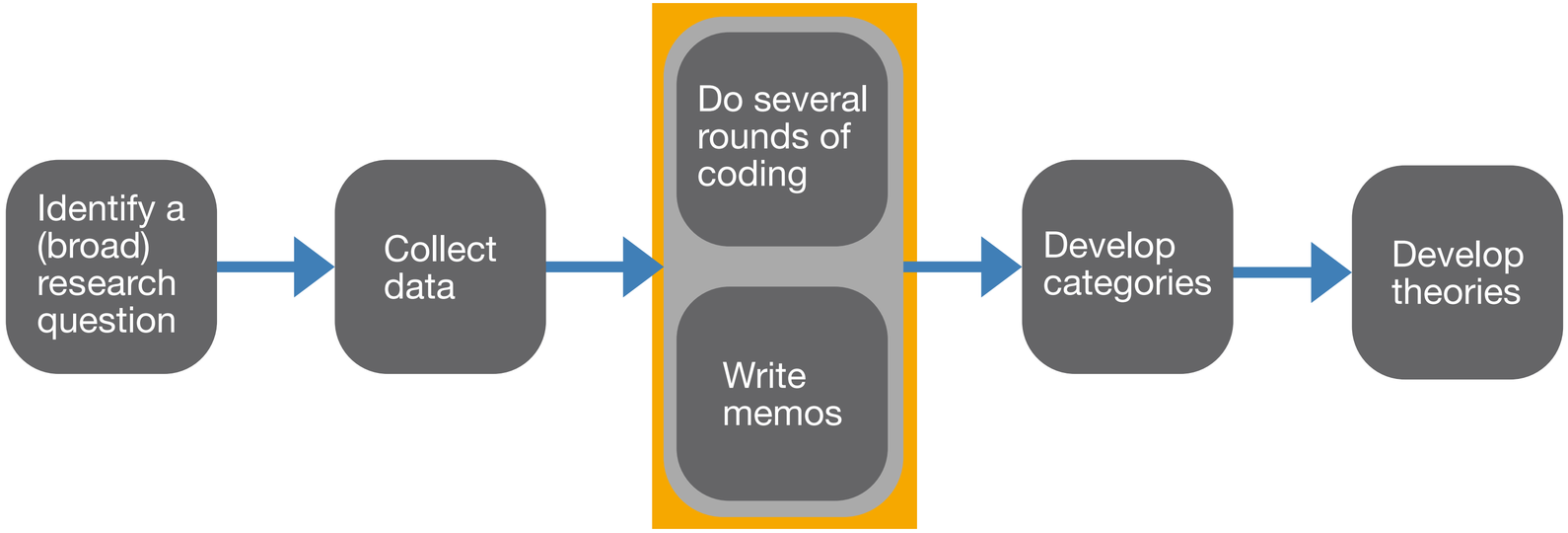


- “How do foreigners use the Deutsche Bahn ticket vending machine?”

needpix.com



- Observation and follow-up interviews
- Questions can be based on what you observed
  - E.g., “I noticed that you took some time to figure out the payment method—could you talk about what happened?”



