Segmentation Approaches in Public Health

Moya Lynn Alfonso, PhD, MSPH
Center for Social Marketing
Florida Prevention Research Center
University of South Florida
College of Education
June 19 and 20, 2007
Segmentation

- Process
- Start with heterogeneous population (i.e., dataset)
- End with smaller homogeneous segments
- Select those most worth pursuing
Social Marketing

- Segment target audience into homogeneous groups
- Analyze characteristics that discriminate segments, such as knowledge, attitudes, social norms, and behavior
- Identify communication channels specific to each segment
- Develop strategies based on analysis of characteristics of each segment
- Pretest materials and interventions with members of each segment
Foundation - Subgroups

- Which subgroups to give highest priority
  - Primary Audience
    - Whose behavior you want to change
  - Secondary Audience
    - Influencers
    - Gate keepers
Logic of Segmentation

- Begin – undifferentiated sample
- Distinguish between groups using defined characteristics
Statistical Segmentation

- *A priori* segmentation - variables and categories are determined before data are gathered
- Cluster-based - responses to a number of variables are used to determine segments
Looking for Structure

- Multiple regression
- Factor analysis
- Multidimensional scaling
- Discriminant analysis
- Logistic regression
- Log-linear modeling

- Cluster analysis
- Latent class analysis
- Chi-Square Automatic Interaction Detection (CHAID)
CHAI D

- Predictive cluster analysis approach
- Set of independent variables (i.e., predictors)
- Group participants based on responses to a categorical or polytomous dependent variable
Variable Types

- **General observable**
  - Demographics

- **Product (behavior)-specific observable**
  - Frequency

- **General unobservable**
  - Values, beliefs, and attitudes

- **Product (behavior)-specific unobservable**
  - Benefits, preferences, intentions, readiness
Why CHAID

- Used in fields of marketing research and public health
- Can handle a large number of variables
- Designed to identify potentially meaningful patterns in a dataset
- Easy to use and understand
CHAID Benefits

- Normal distribution not required
- Independent variable categories that do not differ statistically significantly are merged
- Exploratory approach
  - Confirm using regression
  - Bonferroni adjustment
- Controls for Type I error rate inflation
CHAID Concerns

- CHAID is a forward stepwise approach
  - Results depend upon the order in which variables enter the model

- Once a predictor enters, it’s in

- Segments developed using statistical criteria
  - Use manual approach for theory

- Irrelevant variables decrease validity

- No agreed upon stopping rule
  - Use effect size (Cramer’s V or Phi (2X2))
CHAI D Technical Information

- Hierarchical, criterion-based approach
- Defines segments based on combinations of predictor variables
- Results: mutually exclusive and exhaustive segments
- Iterative, chi-square test of independence based analysis of the interactions among predictor variables
Advantage Over Regression

- Assumes that the predictor variables will interact
- Enables identification of the most significant predictors from large number
- Simplifies interpretation of complex interactions
3 Components of Analysis

- Categorical or polytomous dependent variable
- Set of predictor variables
- Settings for CHAID parameters
  - Variable classifications (e.g., floating)
  - Stopping criterion (i.e., smallest segment size)
Variable Types

- **Free**
  - No inherent ordering (e.g., occupation)

- **Monotonic**
  - Inherent ordering (ordinal)

- **Floating**
  - Monotonic
  - Last category (missing/unknown) merged with most similar distribution
CHAID Algorithm

- Merging of categories based on their similarity in relation to the dependent variable
- Splitting the overall group on the ‘best’ predictor
  - The lowest statistically significant, Bonferroni adjusted $p$-value
- Returning to the merging step
  - If the stopping criterion has not been met
  - There are more subgroups to analyze
Merging

- Categories are merged within and across independent variables

- Process
  - Two-way cross-tabulations are formed between each independent variable and the dependent variable
  - Categories are merged where appropriate
  - Bonferroni adjusted p-value is calculated for the merged cross-tab
Results of a CHAID Analysis

- A tree diagram
- Gains table
  - Ranks each segment in terms of its likelihood of response to the behavior of interest
- Risk table (i.e., classification matrix)
- Rules (node definition)
- Summary (model information)
Trees

- A root node
- Parent nodes
- Child nodes
- Terminal nodes (segments)
- each node contains
  - Categories (definition)
  - Percentage response for the particular group
  - Sample size for the group

