

**SAMPLE SURVEY DESIGN AND ANALYSIS:
A COMPREHENSIVE THREE-DAY COURSE
WITH APPLICATION TO MONITORING AND EVALUATION**

by

Joseph George Caldwell, PhD
503 Chastine Drive
Spartanburg, SC 29301-5977 USA
(001)(864)439-2772
jcaldwell9@yahoo.com
<http://www.foundationwebsite.org>

COURSE NOTES: DAY THREE

SPECIAL TOPICS / PRACTICAL PROBLEMS IN SURVEY DESIGN

3 April 2007
(Updated 4 May 2007, 26 March 2009)

NO RECORDING DEVICES ALLOWED

© 1980 - 2009 Joseph George Caldwell. All rights reserved.

Posted at Internet website
<http://www.foundationwebsite.org/SampleSurvey3DayCourseDayThree.pdf> . May be
copied or reposted for noncommercial use (or for evaluation by those considering
attending the course), with attribution.

DAY 3: SPECIAL TOPICS / PRACTICAL PROBLEMS IN SURVEY DESIGN

SURVEY DESIGN FOR MONITORING AND EVALUATION

INSTRUMENTATION, DATA COLLECTION, AND SURVEY FIELD PROCEDURES

PREPARATION OF OMB CLEARANCE FORMS

LONGITUDINAL SURVEYS

SAMPLE FRAME PROBLEMS

SAMPLING FOR RARE ELEMENTS

TREATMENT OF NONRESPONSE

NONSAMPLING ERRORS

RANDOMIZED RESPONSE

RANDOM DIGIT DIALING

MAJOR NATIONAL AND INTERNATIONAL SURVEYS

QUESTIONS AND ANSWERS

I. SURVEY DESIGN FOR MONITORING AND EVALUATION

1. CONCEPTS IN MONITORING AND EVALUATION

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD, *GLOSSARY OF KEY TERMS IN EVALUATION AND RESULTS-BASED MANAGEMENT*, 2002):

MONITORING IS A CONTINUOUS FUNCTION THAT USES THE SYSTEMATIC COLLECTION OF DATA ON SPECIFIED INDICATORS TO PROVIDE MANAGEMENT AND THE MAIN STAKEHOLDERS OF AN ONGOING DEVELOPMENT INTERVENTION WITH INDICATIONS OF THE EXTENT OF PROGRESS AND ACHIEVEMENT OF OBJECTIVES AND PROGRESS IN THE USE OF ALLOCATED FUNDS.

EVALUATION IS THE SYSTEMATIC AND OBJECTIVE ASSESSMENT OF AN ONGOING OR COMPLETED PROJECT, PROGRAM, OR POLICY, INCLUDING ITS DESIGN, IMPLEMENTATION, AND RESULTS. THE AIM IS TO DETERMINE THE RELEVANCE AND FULFILLMENT OF OBJECTIVES, DEVELOPMENT EFFICIENCY, EFFECTIVENESS, IMPACT, AND SUSTAINABILITY. AN EVALUATION SHOULD PROVIDE INFORMATION THAT IS CREDIBLE AND USEFUL, ENABLING THE INCORPORATION OF LESSONS LEARNED INTO THE DECISIONMAKING PROCESS OF BOTH RECIPIENTS AND DONORS.

OTHER TERMS FOR EVALUATION: EVALUATION RESEARCH, EVALUATIVE RESEARCH, EFFECTIVENESS EVALUATION, SUMMATIVE EVALUATION.

OTHER TERMS FOR MONITORING: PROCESS EVALUATION, FORMATIVE EVALUATION.

MONITORING FOCUSES ON DIRECT OUTPUTS; EVALUATION FOCUSES ON OUTCOMES AND IMPACTS (RESULTS RELATING TO ACHIEVEMENT OF THE ULTIMATE GOALS).

WHAT WAS ACCOMPLISHED, FOR WHAT GROUPS, BY WHAT MEANS, AT WHAT COST?

IMPORTANCE OF INDICATORS: "WHAT GETS MEASURED GETS DONE."

MONITORING IS CONCERNED MAINLY WITH MEASUREMENTS OF INPUTS, ACTIVITIES AND OUTPUTS. EVALUATION IS CONCERNED WITH OUTCOMES AND IMPACTS. BOTH ARE CONCERNED WITH EFFICIENCY AND EFFECTIVENESS.

MAJOR TYPES OF EVALUATION STUDIES

- PROCESS EVALUATION (FORMATIVE EVALUATION; EFFICIENCY)
- OUTCOME EVALUATION
- IMPACT EVALUATION (SUMMATIVE EVALUATION; EFFECTIVENESS)

SAMPLE SURVEY IS ESSENTIAL TO BOTH MONITORING AND EVALUATION, E.G., THE DEVELOPMENT OF SAMPLING PLANS FOR MONITORING AGENCY OPERATIONS AND COMPLIANCE WITH STANDARDS, AND FORMAL INVESTIGATIONS OF PROJECT OR PROGRAM IMPACT.

SAMPLE SURVEY CAN SUPPORT BOTH MONITORING (DESCRIPTIVE SURVEY DESIGN, TO PRODUCE DESCRIPTIVE STATISTICS ON A POPULATION AND SUBPOPULATIONS OF INTEREST) AND EVALUATION (ANALYTICAL SURVEY DESIGN, TO DEVELOP MODELS OF A PROCESS AND TEST HYPOTHESES ABOUT IT (E.G., GENERAL LINEAR STATISTICAL MODELS SUCH AS EXPERIMENTAL DESIGNS, QUASI-EXPERIMENTAL DESIGNS AND MULTIPLE REGRESSION MODELS).

EVALUATION -- MEASURING HOW WELL A PROGRAM HAS ACHIEVED ITS OBJECTIVES, OR WHAT CHANGES HAVE TAKEN PLACE AS A RESULT OF THE PROGRAM.

CAROL H. WEISS (*EVALUATING ACTION PROGRAMS*, ALLYN & BACON, 1972):

1. FINDING OUT THE GOALS OF A PROGRAM
2. TRANSLATING THE GOALS INTO MEASURABLE INDICATORS OF GOAL ACHIEVEMENT
3. COLLECTING DATA ON THE INDICATORS FOR THOSE WHO HAVE BEEN EXPOSED TO THE PROGRAM
4. COLLECTING SIMILAR DATA ON AN EQUIVALENT GROUP THAT HAS NOT BEEN EXPOSED TO THE PROGRAM (CONTROL GROUP)
5. COMPARING THE DATA ON PROGRAM PARTICIPANTS AND CONTROLS IN TERMS OF GOAL CRITERIA

JOSEPH S. WHOLEY (*FEDERAL EVALUATION POLICY*, URBAN INSTITUTE, 1975): EVALUATION (1) ASSESSES THE EFFECTIVENESS OF AN ON-GOING PROGRAM IN ACHIEVING ITS OBJECTIVES, (2) RELIES ON THE PRINCIPLES OF RESEARCH DESIGN TO DISTINGUISH A PROGRAM'S EFFECTS FROM THOSE OF OTHER FORCES WORKING IN A SITUATION, AND (3) AIMS AT PROGRAM IMPROVEMENT THROUGH A MODIFICATION OF CURRENT OPERATIONS.

