FINDING ALL POSSIBLE WAYS HOW THINGS CAN GO WRONG.

Lessons learned from usability testing

Martin Schmettow

UNIVERSITY OF TWENTE.

#1 Anything that can go wrong will go wrong

#2 Knowing all the possible ways that things can go wrong is very useful for developing safe and resilient systems.

The purpose of usability testing is to find all possible ways how things can go wrong.

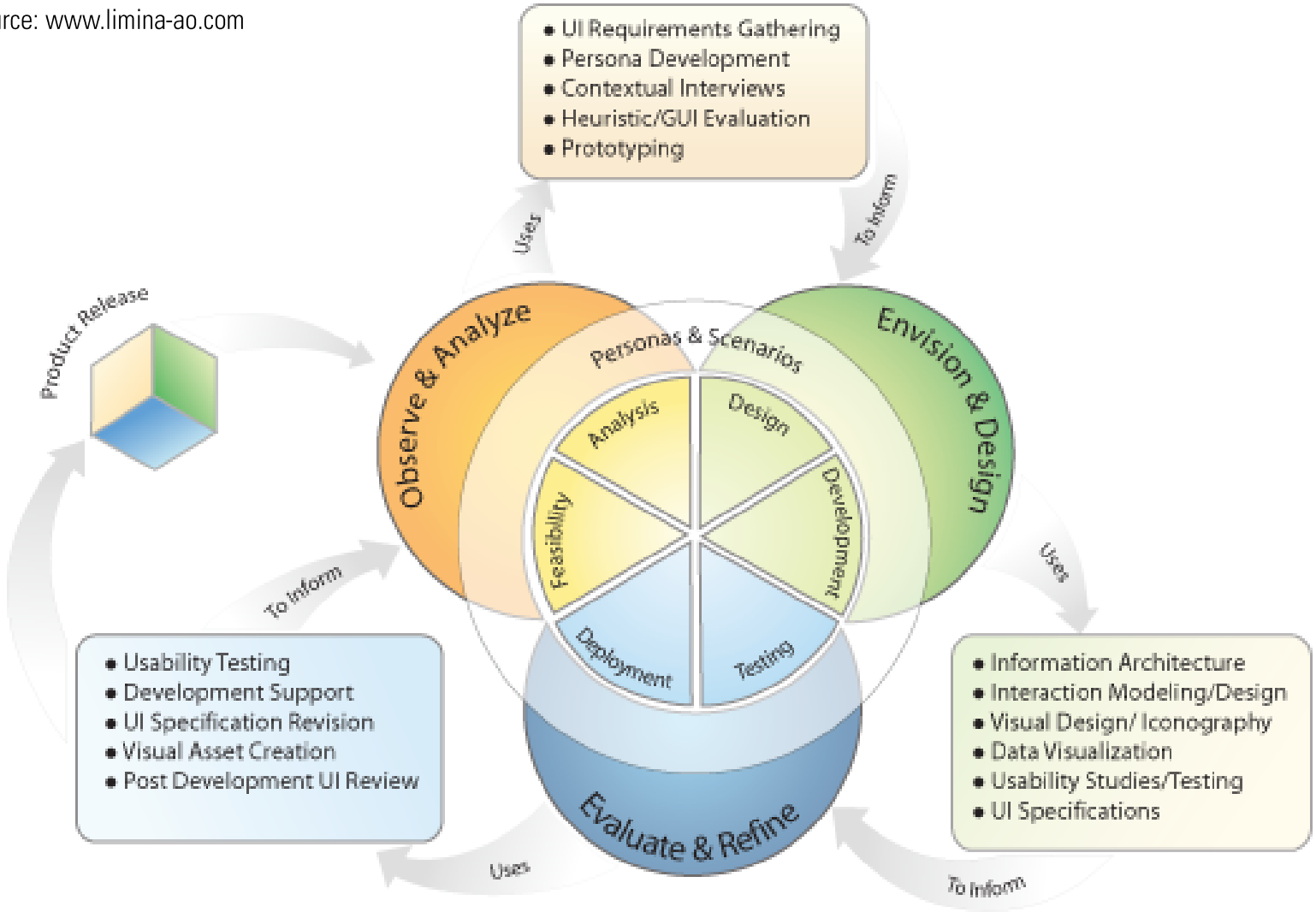
High level usability criteria

- Effectiveness
  - **accuracy** and **completeness** with which users achieve specified goals
- Efficiency
  - **Effort** of achieving results of certain accuracy and completeness
- Satisfaction
  - **freedom from discomfort**, and **positive attitudes** towards the user of the product

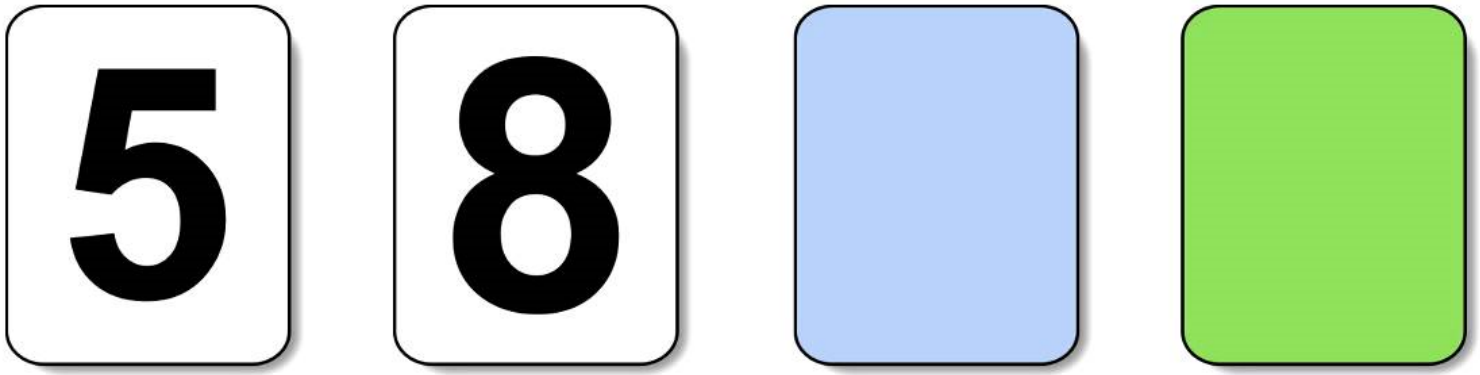
Three Principles of User-Centered Design

1) Iterative Development
   (a) Usability requirements are a **moving target**
   (b) **Iterate** between design and evaluation of design
2) Participation
   (a) Know you **users**, know there **tasks**
   (b) **Involve** users in design early
3) Evidence
   (a) **Measure** performance of interaction
   (b) **Evaluate** design via direct behavioral observation

Source: www.limina-ao.com



IF A CARD SHOWS AN EVEN NUMBER ON ONE FACE, THEN ITS OPPOSITE FACE IS BLUE.