Total Sample
Yes, injured: 25%
No self-injury: 75%
n=2000

Male
Yes, injured: 15%
No self-injury: 85%
n=1000

Female
Yes, injured: 35%
No self-injury: 65%
n=1000
Segmenting Youth

Self-injury During Early Adolescence
The “silent school crisis”

I’m gonna draw a picture,
A picture with a twist!
I’ll draw it with a razor blade.
I’ll draw it on my wrist.
And with the little picture,
A fountain will appear.
And with this flowing fountain,
All my problems disappear!!
Definition

- Self-mutilation is a direct, socially unacceptable, repetitive behavior that causes minor to moderate physical injury; when self-mutilating, the individual is in a psychologically disturbed state but is not attempting suicide or responding to a need for self-stimulation or a stereotypic behavior characteristic of mental retardation or autism. (Suyemoto, 1998, p. 532)
Moderate/Superficial Self-Injury

- Behavioral condition/multiple causes & functions
- Most common form = cutting, burning
- Prior to late 1980s, clinical populations
- No longer located in clinical populations
- Late 1980s self-injury ‘came out’ in media
- “Fastest-growing adolescent behavioral problem”
Why So High in Adolescence?

- Self-injury (the behavior)
  - Offers youth benefits depending on needs
  - Self-injury fits
    - experimentation, imitation, rebellion against institutions that seek to control, shock

- Early adolescents: receptive to influence, impulsive, and difficulties regulating emotion and coping with stress

- Environment: models, media exposure, Internet, social reinforcement
Public Health Significance

- Cultural/peer contagion
  - More youth are now exposed

- Substantial proportion lack support or adaptive coping skills

- Subset at risk for chronic behavioral condition
  - At increased risk of suicide
  - Long-term, negative outcomes and limited possibilities

- Vital steps
  - Determine scope of problem
  - Identify risk and protective factors (segments at risk)
Sample Size

- 2,350 surveys distributed
- 2,003 valid surveys (i.e., not Christmas treed) were completed
- Initial response rate of 85.23%
- 1,907 students (~81% of the original sample) self-reported attendance at one of the 8 regular middle schools
Instrumentation

- 2005 Youth Risk Behavior Survey (middle school version)
- 104 items classroom-based survey
- Monitor priority health-risk behaviors
Predictors

- Gender
- Grade
- Ethnicity
- Media Exposure
- Bullying Experience
- Attitudes toward School
- Parent Communication
- Inhalant use
- Peer self-injury
- Substance use
- Abnormal Eating Behaviors
- Deviance
- Suicidal tendencies
Instrumentation

- 2005 Youth Risk Behavior Survey (middle school version)
- 104 items
- Monitor priority health-risk behaviors
Measures of Self-Injury

Lead in: The next 3 questions ask about self-harm (cutting, scratching, burning, not allowing wounds to heal, pinching). Sometimes people who feel upset hurt themselves on purpose as a way to feel better (less upset).

- Have you ever hurt yourself on purpose (cutting, scratching, burning, not allowing wounds to heal, pinching)? Yes/No
- During the past month, how often have you hurt yourself on purpose (cutting, scratching, burning, not allowing wounds to heal, pinching)? Never/1 time/2 or 3 different times/4 or 5 different times/6 or more different times
- Have any of your friends hurt themselves on purpose (cutting, scratching, burning, not allowing wounds to heal, pinching)? Yes/No
Description of Self-injury in General Middle School Population
Self-injury Prevalence

- 28.4% (± 2.1%, 95% CI)
  - High because of broad definition

- Significant but negligible relationship (i.e., .07) with gender

- Not associated with race or ethnicity, grade level, age, or school attended
30 Day Frequency of Self-injury among Youth Who Tried Self-injury

- Never: 30.5%
- One time: 35%
- Two or three times: 18%
- Four or five times: 5.5%
- Six or more times: 11%
Frequency of Self-injury

- Not associated with race or ethnicity, grade level, age, or school attended
Peer Self-injury

- 46.8% knew a friend who had harmed themselves on purpose
- Varied across schools
Identification of Meaningful Segments of Youth who Self-injure
Selection Criteria

- **Potential impact**
  - Size of group
  - Need or benefit received

- **Responsiveness**
  - Readiness to change
  - Likely response to your efforts
Sources of Segmentation Validity Evidence

- Use of theory and applied knowledge in developing segmentations
- Use of holdout samples
  - Randomly splitting the original sample into two separate samples
- Predictive validity studies