REFERENCES:

1. RUBIN, ALLEN AND EARL BABBIE, *RESEARCH METHODS FOR SOCIAL WORK*, 3rd EDITION, BROOKS/COLE, 1997
2. JODY ZALL KUSEK & RAY C. RIST, *TEN STEPS TO A RESULTS-BASED MONITORING AND EVALUATION SYSTEM* (WORLD BANK, 2004)

EXAMPLES FROM KUSEK/RIST:

MONITORING AND EVALUATION ARE CONCERNED WITH:

- GOALS (LONG-TERM, WIDESPREAD IMPROVEMENTS IN SOCIETY)
- OUTCOMES (INTERMEDIATE EFFECTS OF OUTPUTS ON CLIENTS)
- OUTPUTS (PRODUCTS AND SERVICES PRODUCED)
- ACTIVITIES (TASKS PERSONNEL UNDERTAKE TO TRANSFORM INPUTS TO OUTPUTS)
- INPUTS (FINANCIAL, HUMAN AND MATERIAL RESOURCES)

MILLENNIUM DEVELOPMENT GOALS:

1. ERADICATE EXTREME POVERTY AND HUNGER
2. ACHIEVE UNIVERSAL PRIMARY EDUCATION
3. PROMOTE GENDER EQUALITY AND EMPOWER WOMEN
4. REDUCE CHILD MORTALITY
5. IMPROVE MATERNAL HEALTH
6. COMBAT HIV/AIDS, MALARIA AND OTHER DISEASES
7. ENSURE ENVIRONMENTAL SUSTAINABILITY
8. DEVELOP A GLOBAL PARTNERSHIP FOR DEVELOPMENT

EXAMPLE OF MILLENNIUM DEVELOPMENT GOAL, TARGETS AND INDICATORS

GOAL: ERADICATE EXTREME POVERTY AND HUNGER

- TARGET 1: HALVE, BETWEEN 1990 AND 2015, THE PROPORTION OF PEOPLE WHOSE INCOME IS LESS THAN USD1 PER DAY
 - INDICATOR 1: PROPORTION OF POPULATION BELOW USD1 PER DAY
 - INDICATOR 2: POVERTY GAP RATIO (INCIDENCE X DEPTH OF POVERTY)
 - INDICATOR 3: SHARE OF POORTEST QUINTILE IN NATIONAL CONSUMPTION
- TARGET 2: HALVE, BETWEEN 1990 AND 2015, THE PROPORTION OF PEOPLE WHO SUFFER FROM HUNGER
 - INDICATOR 4: PREVALENCE OF UNDERWEIGHT CHILDREN (UNDER 5 YEARS OF AGE)
 - INDICATOR 5: PROPORTION OF POPULATION BELOW MINIMUM LEVEL OF DIETARY ENERGY CONSUMPTION

EXAMPLE OF INPUTS, ACTIVITIES, OUTPUTS, OUTCOME AND GOAL
(KUSEK/RIST)

- GOAL: REDUCE MORTALITY RATES FOR CHILDREN UNDER 5 YEARS OLD
- OUTCOME: IMPROVE USE OF ORAL REHYDRATION THERAPY (ORT) FOR MANAGING CHILDHOOD DIARRHEA
- OUTPUTS:
 - 15 MEDIA CAMPAIGNS COMPLETED
 - 100 HEALTH PROFESSIONALS TRAINED
 - INCREASED MATERNAL KNOWLEDGE OF ORT SERVICES
 - INCREASED ACCESS TO ORT
- ACTIVITIES:
 - LAUNCH MEDIA CAMPAIGN TO EDUCATE MOTHERS
 - TRAIN HEALTH PROFESSIONALS IN ORT
- INPUTS:
 - TRAINERS
 - ORT SUPPLIES
 - FUNDS
 - PARTICIPANTS

TYPES OF EVALUATIONS (KUSEK/RIST, WORLD BANK'S *MONITORING AND EVALUATION: SOME TOOLS, METHODS AND APPROACHES*):

- PERFORMANCE LOGIC CHAIN ASSESSMENT
 - DETERMINE THE STRENGTH AND LOGIC OF THE CAUSAL MODEL BEHIND THE POLCY, PROGRAM OR PROJECT
- PRE-IMPLEMENTATION ASSESSMENT
 - ARE THE OBJECTIVES WELL-DEFINED SO THAT OUTCOMES CAN BE STATED IN MEASURABLE TERMS?
 - IS THERE A COHERENT AND CREDIBLE IMPLEMENTATION PLAN THAT PROVIDES CLEAR EVIDENCE OFHOW IMPLEMENTATIONIS TO PROCEED AND HOW SUCCESSFUL IMPLEMENTATION CAN BE DISTINGUISHED FROM POOR IMPLEMENTATION?
 - IS THE RATIONALE FOR THE DEPLOYMENT OF RESOURCES CLEAR AND COMMENSURATE WITH THE REQUIREMENTS FOR ACHIEVING THE STATED OUTCOMES?
- PROCESS IMPLEMENTATION ASSESSMENT
 - FOCUSES ON IMPLEMENTATION DETAILS
 - WHAT WAS PLANNED TO BE IMPLEMENTED
 - WHAT GOT IMPLEMENTED
 - COST AND TIME EXPENDITURES, PROBLEMS
- RAPID APPRAISAL
 - KEY INFORMANT SURVEYS
 - FOCUS GROUP INTERVIEWS
 - COMMUNITY INTERVIEWS
 - STRUCTURED DIRECT OBSERVATION
 - SURVEYS
- CASE STUDY
 - IN-DEPTH INVESTIGATION
- IMPACT EVALUTION
 - "CLASSIC" EVALUATION: WHAT CHANGES OCCURRED, AND TO WHAT CAN THOSE CHANGES BE ATTRIBUTED
- META-EVALUATION
 - SUMMARIZE PREVIOUS EVALUATIONS
- COST-BENEFIT ANALYSIS AND COST-EFFECTIVENESS ANALYSIS
 - COST-BENEFIT ANALYSIS REQUIRES BOTH INPUTS AND OUTPUTS IN MONETARY TERMS (E.G., COST-BENEFIT ANALYSIS OF DAY CARE FROM CLIENT, GOVERNMENT AND SOCIETAL VIEWPOINTS)
 - COST-EFFECTIVENESS ANALYSIS ESTIMATES INPUTS IN MONETARY TERMS AND OUTCOMES IN NONMONETARY QUANTITATIVE TERMS (E.G., IMPROVEMENS IN STUDENTS' READING SCORES)
 - DETAILED FINANCIAL/ECONOMIC/OPTIMIZATION MODELS. LAGRANGE MULTIPLIERS / SHADOW VALUES / SHADOW PRICES
 - HIGHLY TECHNICAL: SKILLS IN FINANCE, ECONOMICS, PUBLIC FINANCE, OPTIMIZATION THEORY

ESSENTIAL ACTIONS IN BUILDING AN M&E SYSTEM (KUSEK/RIST):

- FORMULATE OUTCOMES AND GOALS
- SELECT OUTCOME INDICATORS TO MONITOR
- GATHER BASELINE INFORMATION ON THE CURRENT SITUATION
- SET SPECIFIC TARGETS TO REACH, AND DATES FOR REACHING THEM
- REGULARLY COLLECT DATA TO ASSESS WHETHER THE TARGETS ARE BEING MET
- ANALYZE AND REPORT THE RESULTS

KUSEK/RIST 10 STEPS TO BUILDING A RESULTS-BASED M&E SYSTEM:

1. CONDUCT A READINESS ASSESSMENT (EVALUABILITY ASSESSMENT)
2. AGREE ON OUTCOMES TO MONITOR AND EVALUATE
3. SELECT KEY INDICATORS TO MONITOR OUTCOMES
4. COLLECT BASELINE DATA ON INDICATORS
5. PLAN FOR IMPROVEMENT (SELECT RESULTS TARGETS)
6. MONITOR RESULTS
7. CONDUCT EVALUATIONS
8. REPORT FINDINGS
9. USE FINDINGS
10. IMPLEMENT SUSTAINABLE M&E SYSTEM WITHIN THE ORGANIZATION

CONSISTENT WITH QUALITY MANAGEMENT STANDARDS (DEMING, ISO 9000 QUALITY MANAGEMENT, CMU SEI CMM).

