What Analysis Can't Fix: Why Retention is so Critically Important

Karen Bandeen-Roche, Ph.D.

Departments of Biostatistics, Medicine & Nursing

Johns Hopkins Medical Institutions

Introduction: Why Retention Matters
Things you already know

- Failure of retention creates missing data
- Likely: Those not retained differ systematically from those retained
    - Sicker?
    - Less emotionally robust?
    - Cognitively declining?
- Findings' accuracy, precision may suffer

Introduction
More original points I hope to make

- The frail tend not to be retained

- There all missing data, and missing data

- Not all missing data are analytically fixable

- Study design considerations are crucial

- Representative recruitment matters at least as much as retention

The problem
Women's Health & Aging Study

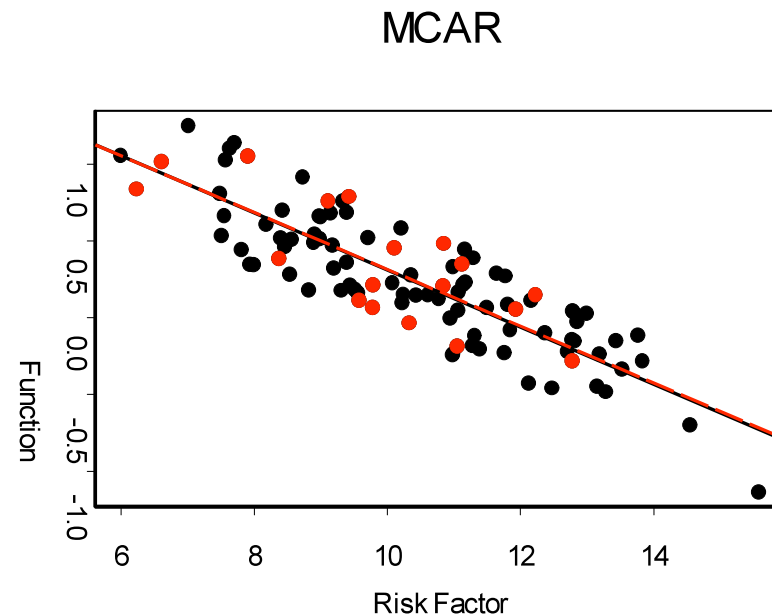| Round 1 Status	Round 3 Status (1 year later)				Round 1 Margin
	Frail	Non-frail	LTF/ NH	Dead	
Frail	149 42%	96 27%	78 22%	33 9%	356 36%
Non-frail	116 18%	403 63%	85 13%	33 5%	637 64%
Round 3 Margin	265 27%/35%	499 50%	163 16%	66 7%	993 100%

Missing data, and Missing data

- A standard hierarchy of three types of missing data (Rubin, 1974)
  - Missing Completely at random (MCAR)
  - Missing at Random (MAR)
  - Not Missing at Random (NMAR)

- The distinctions matter because the type of missing data mechanism determines the analytic sophistication that is needed

Missing Data: Variety 1
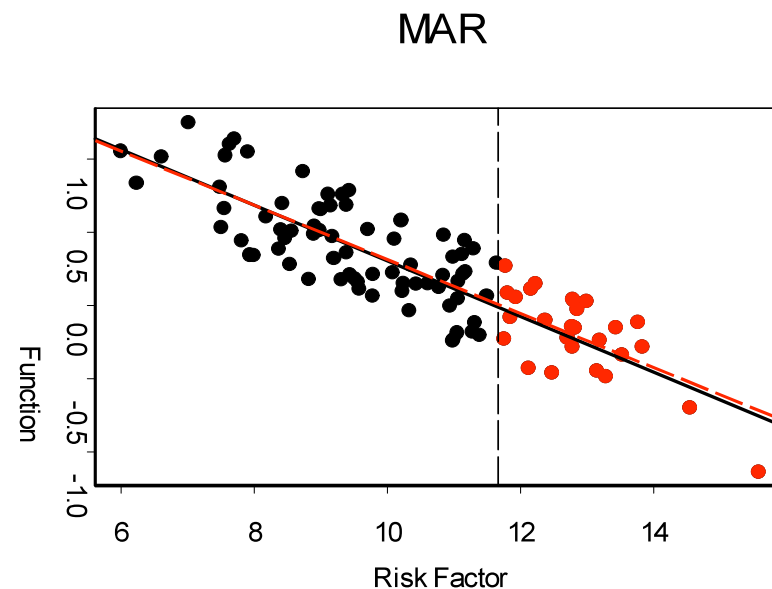
- **Missing completely at random (MCAR)**