Passenger:ExpressTrain

UseCase:OnlyCash

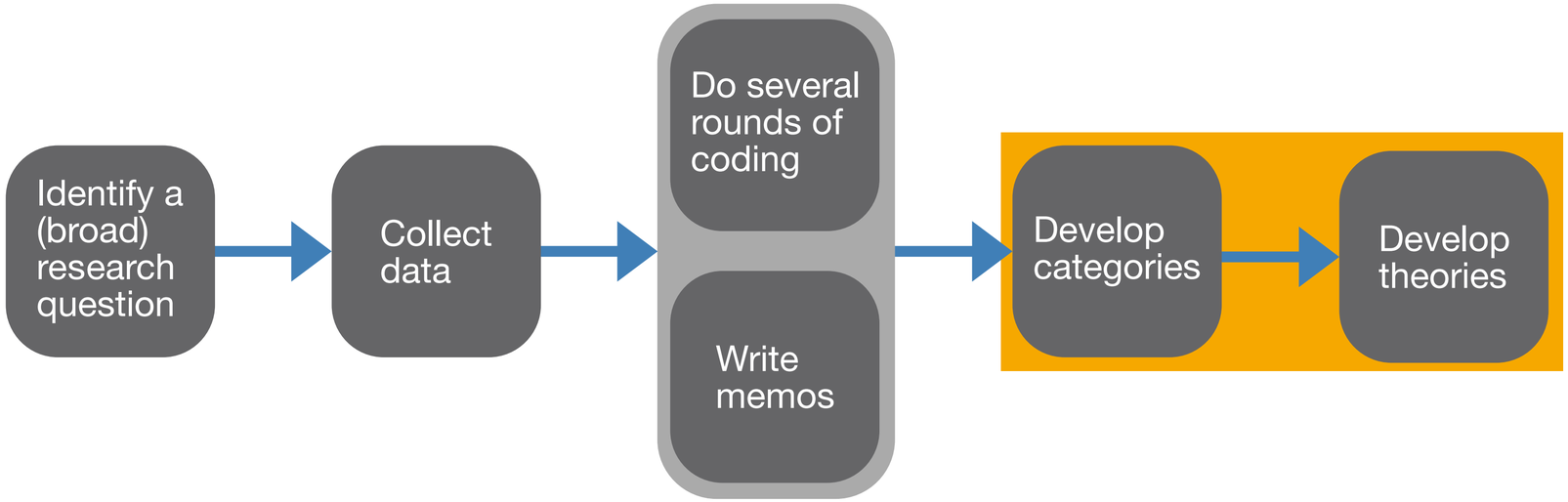
"I had to take an ICE train this morning. I had only 5 minutes left to buy a ticket because I only had cash to pay for the ticket. There was a bit of a queue so I had to wait for a couple of minutes before I could buy the ticket. The process was OK overall, but I had difficulty selecting the destination—I first selected Bochum instead of Bonn."

Problem:Waiting

Problem:CitySelection

Example Memo

Many participants have problems selecting the destination. Is this because of the interface? Maybe I can check if the participants who are not so rushed also make this mistake. Also, does this occur for specific cities? These are things I want to investigate.



- Combine codes to develop categories
- Use categories and memos to develop theories
  - E.g., “Foreigners face problems even before using the interface, which impacts their experience more than usability issues”

I. Problems

Cat. 1: Problems before purchase

I.I. Queues before purchase

I.II. Finding the least used machines

I.III. Determining which type of ticket to buy

Cat. 2: Interface problems

I.IV. Incorrect selection of cities

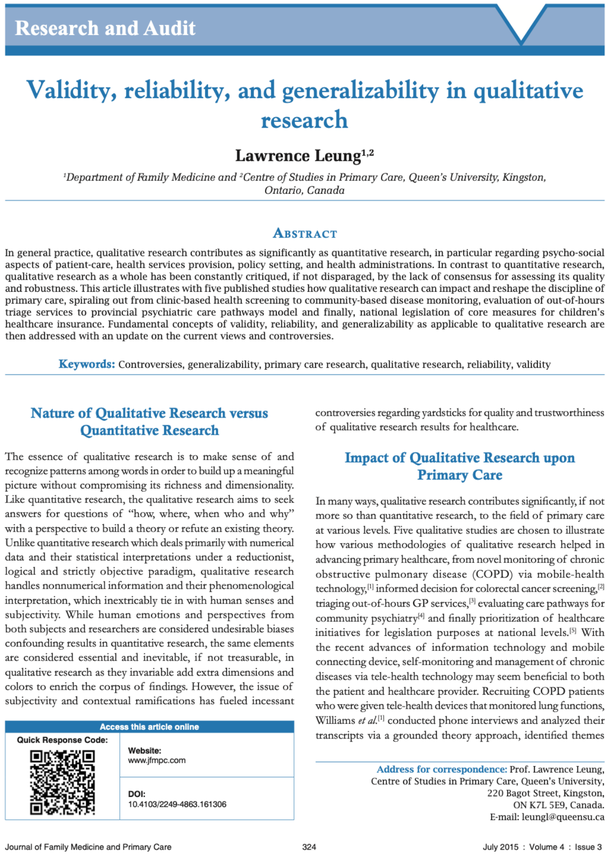
I.V. Unable to pay with the correct payment method