QUESTIONS TO ADDRESS IN BUILDING BASELINE INFORMATION (KUSEK/RIST):

1. WHAT ARE THE SOURCES OF DATA?
2. WHAT ARE THE DATA COLLECTION METHODS?
3. WHO WILL COLLECT THE DATA?
4. HOW OFTEN WILL THE DATA BE COLLECTED?
5. WHAT IS THE COST AND DIFFICULTY TO COLLECT THE DATA?
6. WHO WILL ANALYZE THE DATA?
7. WHO WILL REPORT THE DATA?
8. WHO WILL USE THE DATA?

COMPARISON OF MAJOR DATA COLLECTION METHODS (KUSEK/RIST)

Characteristic	Data collection method			
	Review of program records	Self-administered questionnaire	Interview	Rating by trained observer
Cost	Low	Moderate	Moderate to high	Depends on availability of low-cost observers
Amount of training required for data collectors	Some	None to some	Moderate to high	Moderate to high
Completion time	Depends on amount of data needed	Moderate	Moderate	Short to moderate
Response rate	High, if records contain needed data	Depends on how distributed	Generally good to moderate	High

HISTORICAL BACKGROUND ON MONITORING AND EVALUATION

MUCH WORK IN METHODOLOGIES FOR EVALUATION SINCE THE 1920s (STATISTICS, EXPERIMENTAL DESIGN, SAMPLE SURVEY, ECONOMETRICS).

HEAVY GOVERNMENT INTEREST AND INVESTMENT IN MONITORING AND EVALUATION SINCE THE 1960s. E.G., PLANNING-PROGRAMMING-BUDGETING (PPB, 1965). GENERAL ACCOUNTING OFFICE (GAO). OFFICE OF MANAGEMENT AND BUDGET (OMB).

US GOVERNMENT STANDARDS FOR ACCOUNTABLE MANAGEMENT IN ALL FEDERAL AGENCIES. EVALUATION OF ALL MAJOR SOCIAL AND ECONOMIC PROGRAMS SINCE THE 1960S, WITH THE RENEWED ADVENT OF MASSIVE SOCIAL AND ECONOMIC PROGRAMS, SUCH AS THE "WAR ON POVERTY":

- MATERNAL AND CHILD HEALTH
- VOCATIONAL EDUCATION
- ELEMENTARY AND SECONDARY EDUCATION
- MODEL CITIES
- URBAN RENEWAL
- MANPOWER DEVELOPMENT AND TRAINING
- NEIGHBORHOOD YOUTH CORPS
- WORK INCENTIVE PROGRAM
- HEAD START
- JOB CORPS

FOREIGN ASSISTANCE ACT OF 1961 (AS AMENDED). FORMAL EVALUATION REQUIRED FOR ALL MAJOR US FOREIGN ASSISTANCE PROJECTS

MUCH RESEARCH AND MANY PUBLICATIONS ON MONITORING AND EVALUATION SINCE THE 1960s.

EVALUATION GUIDELINES: *A.I.D. EVALUATION HANDBOOK*

RENEWED INTEREST IN MONITORING AND EVALUATION OF US FEDERAL AGENCIES: THE GOVERNMENT PERFORMANCE AND RESULTS ACT OF 1993 (GPRA) HOLDS FEDERAL AGENCIES ACCOUNTABLE FOR USING RESOURCES WISELY AND ACHIEVING PROGRAM RESULTS.

THE GOVERNMENT PERFORMANCE AND RESULTS ACT OF 1993

GPRA IS NOT LONG (11 PAGES). MAY BE VIEWED AT
<http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html> .

UNDER GPRA, US FEDERAL AGENCIES (HAVING ANNUAL BUDGETS OVER USD20M) MUST:

- DEVELOP PLANS FOR WHAT THEY INTEND TO ACCOMPLISH
- MEASURE HOW WELL THEY ARE DOING
- MAKE APPROPRIATE DECISIONS BASED ON THE INFORMATION THEY HAVE GATHERED
- AND COMMUNICATE INFORMATION ABOUT THEIR PERFORMANCE TO CONGRESS AND THE PUBLIC

GPRA REQUIREMENTS:

- A LONG-RANGE (5-YEAR) STRATEGIC PLAN, WHICH INCLUDES A MISSION STATEMENT, LONG-TERM GOALS AND OBJECTIVES
- ANNUAL PERFORMANCE PLANS, WHICH PROVIDE ANNUAL PERFORMANCE COMMITMENTS TOWARD ACHIEVING THE GOALS AND OBJECTIVES PRESENTED IN THE STRATEGIC PLAN
- ANNUAL PERFORMANCE REPORTS, WHICH EVALUATE AN AGENCY'S PROGRESS TOWARD ACHIEVING PERFORMANCE COMMITMENTS

THE GPRA REQUIREMENTS FORGE LINKS BETWEEN SEVERAL ACTIVITIES:

- PLANNING, TO ACHIEVE GOALS AND OBJECTIVES
- BUDGETING, TO ENSURE THAT RESOURCES ARE AVAILABLE TO CARRY OUT PLANS
- MEASURING, TO ASSESS PROGRESS AND LINK RESOURCES ACTUALLY USED TO RESULTS ACHIEVED
- REPORTING, TO PRESENT PROGRESS ACHIEVED AND IMPACTS ON FUTURE EFFORTS

GPRA WAS INITIATED WITH 10 PILOT PROJECTS IN 1994-96. FIVE AGENCIES SELECTED AS PILOT PROJECTS FOR PERFORMANCE BUDGETING FOR FISCAL YEARS 1998 AND 1999.

A PERFORMANCE BUDGET PRESENTS, FOR ONE OR MORE OF THE MAJOR FUNCTIONS AND OPERATIONS OF THE AGENCY, THE VARYING LEVELS OF PERFORMANCE, INCLUDING OUTCOME-RELATED PERFORMANCE, THAT WOULD RESULT FROM DIFFERENT BUDGETED AMOUNTS.

EXAMPLES OF GPRA ACTIVITIES ON THE INTERNET, E.G., US ENVIRONMENTAL PROTECTION AGENCY (EPA) AT <http://www.epa.gov/ocfo/plan/plan.htm> .

RECENT INCREASES OF INTEREST IN MONITORING AND EVALUATION BY INTERNATIONAL DEVELOPMENT AND ASSISTANCE AGENCIES

IN RECENT YEARS, INTERNATIONAL ORGANIZATIONS SUCH AS THE UNITED NATIONS, THE WORLD BANK, AND THE RED CROSS HAVE INVESTED HEAVILY IN MONITORING AND EVALUATION. IT IS NOW A MAJOR COMPONENT OF ALL PROGRAMS, AND PROJECTS OF SIGNIFICANT SIZE.

MUCH MORE EMPHASIS ON IMPACT EVALUATION.

REFERENCE: *TEN STEPS TO A RESULTS-BASED MONITORING AND EVALUATION SYSTEM*, BY JODY ZALL KUSEK AND RAY C. RIST (WORLD BANK, 2004).

DATA COLLECTION METHODS (FIGURE 4.3)

- CONVERSATION WITH CONCERNED INDIVIDUALS
- COMMUNITY INTERVIEWS
- FIELD VISITS
- REVIEWS OF OFFICIAL RECORDS (MANAGEMENT INFORMATION SYSTEM (MIS) AND ADMINISTRATIVE DATA)
- KEY INFORMANT INTERVIEWS
- PARTICIPANT OBSERVATION
- FOCUS-GROUP INTERVIEWS
- DIRECT OBSERVATION
- QUESTIONNAIRES
- ONE-TIME SURVEY
- PANEL SURVEYS
- CENSUS
- FIELD EXPERIMENTS

SAMPLE SURVEY ADDRESSES THE LAST FIVE METHODS

EVALUATION METHODOLOGIES

THE LOGICAL FRAMEWORK (“LOGFRAME”): AN ANALYTICAL TOOL USED TO PLAN, MONITOR AND EVALUATE PROJECTS.