Which card would you turn to test the rule?

#3 All humans suffer from the confirmatory bias. For finding problems one must take a strictly pessimistic stance.

Usability Evaluation

Performance

- Empirical
  - Effectiveness
  - Efficiency
  - Satisfaction
- Analytic
  - Models

Qualitative assessment

- Empirical
  - User Testing
- Analytic
  - Expert Evaluation
  - Models, Tools

Usability Testing

- Real tasks
- Representative users
- Behavior observation
- Think-aloud interview

The purpose of usability testing is to find all possible ways how things can go wrong.

How man users to test?

The “Five Users” Problem

The „five users“ debate (abridged)

Testing  users is enough -- Nielsen

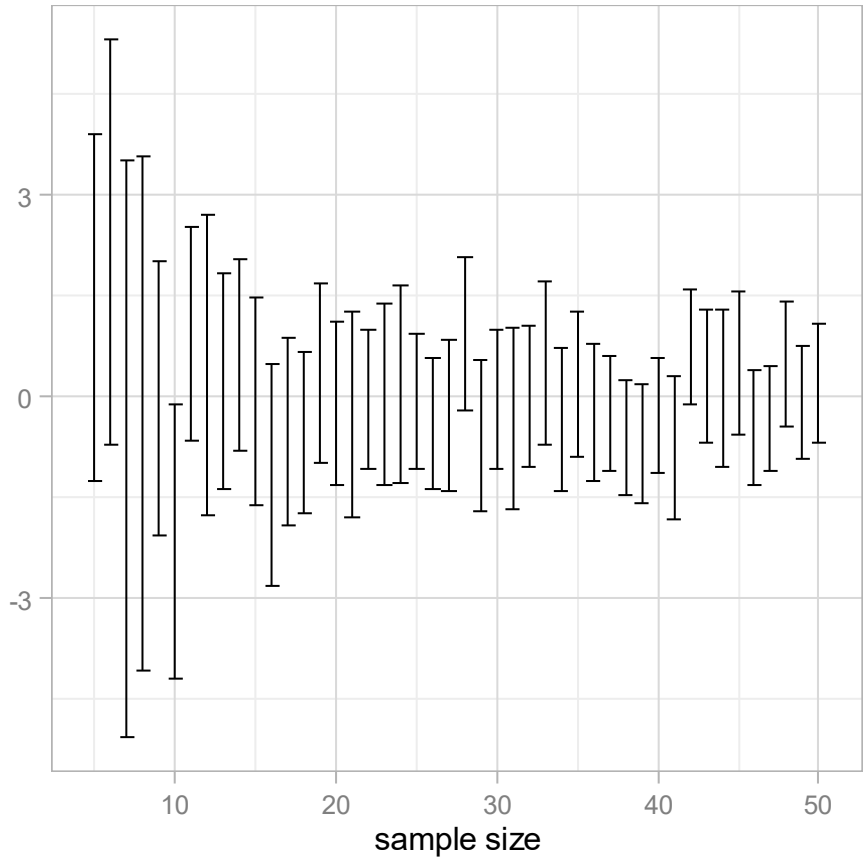
 users is nowhere near enough -- Spool

The magic number really is  -- Salvendy

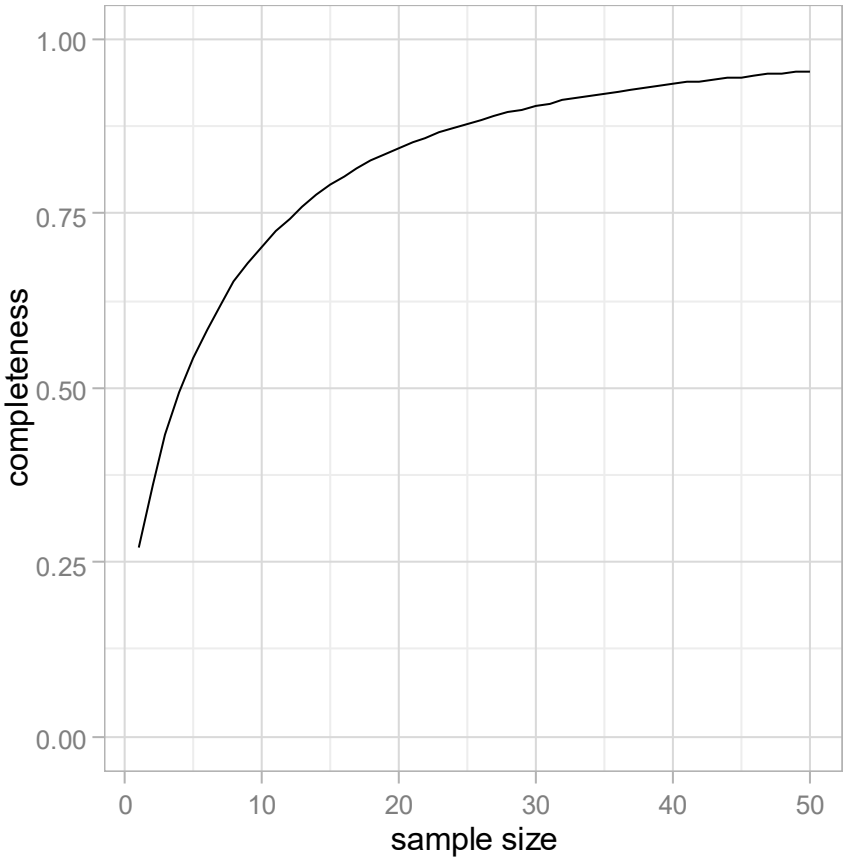
Magic numbers are strictly hocus-pocus -- Me

Stay with the tried-and-true,  users -- Nielsen

#4 Magic numbers for sample size are strictly hocus-pocus.