I.VI. Changing language

Validity, Reliability, and Transparency

- To ensure **validity**
  - Perform **triangulation,** i.e., use more than one source of data to confirm a finding
  - Use a **representative sample,** e.g., make sure you have people of different countries, ethnicities, gender, age group, and experience with technology
- To ensure **reliability,** compute **inter-coder agreement** after coming up with a coding scheme
- Guidelines for **transparent** qual. research: [Aguinis et al., Transparency and Replicability in Qualitative Research: The Case of Interviews With Elite Informants, Strategic Management Journal ’19]

Further Reading



Research and Audit

Validity, reliability, and generalizability in qualitative research

Lawrence Leung[1,2]

[1]*Department of Family Medicine and* [2]*Centre of Studies in Primary Care, Queen's University, Kingston, Ontario, Canada*

ABSTRACT

In general practice, qualitative research contributes as significantly as quantitative research, in particular regarding psycho-social aspects of patient-care, health services provision, policy setting, and health administrations. In contrast to quantitative research, qualitative research as a whole has been constantly critiqued, if not disparaged, by the lack of consensus for assessing its quality and robustness. This article illustrates with five published studies how qualitative research can impact and reshape the discipline of primary care, spiraling out from clinic-based health screening to community-based disease monitoring, evaluation of out-of-hours triage services to provincial psychiatric care pathways model and finally, national legislation of core measures for children's healthcare insurance. Fundamental concepts of validity, reliability, and generalizability as applicable to qualitative research are then addressed with an update on the current views and controversies.

Keywords: Controversies, generalizability, primary care research, qualitative research, reliability, validity

Nature of Qualitative Research versus Quantitative Research

The essence of qualitative research is to make sense of and recognize patterns among words in order to build up a meaningful picture without compromising its richness and dimensionality. Like quantitative research, the qualitative research aims to seek answers for questions of "how, where, when who and why" with a perspective to build a theory or refute an existing theory. Unlike quantitative research which deals primarily with numerical data and their statistical interpretations under a reductionist, logical and strictly objective paradigm, qualitative research handles nonnumerical information and their phenomenological interpretation, which inextricably tie in with human senses and subjectivity. While human emotions and perspectives from both subjects and researchers are considered undesirable biases confounding results in quantitative research, the same elements are considered essential and inevitable, if not treasurable, in qualitative research as they invariable add extra dimensions and colors to enrich the corpus of findings. However, the issue of subjectivity and contextual ramifications has fueled incessant controversies regarding yardsticks for quality and trustworthiness of qualitative research results for healthcare.

Impact of Qualitative Research upon Primary Care

In many ways, qualitative research contributes significantly, if not more so than quantitative research, to the field of primary care at various levels. Five qualitative studies are chosen to illustrate how various methodologies of qualitative research helped in advancing primary healthcare, from novel monitoring of chronic obstructive pulmonary disease (COPD) via mobile-health technology,[1] informed decision for colorectal cancer screening,[2] triaging out-of-hours GP services,[3] evaluating care pathways for community psychiatry[4] and finally prioritization of healthcare initiatives for legislation purposes at national levels.[5] With the recent advances of information technology and mobile connecting device, self-monitoring and management of chronic diseases via tele-health technology may seem beneficial to both the patient and healthcare provider. Recruiting COPD patients who were given tele-health devices that monitored lung functions, Williams *et al.*[1] conducted phone interviews and analyzed their transcripts via a grounded theory approach, identified themes