ORIGINALLY DEVELOPED BY THE US DEPARTMENT OF DEFENSE, USED BY US AGENCY FOR INTERNATIONAL DEVELOPMENT SINCE THE 1960s. USED BY MANY INTERNATIONAL ASSISTANCE ORGANIZATIONS (US, CANADA, AUSTRALIA, UNITED KINGDOM, GERMANY, AND THE EUROPEAN UNION).

THE LOGFRAME IS A MATRIX:

	Narrative Summary	Objectively Verifiable Indicators	Means of Verification	Important Assumptions
Goal				
Purpose (or Objective)				
Outputs				
Activities		Inputs		

(Source: “Engendering the Logical Framework,” by Helen Hambly Odame, International Service for National Agricultural Research (ISNAR), May 25, 2000.

THE BALANCED SCORECARD

PROPOSED BY ROBERT S. KAPLAN AND DAVID P. NORTON IN 1992: “THE BALANCED SCORECARD: MEASURES THAT DRIVE PERFORMANCE,” *HARVARD BUSINESS REVIEW*, JAN-FEB 1992, 71-79.

THE BALANCED SCORECARD (BSC) IS A MANAGEMENT SYSTEM (MANAGEMENT FRAMEWORK, PERFORMANCE MEASUREMENT SYSTEM), THAT BALANCES STRATEGIC OBJECTIVES / PERSPECTIVES:

- FINANCIAL PERSPECTIVE (“TO SUCCEED FINANCIALLY, HOW SHOULD WE APPEAR TO OUR SHAREHOLDERS”)
 - RETURN ON CAPITAL
 - RETURN ON SALES
 - SALES GROWTH
 - CUSTOMER VALUE ADDED
 - HUMAN RESOURCES (EMPLOYEES)
- EXTERNAL CUSTOMER PERSPECTIVE (“TO ACHIEVE OUR VISION, HOW SHOULD WE APPEAR TO OUR CUSTOMERS”)
 - CUSTOMER SATISFACTION
 - CUSTOMER COMPLAINTS
 - CUSTOMERS LOST / WON (MARKET SHARE)
 - SALES FROM NEW PRODUCTS
 - ON-TIME DELIVERY

- INTERNAL BUSINESS PROCESS PERSPECTIVE (“TO SATISFY OUR SHAREHOLDERS AND CUSTOMERS, WHAT BUSINESS PROCESSES MUST WE EXCEL AT?")
 - ORDER CONVERSION RATE
 - ON-TIME DELIVERY FROM SUPPLIERS
 - COST OF NON-CONFORMANCE
 - LEAD TIME REDUCTION
 - QUALITY MANAGEMENT (STANDARDS, E.G., ISO 9000, CMM)
- LEARNING, GROWTH AND INNOVATION (“TO ACHIEVE OUR VISION, HOW WILL WE SUSTAIN OUR ABILITY TO CHANGE AND IMPROVE?")
 - APPRAISALS COMPLETED ON TIME
 - TRAINING PLANS COMPLETED
 - NEW PRODUCT DEVELOPMENT ON TIME

THE BSC RECOGNIZES THAT MOST ASSETS OF ORGANIZATIONS TODAY ARE NOT FINANCIAL (PHYSICAL ASSETS), AND THAT CONCENTRATING ON FINANCIAL ACCOUNTING WILL NOT GUIDE MOST FIRMS TO WHERE THEY WANT TO GO.

THE CRITICAL MANAGEMENT PROCESSES ARE:

- CLARIFY AND TRANSLATE VISION AND STRATEGY
- COMMUNICATE AND LINK STRATEGIC OBJECTIVES AND MEASURES
- PLAN, SET TARGETS, AND ALIGN STRATEGIC OBJECTIVES
- ENHANCE STRATEGIC FEEDBACK AND LEARNING

ORIGINALLY DEVELOPED TO ASSIST COMMERCIAL BUSINESSES IN TRANSLATING STRATEGIC GOALS INTO ACTION, IT HAS NOW BEEN EMBRACED BY MANY PUBLIC INSTITUTIONS, INCLUDING STATE AND LOCAL GOVERNMENTS AND INTERNATIONAL ORGANIZATIONS.

REFERENCES ON THE BALANCED SCORECARD:

Kaplan, Robert S. and David P. Norton, *The Balanced Scorecard: Translating Strategy into Action*, Harvard Business School Press, 1996

Niven, Paul R., *Balanced Scorecard Step-by-Step*, John Wiley & Sons, 2002 (2nd edition 2006)

Niven, Paul R., *Balanced Scorecard Step-by-Step for Government and Nonprofit Agencies*, John Wiley & Sons, 2003

Nair, Mohan, *Essentials of Balanced Scorecard*, John Wiley & Sons, 2004

Bourne, Mike and Pippa Bourne, *Balanced Scorecard in a Week*, Hodder Arnold, 2000, 2002

EVALUATION GUIDELINES

THERE ARE MANY SOURCES OF INFORMATION ABOUT MONITORING AND EVALUATION ON THE INTERNET, INCLUDING:

<http://www.unaids.org>
<http://www.who.int>
<http://www.cpc.unc.edu/measure>
<http://www.fhi.org>
<http://www.cdc.gov>
<http://www.usaid.gov>

MAJOR EVALUATION GUIDELINES THAT MAY BE DOWNLOADED FROM THE INTERNET INCLUDE:

Monitoring and Evaluation: Some Tools, Methods and Approaches, The World Bank, 2004

National AIDS Councils' Monitoring and Evaluation Operations Manual, The World Bank and United Nations, 2002

Handbook on Monitoring and Evaluating for Results, United Nations Development Programme, 2002

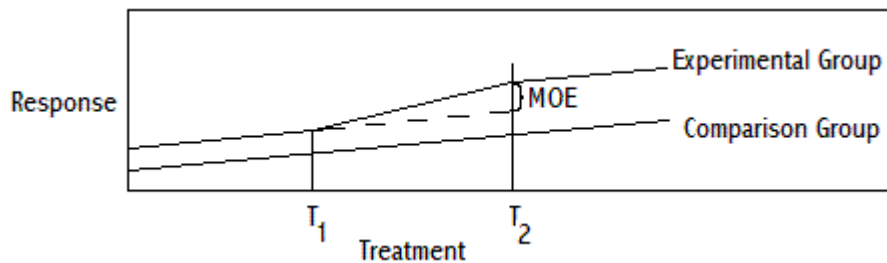
Handbook for Monitoring and Evaluation, International Federations of Red Cross and Red Crescent Societies, 2002

Many additional Internet sources for M&E guidelines are listed at the UNDP website, http://stone.undp.org/undpweb/eo/evalnet/docstore3/yellowbook/template/bibliography/bib_c.htm

THE VARIOUS MONITORING AND EVALUATION GUIDELINES DESCRIBE FRAMEWORKS FOR CONDUCTING EVALUATIONS. THEY IDENTIFY VARIOUS METHODOLOGIES, AND DISCUSS WHEN IT IS APPROPRIATE TO USE THEM. THEY PRESENT LITTLE METHODOLOGICAL INFORMATION ON HOW TO USE THEM (SUCH AS IS ADDRESSED IN THIS COURSE ON SAMPLE SURVEY).