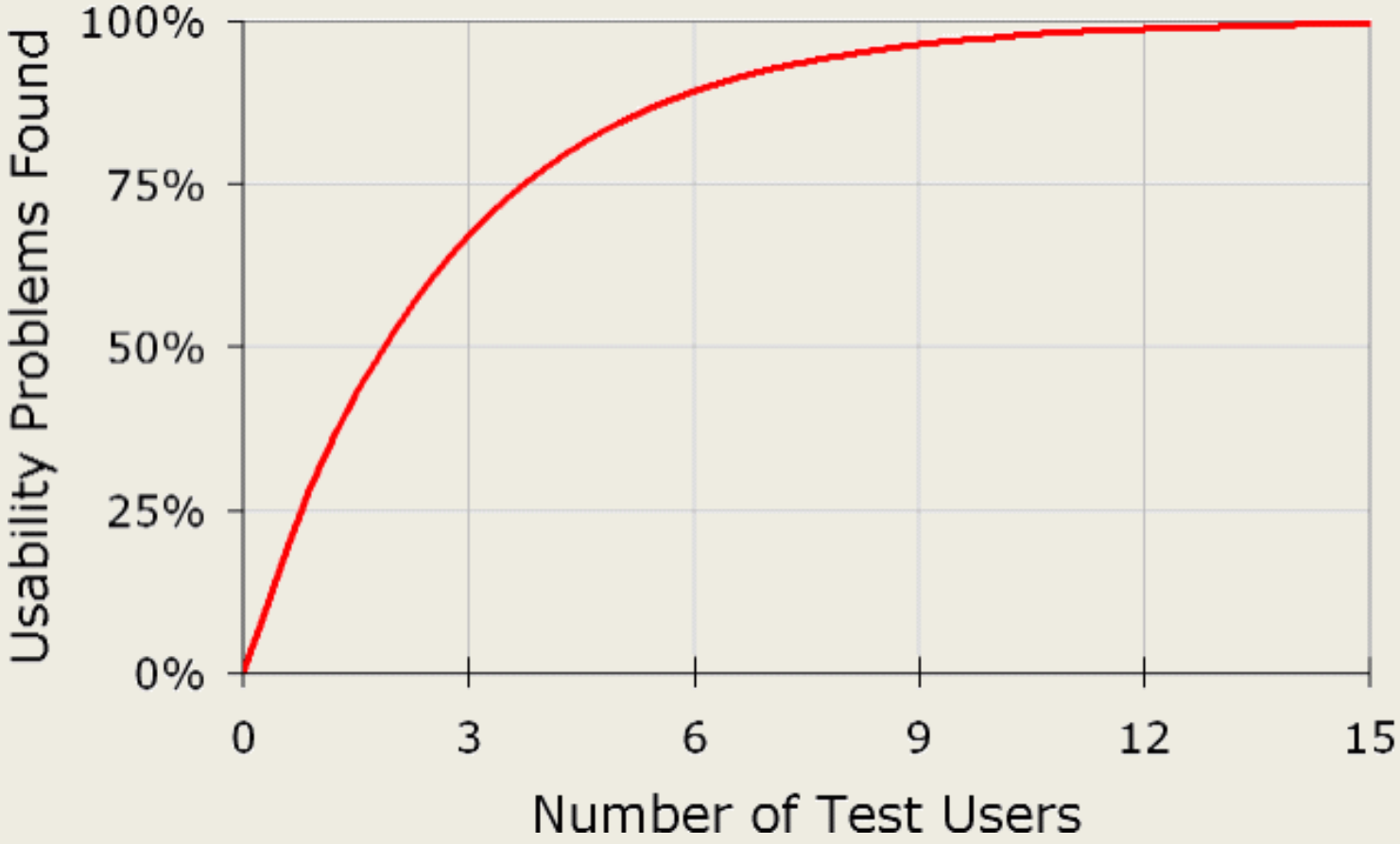


experimental research:
precision of estimates



qualitative research:
completeness of discoveries

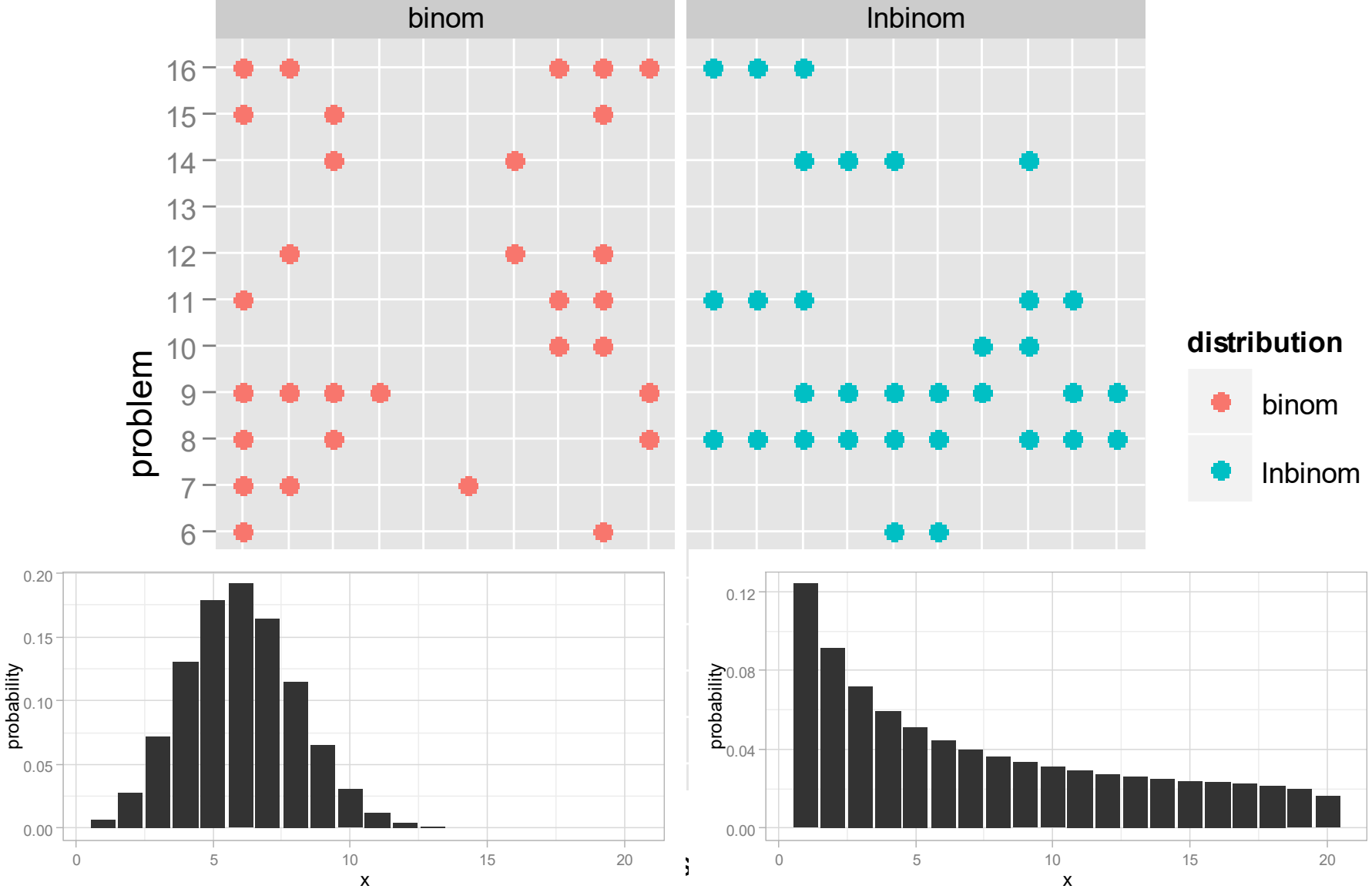
impact of sample size



How many users to test? $1-(1-p)^n$

Nielsen, J. Why you only need to test with 5 users. 2000.
http://www.useit.com/alertbox/20000319.html.