Access this article online

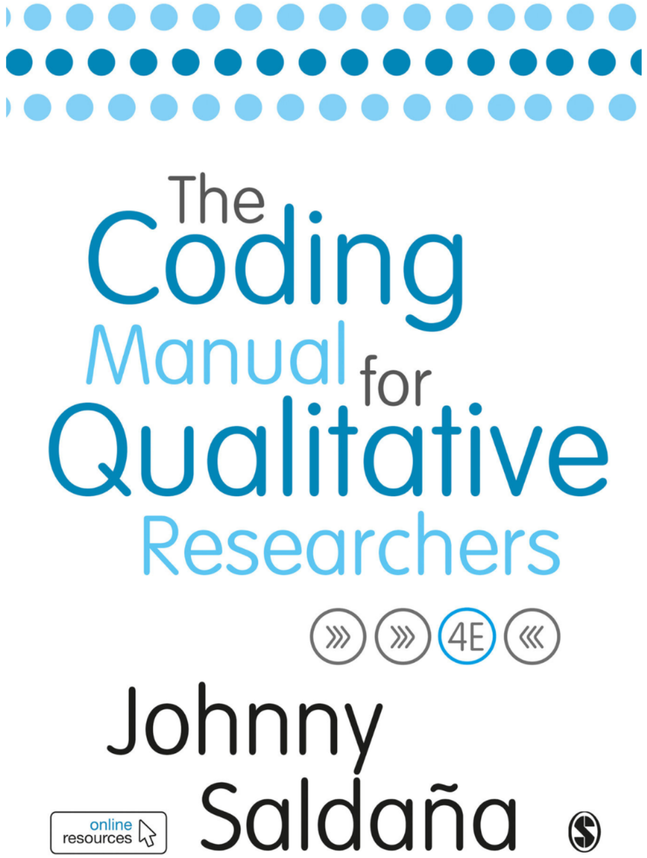
Quick Response Code:

Website: www.jfmpc.com

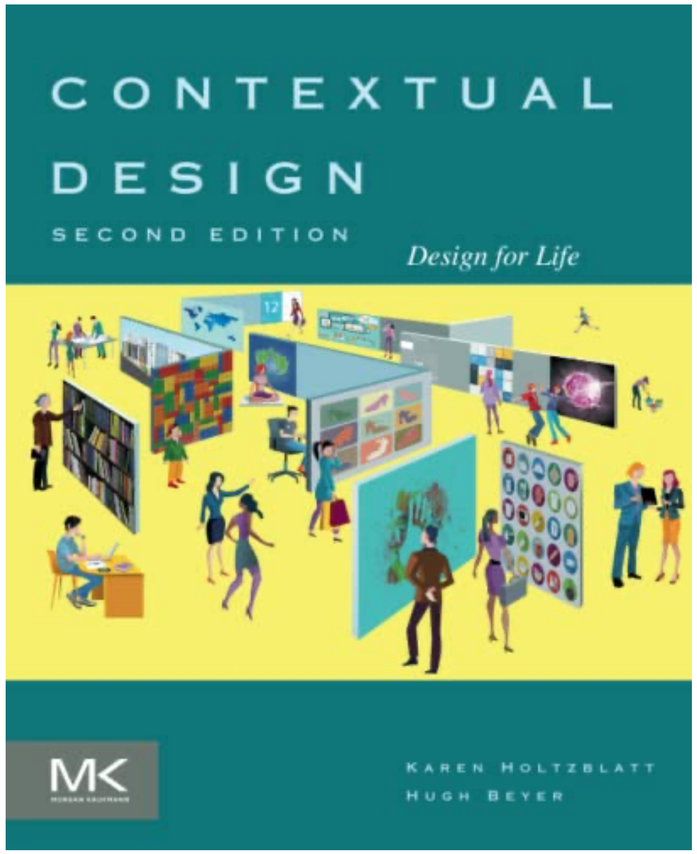
DOI: 10.4103/2249-4863.161306

Address for correspondence: Prof. Lawrence Leung, Centre of Studies in Primary Care, Queen's University, 220 Bagot Street, Kingston, ON K7L 5E9, Canada. E-mail: leungl@queensu.ca

Lawrence, Validity, Reliability, and Generalizability in Qualitative research, Journal of Family Medicine and Primary Care ’15



For Grounded-theory analysis: Johny Saldaña, The Coding Manual for Qualitative Researchers, 2021



For affinity diagramming and contextual design: Beyer and Holtzblatt, Contextual Design: Design for Life, 2016

RWTH AACHEN UNIVERSITY

Summary

- Use qualitative analysis to answer “how” and “why” questions
- Often a useful first step to better understand problems or to identify problems
- Can also be a useful tool for evaluating an artifact
- Requires qualitative coding, inductive reasoning, and rigorous comparisons of data