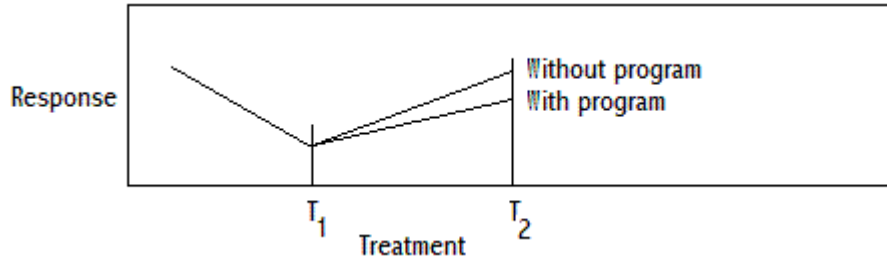
2. MEASURES OF EFFECTIVENESS (MOE's)

- MISSION → GOALS → OBJECTIVES → MOE' S
- PROBLEM IN MEASURING THE EFFECTIVENESS OF A PROGRAM:
 - VALIDITY OF MEASURE (E.G., CHILDREN RETURNED HOME, CHANGE IN EARNINGS BEFORE AND AFTER PROGRAM)
 - HAWTHORNE EFFECT
 - SURVEY DESIGN PROBLEMS
- NEED CONTROLLED EXPERIMENT
- CONTROL OFTEN NOT POSSIBLE
 - DENIAL OF SERVICES
 - AWARENESS OF "TEST"
 - NO CONTROL OVER EXOGENOUS VARIABLES
- INTERACTION VARIABLE OFTEN USED AS MEASURE OF EFFECTIVENESS

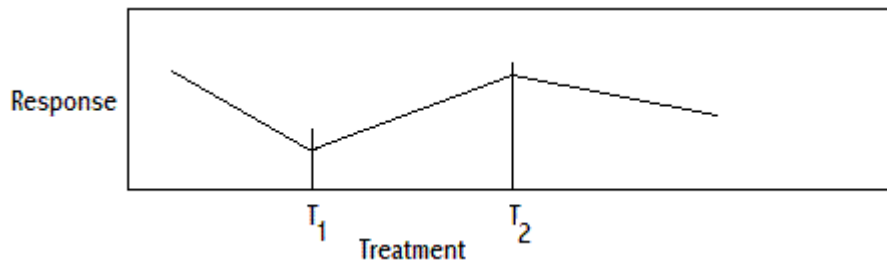


MOE = DIFFERENCE IN RESPONSE FOR EXPERIMENTAL AND COMPARISON GROUPS (INTERACTION OF TREATMENT WITH GROUP)

PRE-POST MEASURES OFTEN SUSPECT



FOLLOW-UP MEASURES OFTEN SUSPECT



DISPLACEMENT EFFECTS (NET IMPACT)

CLIENT GETS JOB, DISPLACES SOMEONE ELSE IN THE LABOR MARKET, NET IMPACT IS ZERO.

EXAMPLE: WAL-MART COMES TO TOWN, AND IT IS ADVERTISED AS SAVING EVERY FAMILY SEVERAL THOUSAND DOLLARS A YEAR, AND GENERATING EMPLOYMENT – BUT WHAT ABOUT THE EMPLOYEES OF ALL OF THE LOCAL HARDWARE STORES WHO NOW HAVE NO JOBS, OR WORK AT A FRACTION OF THEIR PREVIOUS SALARIES / WAGES.

EXAMPLE: EARLY MAMMOGRAMS CAUSED MORE BREAST CANCER THAN THEY DETECTED.

EXAMPLE: BETTER CRIME REPORTING / MONITORING SYSTEM APPEARS TO LEAD TO INCREASE IN CRIME.

3. RESEARCH DESIGN IN EVALUATION RESEARCH

- EXPERIMENTAL DESIGN IS RECOMMENDED (RANDOM ASSIGNMENT OF TREATMENTS TO UNITS)
- QUASI-EXPERIMENTAL DESIGNS AND NON-EXPERIMENTAL DESIGNS OFTEN USED IN SOCIAL SCIENCES RESEARCH FOR "PRACTICAL" REASONS:
 - DENIAL OF TREATMENT OR SERVICES CONSIDERED UNETHICAL OR POLITICALLY INEXPEDIENT
 - NO ADDITIONAL PEOPLE TO SERVE AS CONTROLS (PROGRAM SERVES ALL INTERESTED PERSONS)
 - PARTICIPANTS MAY DROP OUT OF PROGRAM
 - TREATMENT AND CONTROL GROUPS ARE LIKELY TO KNOW THEIR STATUS
- SAMPLE SURVEY OFTEN USED TO IMPLEMENT QUASI-EXPERIMENTAL DESIGNS -- AN "ANALYTICAL SURVEY"
 - DESCRIBE RELATIONSHIPS BETWEEN NUMEROUS INTERVENING VARIABLES (E.G., EDUCATION, AGE, EXPERIENCE, AND QUALITY OF SCHOOL ON EARNINGS)
 - FPC IS IRRELEVANT -- WE WISH TO EVALUATE THE PROCESS GENERATING THE POPULATION NOT DESCRIBE THE CURRENT POPULATION

4. EXPERIMENTAL DESIGN vs SAMPLE SURVEY DESIGN

EXPERIMENTAL DESIGN IS CONCERNED WITH ESTABLISHING CAUSAL RELATIONSHIPS BETWEEN VARIABLES -- THE POPULATION UNDER STUDY IS FORMED BY THE EXPERIMENTER.

DESCRIPTIVE SURVEY DESIGN IS CONCERNED WITH DESCRIPTION OF A GIVEN POPULATION, AND THE MEASUREMENT OF ASSOCIATIONS BETWEEN CHARACTERISTICS IN THE POPULATION. IT IS CONCERNED WITH ESTIMATION OF POPULATION AND SUBPOPULATION CHARACTERISTICS OF INTEREST, SUCH AS MEANS AND TOTALS, NOT WITH TESTS OF HYPOTHESIS.

ANALYTICAL SURVEY DESIGN ("MODEL-BASED," OR "MODEL-ASSISTED" DESIGN) IS CONCERNED WITH DEVELOPING A MODEL OF THE PROCESS THAT (HYPOTHETICALLY) GENERATED THE POPULATION AT HAND. IT IS CONCERNED WITH ESTIMATION OF RELATIONSHIPS AND TESTS OF HYPOTHESIS.

5. CHARACTERISTICS OF A GOOD EXPERIMENTAL DESIGN

HIGH INTERNAL VALIDITY: DID THE EXPERIMENTAL TREATMENTS MAKE A DIFFERENCE IN THE SPECIFIC EXPERIMENT?

HIGH EXTERNAL VALIDITY (GENERALIZABILITY): TO WHAT EXTENT CAN THE RESULTS OF THE EXPERIMENT BE GENERALIZED (TO OTHER POPULATIONS, SETTINGS, CONDITIONS)?

6. THREATS TO INTERNAL VALIDITY

(REFERENCE: DONALD T. CAMPBELL AND JULIAN C. STANLEY, *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH*, RAND McNALLY, 1963)

LACK OF CONTROL OF EXTRANEOUS VARIABLES

- HISTORY (EVENTS OCCURRING DURING THE COURSE OF THE EXPERIMENT)
- MATURATION (AGING)
- TESTING (EFFECTS OF TAKING ONE TEST ON SCORE OF A SECOND TEST)
- INSTRUMENTATION (CHANGES IN CALIBRATION, SCORERS)
- STATISTICAL REGRESSION EFFECTS (MOVEMENT OF GROUP MEAN SCORES CAUSED BY SELECTION ON THE BASIS OF EXTREME SCORES)
- SELECTION BIAS (CHOOSING TREATMENT AND CONTROL GROUPS WITH DIFFERENT CHARACTERISTICS)
- EXPERIMENTAL MORTALITY (DIFFERENTIAL LOSSES FROM TREATMENT AND COMPARISON GROUPS)
- SELECTION - HISTORY (OR -MATURATION OR -TESTING) INTERACTION (EVENTS OCCURRING DURING THE COURSE OF THE EXPERIMENT FOR ONE OF THE COMPARISON GROUPS (E.G., THE EXPERIMENTAL AND QUASI-EXPERIMENTAL GROUPS) BUT NOT THE OTHERS)

RANDOMIZATION OFFERS PROTECTION AGAINST ALL OF THESE THREATS TO INTERNAL VALIDITY.