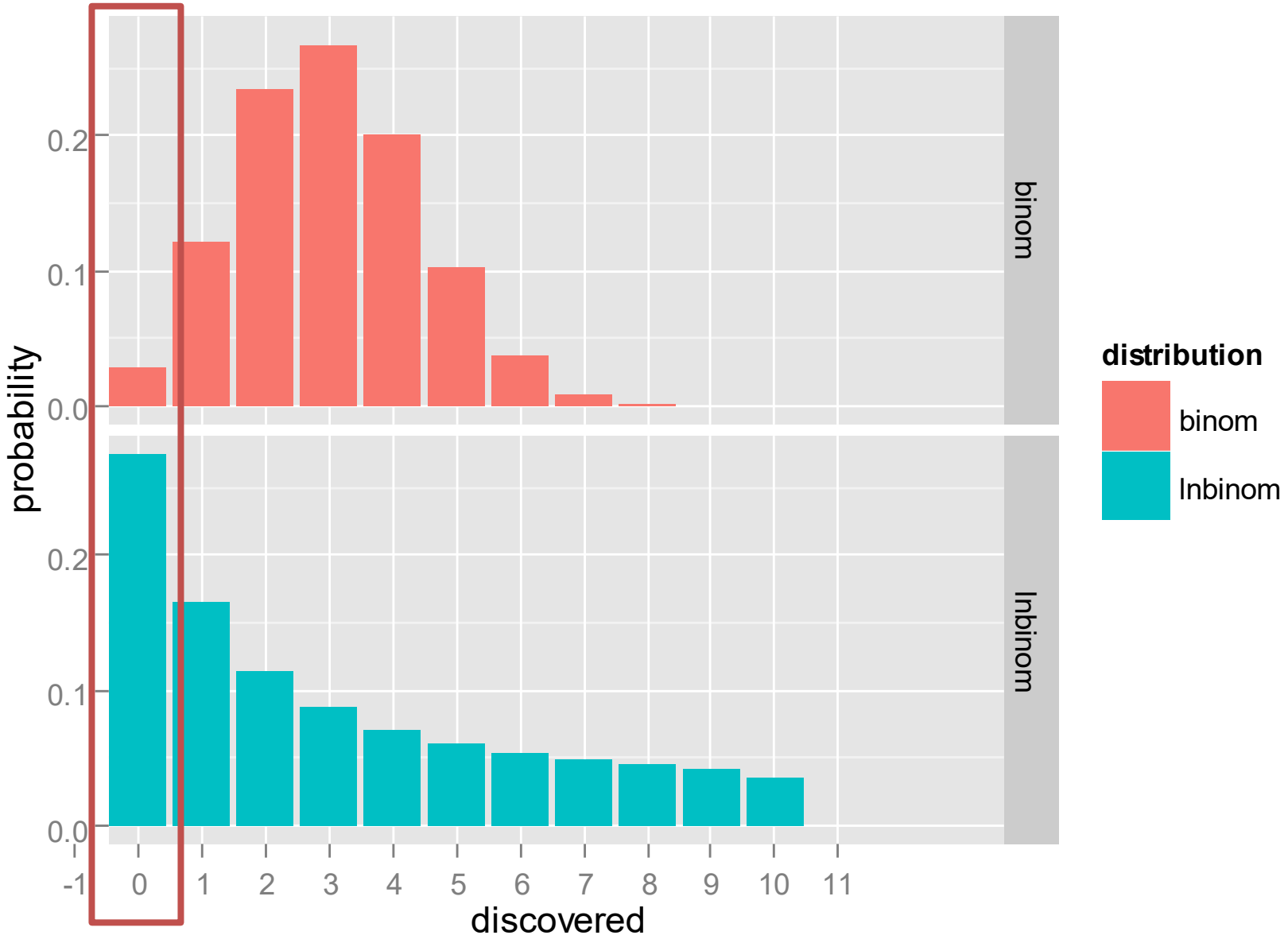
#5 Proper mathematical models for problem discovery must regard visibility variance and incompleteness.