  The probability of being currently retained is unrelated to previous or current frailty status or any other measured characteristic.

MCAR

Missing Data: Variety 2
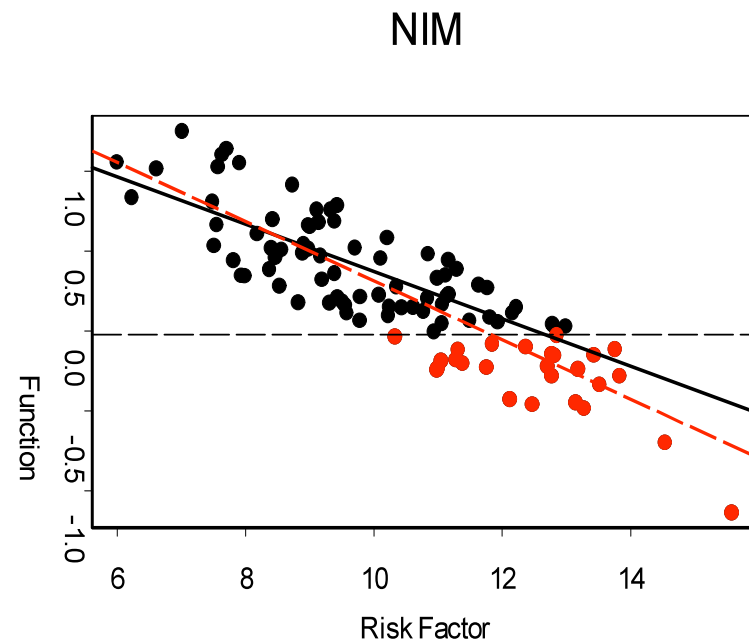
- ## Missing at random (MAR)

  The probability of being currently retained is unrelated to current frailty status, after controlling for observed frailty history and other measured characteristics.

MAR

Missing Data: Variety 3

- ## Non-ignorable missing or not missing at random (NMAR)

  The probability of being currently retained depends on current frailty status even after controlling for all observed characteristics

NIM

Function

Risk Factor

Analytic fixes

- ## MCAR
  - Complete case analysis
  - If item-wise missing: multiple imputation

- ## MAR
  - Key 1: Careful analytic inclusion of variables that predict dropout
    - Even if they are not of primary interest
    - Multiple imputation helpful if "control" not desirable
  - Key 2: Right analytic methods
    - That accurately account for correlation over time

Analytic fixes

- NMAR
  - None

- Debate: Methods for NMAR exist
  - Rebuttal: those methods require
    - External information on relation of current retention to current frailty status
    - Unverifiable statistical assumptions
  - Best one can do: Sensitivity analysis
    - Set bound on reasonable strength of NMAR relation: B
    - Evaluate findings change as strength varies from 0 to B

One more loss issue
Censoring due to death

- Competing risk analysis
- At least: Parallel analyses of incident frailty, death ("cause-specific failures")
- Other options
  - Death as informative censoring (NMAR)
  - Frailty "net" death?
  - Frailty, death as a joint process

Representative recruitment matters too

- Obvious reasons: Already discussed

- Also: "measurement" of frailty
  - Theory implies relationships among criteria
  - If participation particularly contingent on individual criteria, then evaluation of such relationships may be particularly biased

Conclusion
Study design arguably most important

- Measure covariates that might reasonably create an MAR situation
- Foot-in-door questionnaire
- Intensive follow-up for a random subset
- Limit the burden on participants
  - Many smaller studies may be better than a grand, broad-sweeping study