7. REGRESSION EFFECT (MATCHING)

TEMPTATION: SELECT A COMPARISON GROUP THAT IS SIMILAR TO THE TREATMENT GROUP ON A NUMBER OF VARIABLES, AND COMPARE PERFORMANCE OF TWO GROUPS AT END OF PROJECT.

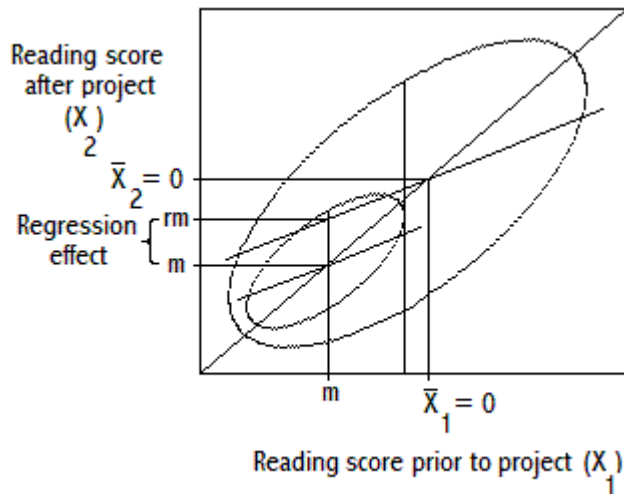
PROBLEM: IF THE SELECTION VARIABLES ARE UNRELIABLE PRE-MEASURES OF OUTCOME (E.G., A PRE-PROJECT READING SCORE), A COMPARISON GROUP SELECTED FROM THE GENERAL (MORE ABLE) POPULATION WILL AUTOMATICALLY IMPROVE, INDEPENDENT OF THE PROGRAM.

NOTE:

IN MATCHING IN AN EXPERIMENTAL DESIGN (PAIRED COMPARISONS) THE UNITS ARE MATCHED AND THEN RANDOMLY ASSIGNED TO THE TREATMENT AND CONTROL GROUPS.

REASON: THE PRETEST MEASURE CONTAINS AN ERROR TERM. ON ANY PARTICULAR TESTING, SOME INDIVIDUALS WILL SCORE PARTICULARLY HIGH OR LOW. ON A RETEST, THE SCORES WILL TEND TO BE NEARER THE MEAN, HENCE, A COMPARISON GROUP SELECTED ON THE BASIS OF EXTREME SCORES WILL REGRESS TO THE MEAN.' INDEPENDENT OF THE PROGRAM EFFECT,

ILLUSTRATION OF THE REGRESSION EFFECT:



IN STANDARDIZED UNITS,

$$E(X_2|X_1) = rX_1$$

IF $r < 1$ AND $X_1 < 0$ (THE MEAN, IN THIS EXAMPLE), THEN

$$E(X_2|X_1) > X_1$$

REGRESSION EFFECT: A GROUP SELECTED ON THE BASIS OF AN IMPRECISE TEST SCORE (I.E., A SCORE CONTAINING RANDOM ERROR, THAT IS NOT PERFECTLY REPEATABLE) WILL TEND TO “REGRESS” TO THE MEAN UPON RETESTING, EVEN THOUGH NO CHANGE HAS OCCURRED IN THEIR ABILITIES OR THE ABILITIES OF THE GENERAL POPULATION. IF THE GROUP HAS BEEN SELECTED BASED ON LOW TEST SCORES, IT WILL TEND TO PERFORM BETTER (ON AVERAGE) UPON RETESTING. IF THE GROUP HAS BEEN SELECTED BASED ON HIGH TEST SCORES, IT WILL TEND TO PERFORM WORSE (ON AVERAGE) UPON RETESTING.

8. THREATS TO EXTERNAL VALIDITY

LACK OF "REPRESENTATIVENESS," OR GENERALIZABILITY:

- INTERACTION EFFECT OF PRETESTING (PRETEST MIGHT INCREASE OR DECREASE SENSITIVITY TO TREATMENT vs, UNPRETESTED POPULATION)
- INTERACTION EFFECTS OF SELECTION BIASES AND THE EXPERIMENTAL VARIABLE (SELECTION BIASES DIFFERENT FOR DIFFERENT TREATMENT GROUPS)
- INTERACTION EFFECTS OF EXPERIMENTAL SETTING (EFFECTS DIFFERENT IN EXPERIMENTAL AND NON-EXPERIMENTAL SETTINGS)
- MULTIPLE TREATMENT INTERFERENCE (CARRY-OVER EFFECTS)

9. NON-EXPERIMENTAL DESIGN

ONE-SHOT CASE STUDY

X O

ONE-GROUP PRETEST-POSTTEST DESIGN

O X O

STATIC GROUP COMPARISON

$\frac{X \quad O}{O}$

10. EXPERIMENTAL DESIGN

USES RANDOMIZATION TO ASSIGN MEMBERS TO EXPERIMENTAL AND CONTROL GROUPS

POSTTEST-ONLY CONTROL GROUP DESIGN:

$$\text{RANDOMIZATION} \begin{cases} X & O \\ & O \end{cases}$$

PRETEST-POSTTEST CONTROL GROUP DESIGN:

$$\text{RANDOMIZATION} \begin{cases} O & X & O \\ O & & O \end{cases}$$

(CAN'T MEASURE INTERACTION EFFECT OF TESTING WITH X)

SOLOMON FOUR-GROUP DESIGN:

$$\text{RANDOMIZATION} \begin{cases} O & X & O \\ O & & O \\ & X & O \\ & & O \end{cases}$$

(CAN MEASURE MAIN EFFECT OF TESTING AND INTERACTION OF TESTING WITH X)

PLUS MANY OTHERS.

11. QUASI-EXPERIMENTAL DESIGNS

TIME SERIES DESIGN

O O O O X O O O O

MULTIPLE TIME SERIES DESIGN

O O O X O O O O
O O O O O O O O

NONEQUIVALENT CONTROL GROUP

O X O
O O

- NO ATTEMPT TO MATCH ON PRETEST
- POSSIBLE MATCH ON BROAD RANGE OF SOCIO-DEMOGRAPHIC CHARACTERISTICS

PLUS MANY OTHERS.

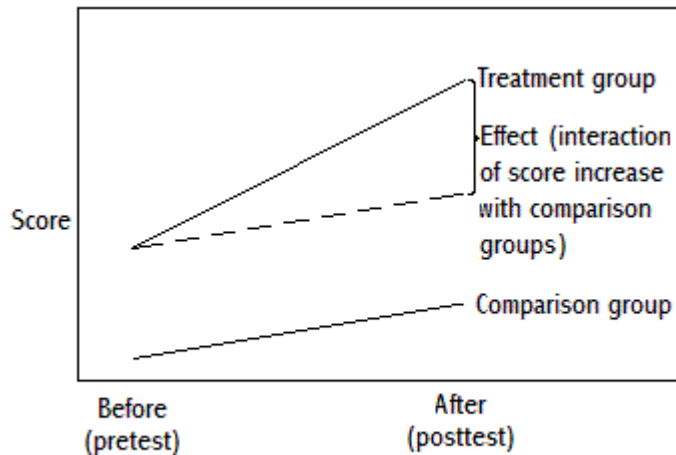
12. SOURCES OF INVALIDITY FOR RESEARCH DESIGNS (REFERENCE: CAMPBELL AND STANLEY, *OP. CIT.*)

Sources of Invalidity for Research Designs												
Sources of Invalidity												
Internal											External	
History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and Treatment	Interaction of Selection and Treatment	Reactive Arrangements	Multiple-Treatment Interference	
Research Design												
Pre-Experimental Designs												
One-Shot Case Study X O	x	x				x	x		x			
One-Group Pretest- Posttest Design O X O	x	x	x	x	(x)		x	x	x	(x)		
Static-Group Comparison X O O		(x)			x	x	x		x			
Experimental Designs												
Pretest-Posttest Control Group Design R O X O R O O								x	(x)	(x)		
Solomon Four-Group Design R O X O R O R X O R O									(x)	(x)		