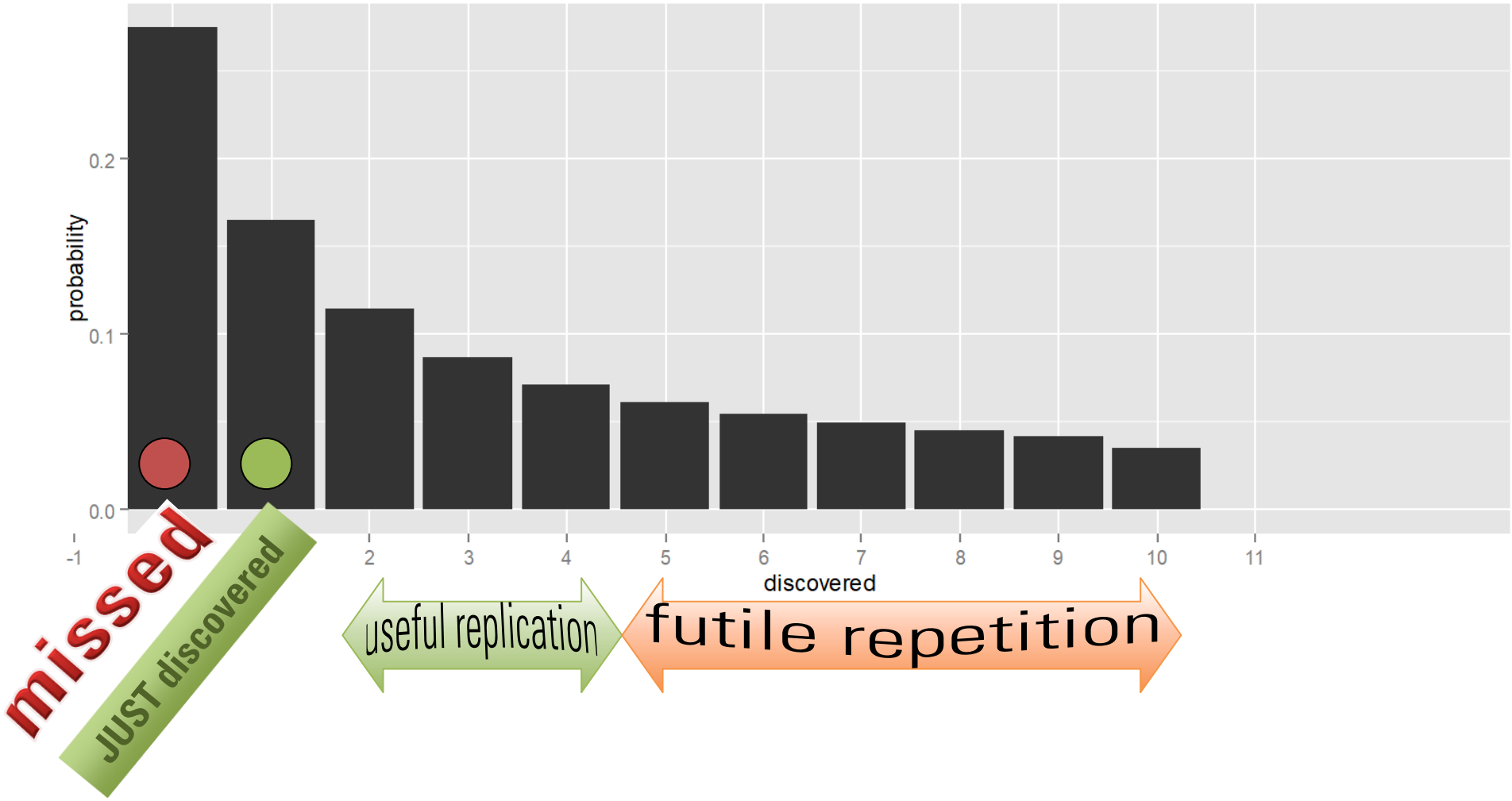
Binomial

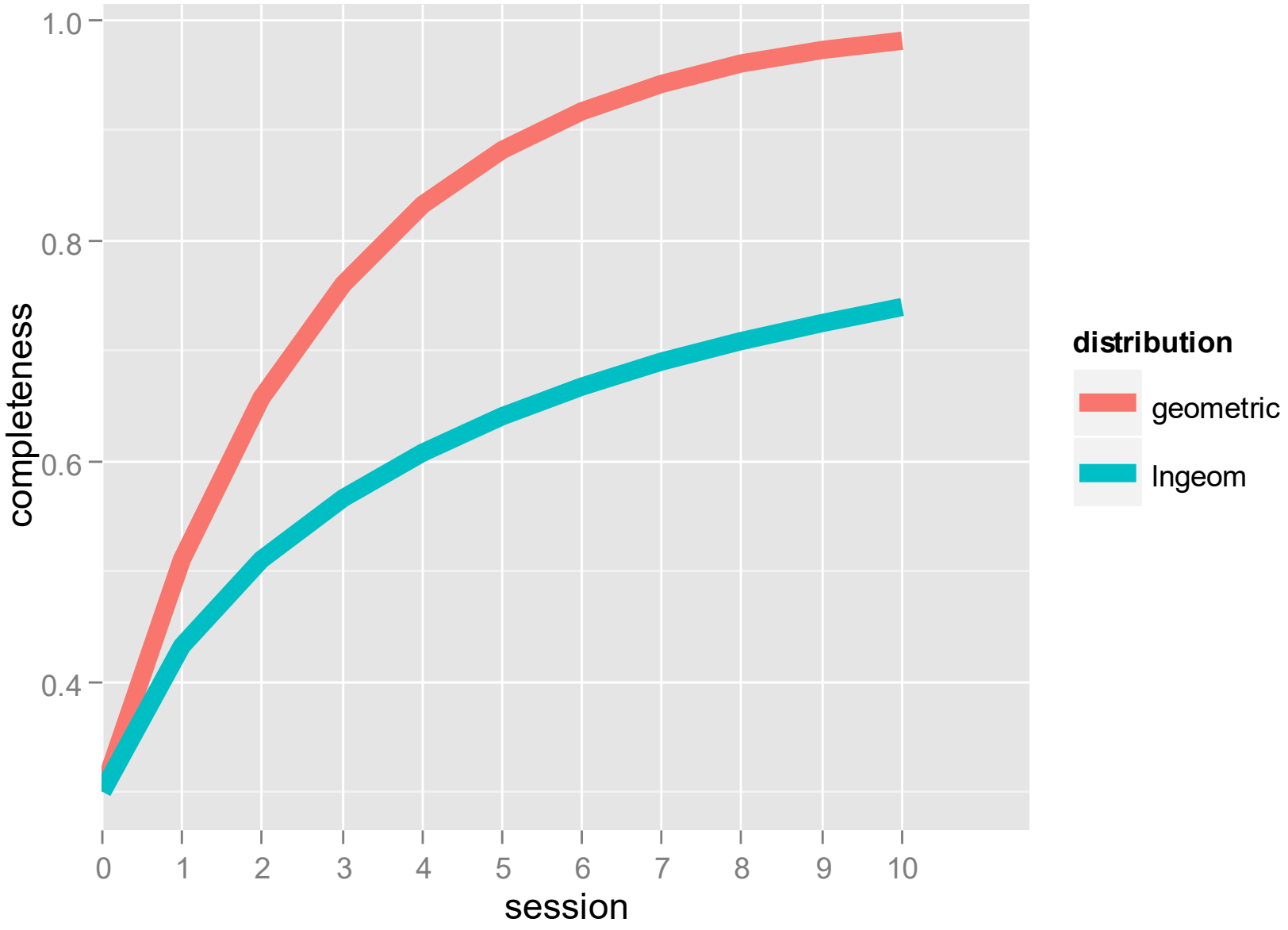
logit-normal Binomial $(\mathrm{k}, \mu, \sigma)$

The effect of visibility variance



discovered and undiscovered


probability
0.2
0.1
0.0
-1
2
3
4
5
6
7
8
9
10
11
discovered
missed
JUST discovered
useful replication
futile repetition



progress of discovery

#6 Statistical models can be used to control the discovery process.