Posttest-Only Control Group Design R X O R O										(x)	(x)	
Quasi-Experimental Designs												
Time Series O O O O X O O O O	x			(x)					x	(x)	(x)	
Multiple Time Samples Design O O O O X O O O O O O O O O O									x	x	(x)	x
Nonequivalent Control Group Design O X O O O					(x)			x	x	(x)	(x)	

13. ANALYSIS OF DATA FROM QUASI-EXPERIMENTAL DESIGNS
(NONRANDOMIZED COMPARISON GROUP)

1. STRAIGHTFORWARD COMPARISON OF MEANS -- NO GOOD (REGRESSION .EFFECT)
2. ADJUSTMENT OF MEANS BY REGRESSION ANALYSIS OR ANALYSIS OF COVARIANCE -- NO GOOD (USUALLY UNDERADJUSTS)
3. USE OF INTERACTION TERM TO REPRESENT TREATMENT EFFECT -- BEST PROCEDURE



THIS APPROACH ESSENTIALLY AVOIDS THE ISSUE OF ATTEMPTING TO EXPLAIN THE LARGE DIFFERENCES IN MEANS BETWEEN THE TREATMENT AND COMPARISON GROUPS (WHETHER MATCHED OR NOT), AND USES THE DIFFERENCE IN CHANGES IN PERFORMANCE AS THE MEASURE OF EFFECTIVENESS.

STATISTICAL TERMINOLOGY: INTERACTION EFFECT OF TREATMENT AND TIME.

EVALUATION TERMINOLOGY: DOUBLE-DIFFERENCE ESTIMATE OF PROGRAM IMPACT; DIFFERENCE-IN-DIFFERENCE ESTIMATE OF PROGRAM IMPACT.

References on Monitoring and Evaluation

Streuning, Elmer L. and Marcia Guttentag, *Handbook of Evaluation Research*, vols. I and II, Sage Publications, 1975

Weiss, Carol H., *Evaluation Research*, Prentice-Hall, 1972

Suchman, Edward A., *Evaluative Research*, Russell Sage Foundation, 1967

Campbell, D. T. and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, 1966

Cook, Thomas D. and Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Houghton Mifflin Company, 1979

Rubin, Allen and Earl Babbie, *Research Methods for Social Work*, 3rd edition, Brooks/Cole, 1997

Cochran, William G. and Gertrude Cox, *Experimental Designs*, 2nd edition, John Wiley and Sons, 1950, 1957

Kusek, Jody Zall and Ray C. Rist, *Ten Steps to a Results-Based Monitoring and Evaluation System*, World Bank, 2004

Klein, Robert E., et al., editors, *Evaluating the Impact of Nutrition and Health Programs*, Plenum Press, 1979

Schmidt, Richard E. Schmidt, John W. Scanlon and James B. Bell, *Evaluability Assessment: Making Public Programs Work Better*, Project Share Human Services Monograph Series Number 14, November 1979, Project Share, PO Box 2309, Rockville, MD 20852, US Department of Health, Education and Welfare, Office of the Secretary, Office of Assistant Secretary for Planning and Evaluation, DHEW Publication Number OS-76-130.

II. INSTRUMENTATION, DATA COLLECTION AND SURVEY FIELD PROCEDURES

A. SELECTION OF DATA COLLECTION PROCEDURES

TELEPHONE VS. MAIL VS. PERSONAL INTERVIEW

MAIL

- INEXPENSIVE
- LOW RESPONSE RATES (OFTEN LESS THAN 50%)

TELEPHONE

- MODERATE COST
- SOME NONCOVERAGE

PERSONAL INTERVIEW

- USUALLY MOST VALID (NOT ALWAYS, E.G., CLIENT SATISFACTION SURVEY)
- VERY EXPENSIVE (E.G., \$100 PER INTERVIEW)

DATA COLLECTION FORM APPLIED TO RECORDS

- HIGH COST, IF MANUAL SEARCH REQUIRED

USE OF SERVICE WORKERS NOT RECOMMENDED AS INTERVIEWERS

ON-LINE, OR WEB-BASED, SURVEYS (SURVEYS DISPLAYED ON THE INTERNET / WORLD WIDE WEB. FOR MUCH INFORMATION ON THIS TOPIC, SEE THE *ONLINE SURVEY DESIGN GUIDE* AT http://lap.umd.edu/survey_design/ . THIS IS A PROJECT OF THE LABORATORY OF AUTOMATION PSYCHOLOGY AND DECISION PROCESSES (LAP), FOUNDED IN 1983 BY DRS. KENT NORMAN AND NANCY ANDERSON AT THE UNIVERSITY OF MARYLAND. THIS GUIDE PROVIDES GOOD INFORMATION ABOUT QUESTIONNAIRE DESIGN IN GENERAL, NOT JUST FOR ON-LINE SURVEYS.

B. QUESTIONNAIRE DEVELOPMENT

1. QUESTION CONTENT

- SUBSTANTIVE THEORY GUIDES QUESTION CONTENT
- IN MANY SURVEYS, QUESTIONS HAVE HIGH FACE VALIDITY ("HOW OLD ARE YOU")
- IN SOME SURVEYS (E.G., LEVEL OF FUNCTIONING), MANY RELATED QUESTIONS ("DO YOU FEEL LONELY") "HOW OFTEN DOES SOMEONE VISIT YOU," ETC.)
- PERFORM ISSUES ANALYSIS; RELATE QUESTIONS TO ISSUES AND VICE VERSA

2. QUESTION ORDER

- LOGICAL FLOW
- GROUP SIMILAR QUESTIONS
- BREAK UP ALONG INTERVIEW INTO PARTS
- RAPPORT-BUILDING QUESTIONS FIRST
- SENSITIVE QUESTIONS LAST

3. QUESTION WORDING

- LEADING
- BIASED
- DIFFICULT TO UNDERSTAND
- AMBIGUOUS
- INFLAMMATORY
- PREVIOUSLY USED

4. QUESTION STRUCTURE/FORMAT

- OPEN VS, CLOSED
- NUMBER OF CATEGORIES
- OPINIONS - ODD OR EVEN?
- COMPARABILITY OF CATEGORIES WITH TRADITIONAL GOVERNMENT RESPONSE CATEGORIES

5. QUESTIONNAIRE LENGTH

- BURDEN
- NONRESPONSE
- VALIDITY

6. QUESTIONNAIRE LAYOUT

- SELF-CODING, IDENTIFIABLE
- CLEAR SKIP PATTERNS
- RESPONSE SHEETS

7. QUESTIONNAIRE INSTRUCTIONS

- INSTRUCTIONS/PROCEDURES
- PROBES

8. OPPORTUNITIES FOR INTERVIEWER COMMENT ON VALIDITY OF RESPONSE

9. RELIABILITY ANALYSIS (PSYCHOLOGICAL-TYPE QUESTIONNAIRES)

- ITEM-ITEM CORRELATIONS
- ITEM-TOTAL CORRELATIONS
- SPLIT-HALF RELIABILITY
- FACTOR ANALYSIS TO VERIFY GROUPING
- INTER-RATER RELIABILITY
- TIME-RELIABILITY

10. VALIDITY ANALYSIS

- CONSTRUCT VALIDITY
- COMPARISON TO “STANDARD”
- WHO PROVIDED ANSWERS

11. RELATE TO ANALYSIS PLAN

- ALL REQUIRED DATA ARE AVAILABLE
- NO UNUSED DATA

12. REVIEW

- BY INDEPENDENT PANEL (SUBSTANTIVE)
- BY DATA PROCESSING PERSONNEL
- BY CODING/EDITING PERSONNEL

C. DEVELOPMENT OF FIELD PROCEDURES

1. TREATMENT OF NONRESPONSE

- ENDORSEMENTS
- CONTACT BY HIGH-LEVEL PERSONNEL
- CALL-BACKS
- INCENTIVES FOR INTERVIEW TEAM
- DYNAMIC REVIEW OF RESPONSE RATES