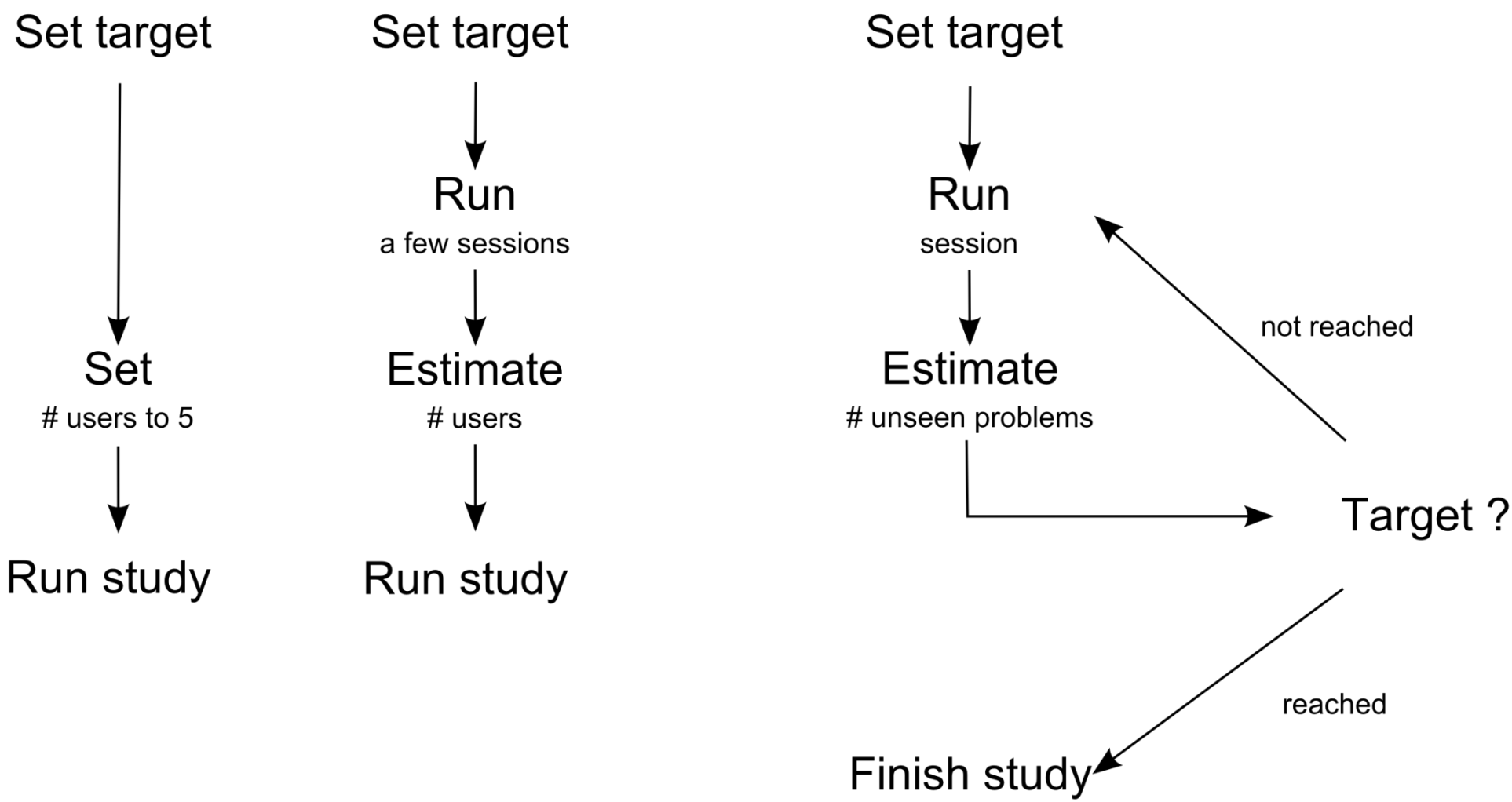
This device is a killer!

- Dozens of killed patients
- Hundreds of harmed patients
- Nurses lost their jobs

- Why? Abysmal usability!

Study 1: Usability Testing a Medical Infusion Pump

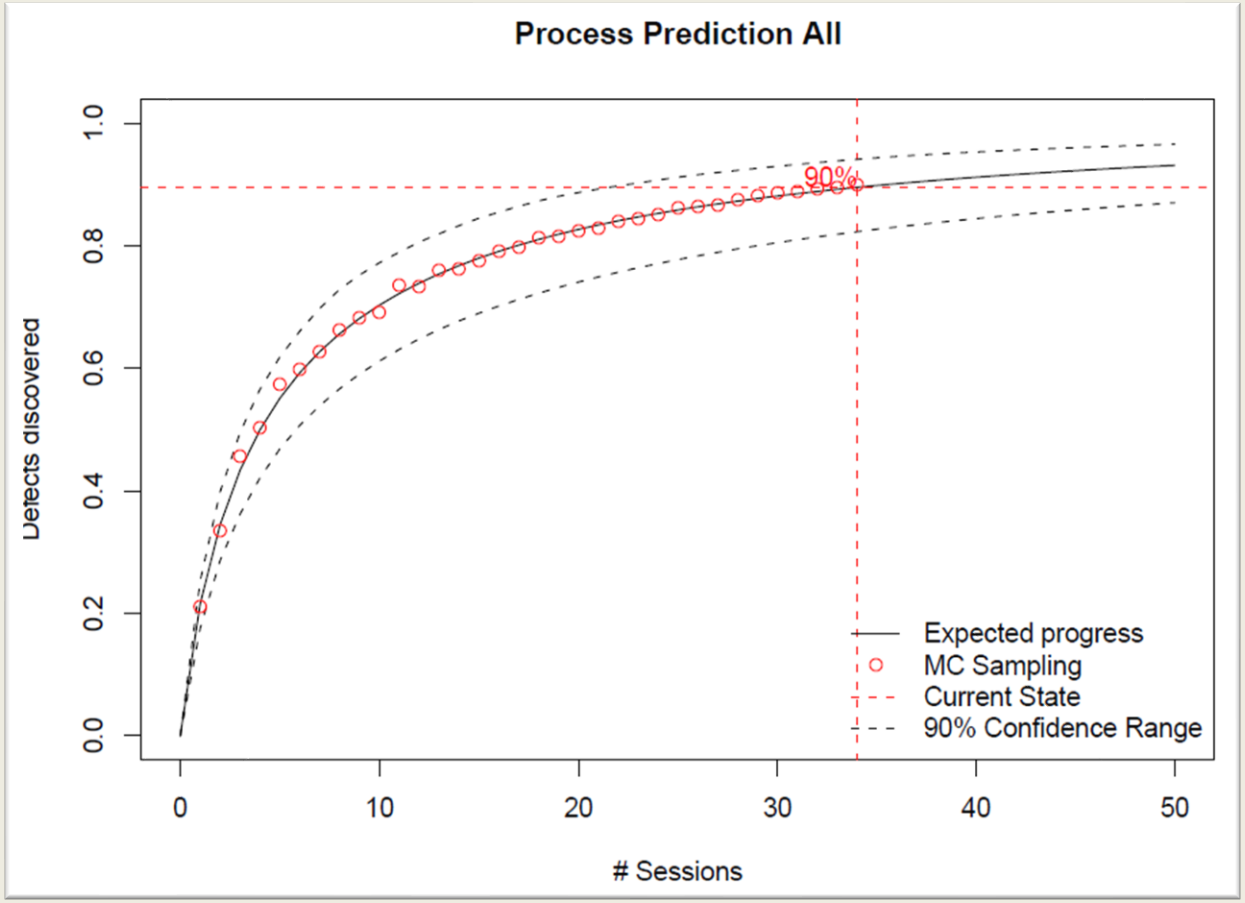
- Prototype developed at TNO
- 34 professionals tested
- 107 usability problems discovered



magic — Nielsen, 2000

early — Lewis, 2003

late — Schmettow, 2012



90% problems with 34 users

ad #4 Magic numbers are not even close.

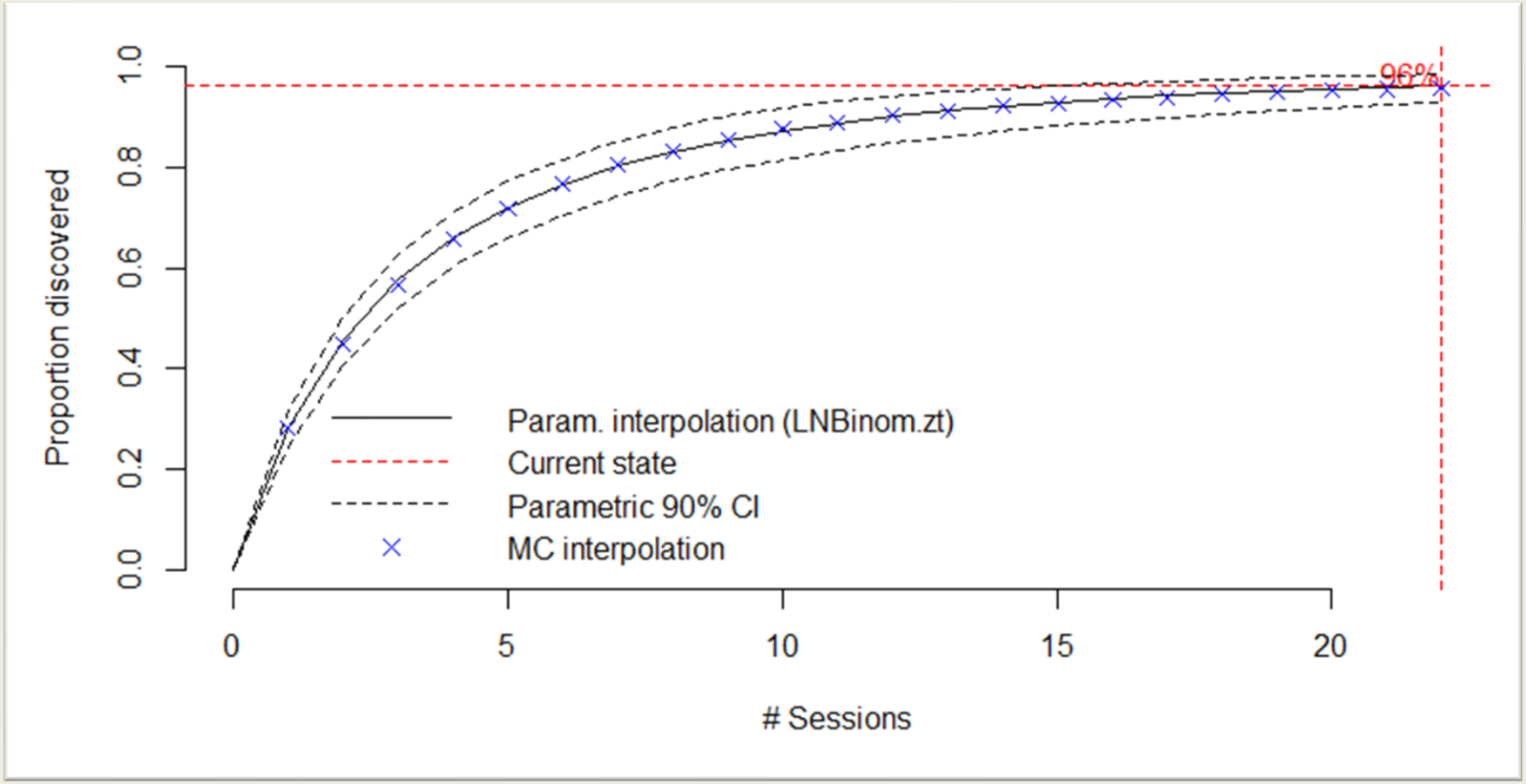
“In interview studies, sample size is often justified by interviewing participants until reaching ‘data saturation’. However, there is no agreed method of establishing this.”

Francis, *et al,* (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & health*, *25*(10), 1229–45.

#7 Discovery process models transfer well to other qualitative elicitation methods.

Study 2: User requirements

- Requirements for a medical information system
- 22 professionals interviewed
- 69 user requirements classified
- Are we complete?



96% requirements discovered with 22 interviews

On a higher level …

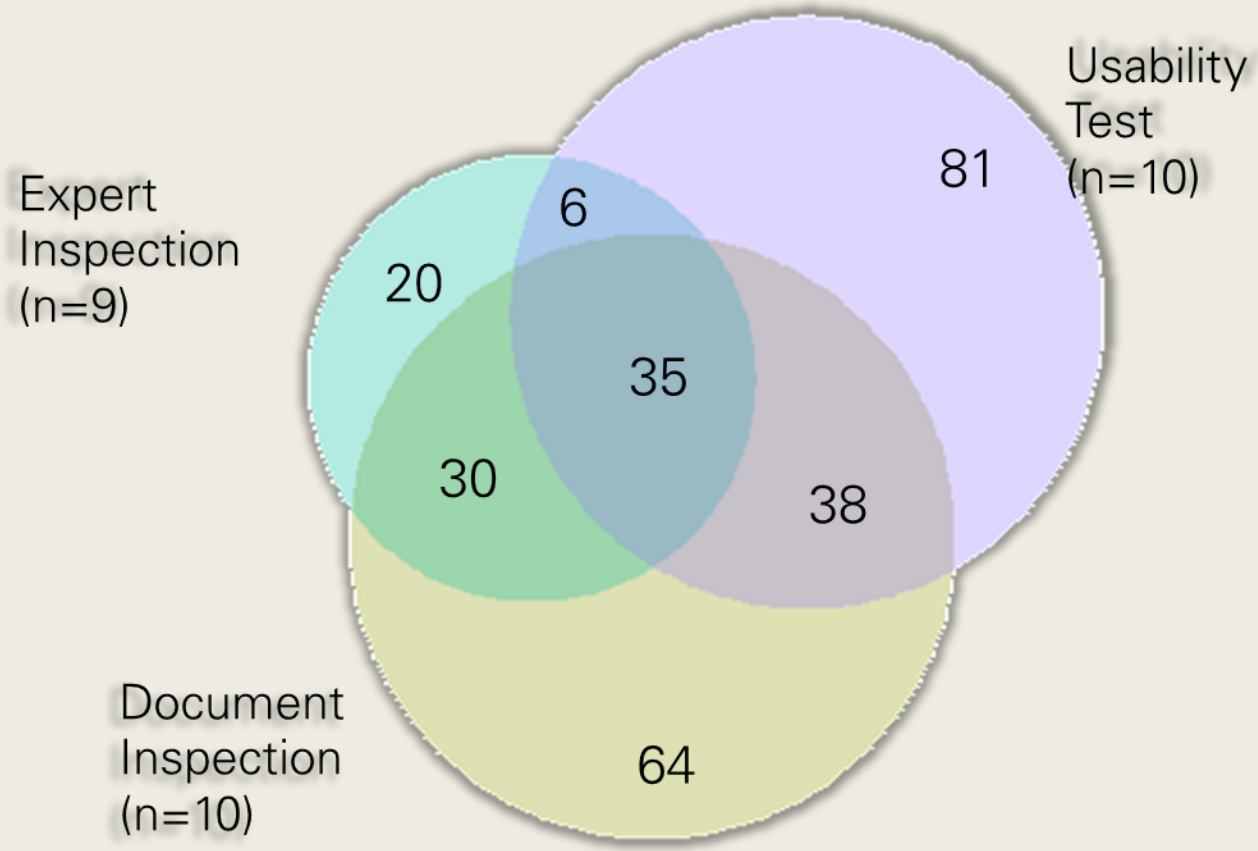
METHOD	Usability Testing	Requirement Elicitation	Expert interviews	Quality Assurance
DISCOVERS	Usability problems	Requirements	Domain concepts	System failures
BY	Users	Stakeholders	Experts	Testers

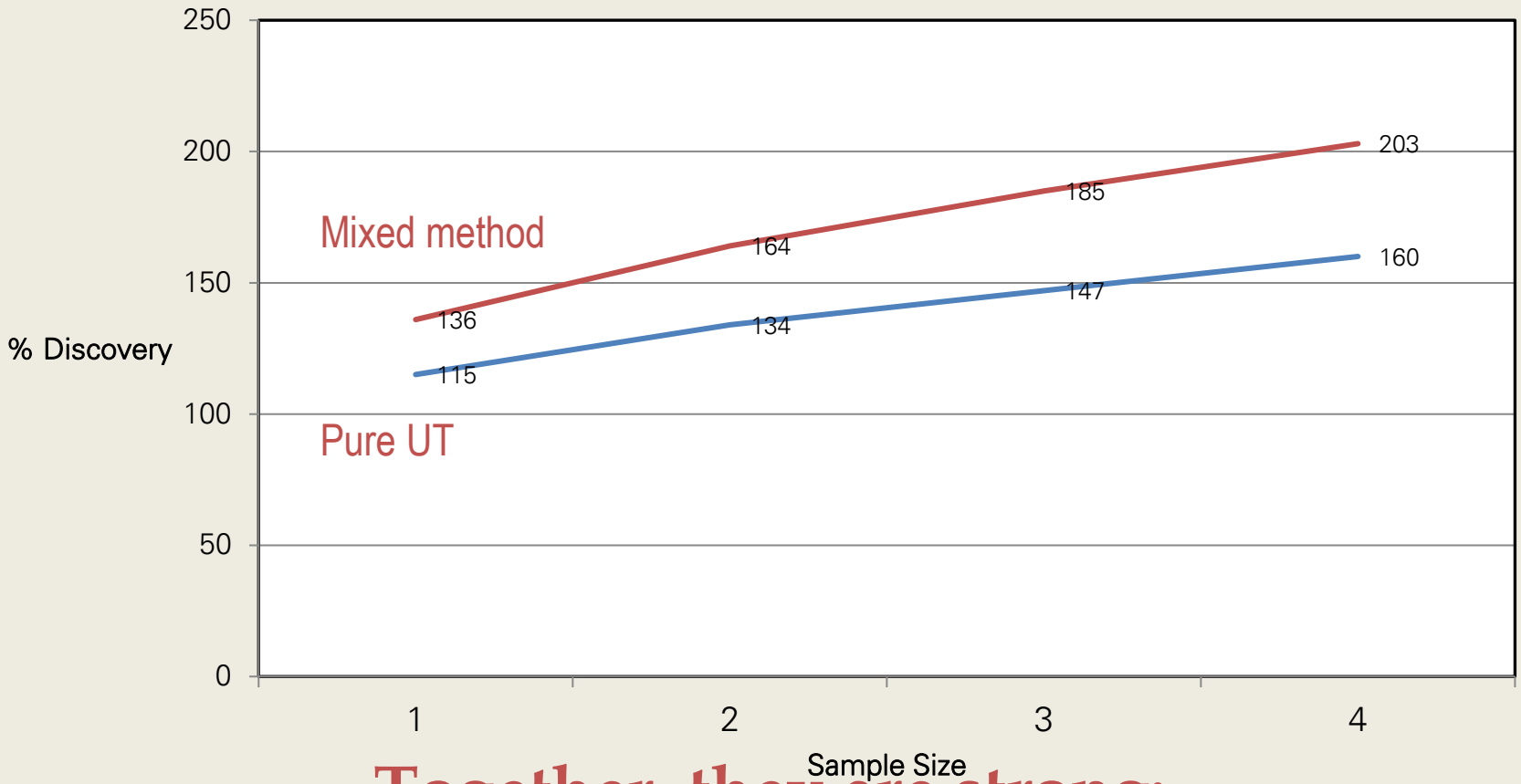
#8 All discovery methods have blind spots and are essentially incomplete.

Study 3:
Comparison of three discovery methods

- two Virtual Environment applications
- 3 evaluation methods
  - Usability Test n=10
  - Document Inspection n=10
  - Expert Inspection n= 9
- Overall problems found: 274

Overlap in problem discovery





Together, they are strong:
20 % better discovery through complementary methods

#9 Mixed-method discovery processes are more effective.

“Outlier data […] is often informative and should be investigated to determine the nature and pattern of the use scenarios associated with them.”

FDA Guidelines on
Medical Device Use-Safety
(2000)

ad #1 People who use the term „outlier“ have not quite understood Murphy‘s law.

#10 Provoking so-called outliers is an efficient way to find all possible way things can go wrong.

Summary

1. Murphy's law
2. Confirmation bias
3. Knowing how things can go wrong
4. Magic numbers are hocus-pocus
5. Visibility variance and incompleteness
6. Statistical control of discovery
7. Domain transfer
8. Blind spots
9. Mixed-methods more effective
10. Provoking outliers

References

Schmettow, M., Bach, C., & Scapin, D. (2014). Optimizing usability studies by complementary evaluation methods. In *Proceedings of the 28th International BCS Human Computer Interaction Conference: Sand, Sea and Sky - Holiday HCI, HCI 2014*. https://doi.org/10.14236/ewic/hci2014.12

Schmettow, M., Vos, W., & Schraagen, J. M. (2013). With how many users should you test a medical infusion pump? Sampling strategies for usability tests on high-risk systems. *Journal of Biomedical Informatics*, *46*(4), 626–641. https://doi.org/10.1016/j.jbi.2013.04.007

Schmettow, M. (2012). Sample size in usability studies. *Communications of the ACM*, *55*(4), 64. doi:10.1145/2133806.2133824

Schmettow, M. (2009). Controlling the usability evaluation process under varying defect visibility. *BCS-HCI '09: Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology* (pp. 188–197). Swinton, UK: British Computer Society.