2. IN-PLACE INTERVIEWERS OR TRAVELING TEAM

- INPLACE: LESS EXPENSIVE; LESS CONTROL
- TRAVELING TEAM: MORE EXPENSIVE; NOT ALWAYS SKILLED IN INTERVIEWING

D. PRETESTING AND PILOT TESTING

1. PRETESTING

CHECK FOR:

- AVAILABILITY OF INFORMATION
- UNDERSTANDABILITY OF QUESTIONS
- CORRECT RESPONSE CATEGORIES
- LENGTH OF TIME TO ADMINISTER
- SMOOTHNESS OF FLOW

2. PILOT TEST

TEST OF:

- OPERATIONAL PROCEDURES (SAMPLE ASSIGNMENTS, TRANSMITTAL, LOGGING, CODING, EDITING)
- RESPONSE CATEGORIES
- RELIABILITY/VALIDITY TEST

E. EDITING, CODING, DATA BASE DESIGN AND DEVELOPMENT

1. MANUAL EDIT

CHECK FOR:

- IDENTIFICATION
- RESPONSE
- DATA TYPE (ALPHA OR NUMERIC)
- RANGE CHECKS
- REASONABLENESS
- CODE OPEN-ENDED QUESTIONS

2. CODING

- SELF-CODING FORM
- VERIFICATION
- MACHINE EDIT (CONSISTENCY CHECKS, RANGE CHECKS)

3. DATA BASE DESIGN AND DEVELOPMENT

- UPDATE FILE
- AUDIT TRAIL
- NONRESPONSE INDICATORS
- IMPUTATION (NOT RECOMMENDED)
- SAMPLE WEIGHTS
- MERGING WITH OTHER DATA
- HIERARCHICAL FILE (DISTRICTS, SCHOOLS)
- DATA DICTIONARY, CODE BOOK, TABLE KEYS (UNIQUE IDENTIFIERS) AND RELATIONSHIPS
- DOCUMENTATION

III. PREPARATION OF OMB CLEARANCE FORMS; CONFIDENTIALITY AND PRIVACY ISSUES

1. OMB CLEARANCE

- COLLECTION OF DATA FROM MORE THAN 9 RESPONDENTS
- GRANTS EXEMPTED
- TIME DELAY – 4 MONTHS
- PROCESS
- PROJECT OFFICER → AGENCY CLEARANCE → OMB CLEARANCE
- LEGAL CLEARANCES
- PRIOR CONTACT HELPFUL
- CLEARANCES FOR PRETEST INSTRUMENT AND FINAL INSTRUMENT, IF REVISED

ITEMS REQUESTED IN OMB CLEARANCE REQUEST PACKAGE

A. STANDARD FORM 83 A (2 COPIES)

B. CERTIFICATION BY AUTHORIZED OFFICIALS SUBMITTING REQUEST (3 COPIES OF THIS, AND THE FOLLOWING ITEMS)

C. SUMMARY OF SUPPORTING STATEMENT

D. SUPPORTING STATEMENT

1. JUSTIFICATION

- NATURE OF REQUIREMENT FOR DATA
- INTENDED USE OF DATA
- AVAILABILITY OF SIMILAR EXISTING DATA

2. DESCRIPTION OF SURVEY PLAN

- DESCRIPTION OF RESPONDENT UNIVERSE
- SURVEY DESIGN AND SAMPLING PLAN
- NAME OF AGENCY OR CONSULTING STATISTICIAN
- NAME AND ROLE OF CONTRACTOR, AND PROVISIONS FOR PROTECTING AND DISPOSING OF DATA FORMS

3. TABULATION AND PUBLICATION PLANS

- PLANS FOR PUBLICATION (TIME, TYPE, AND CONTENT)
- SUMMARY OF TABULATION PLANS

4. TIME SCHEDULE FOR DATA COLLECTION AND PUBLICATION
 - PROJECT SCHEDULE
 - TIME BETWEEN COMPLETION OF DATA COLLECTION AND ISSUANCE OF FINAL PUBLISHED RESULTS
5. CONSULTATIONS OUTSIDE THE AGENCY
 - NAMES OF OUTSIDE CONSULTANTS, SUMMARY OF PROBLEMS
 - DATA AVAILABILITY LEARNED FROM CONSULTATIONS
 - CONSULTATION WITH STATE AND LOCAL GOVERNMENT OFFICIALS
6. ESTIMATION OF RESPONDENT REPORTING BURDEN
 - BASIS FOR ESTIMATE OF RESPONDENT BURDEN
 - TIME FOR COLLECTION AND COMPILATION OF DATA
 - RANGE OF ESTIMATED BURDEN
7. SENSITIVE QUESTIONS
8. ESTIMATE OF COST TO FEDERAL GOVERNMENT
- E. RELATED BACKGROUND AND DOCUMENTS
- F. COPIES OF INSTRUMENTS AND FORMS (INCLUDING INSTRUCTIONS)

2. HINTS

- CAREFUL DEVELOPMENT OF SURVEY INSTRUMENTS
- CAREFUL DEVELOPMENT OF SURVEY DESIGN
- METICULOUS ATTENTION TO "INSTRUCTIONS FOR REQUESTING OMB APPROVAL UNDER THE FEDERAL REPORTS ACT"
- CONSIDERATION OF TERMS OF RELEVANT LEGISLATION, SUCH AS PRIVACY ACT OF 1974, FREEDOM OF INFORMATION ACT, AGENCY REGULATIONS
- COORDINATION WITH OMB CLEARANCE STAFF AND AGENCY CLEARANCE STAFF, BEFORE SUBMISSION OF CLEARANCE REQUEST FORM
- JUSTIFY THE STUDY
- JUSTIFY EACH DATA ELEMENT USED

PRIVACY ACT OF 1974

- RESTRICTIONS ON RELEASE OF INFORMATION
- NO SYSTEM OF RECORDS BASED ON SSN (EXCEPT FOR GRANTS)
- GUARANTEES OF CONFIDENTIALITY -- GET WORDING FROM LEGAL COUNSEL

IV. LONGITUDINAL SURVEYS

1. REASON FOR LONGITUDINAL SURVEYS

- DIFFERENCE BETWEEN A LONGITUDINAL SURVEY AND A SERIES OF CROSS-SECTIONAL SURVEYS
- PURPOSE OF LONGITUDINAL SURVEY
 - MEASUREMENT ON THE SAME INDIVIDUAL
 - PRECISE MEASURE OF CHANGE (CORRELATED OBSERVATIONS)
 - PRECISE ESTIMATES OF RELATIONSHIPS (DISAGGREGATED)
 - ANALYSIS ON VARIABLES THAT CAN BE DEFINED ONLY BY HAVING MEASUREMENTS ON THE SAME INDIVIDUAL

2. PRECISION IMPROVEMENT FROM LONGITUDINAL SURVEYS

CORRELATION:

$$\rho(X_{t1}, X_{t2})$$

$$\text{var}(X_{t2} - X_{t1}) = \text{var}(X_{t2}) + \text{var}(X_{t1}) - 2\rho\sqrt{\text{var}(X_{t2})\text{var}(X_{t1})}$$

IF INDEPENDENT SAMPLES:

$$\text{var}(\bar{X}_{t2} - \bar{X}_{t1}) = \frac{\text{var}(X_{t2})}{n} - \frac{\text{var}(X_{t1})}{n}$$

IF CORRELATED:

$$\text{var}(\bar{X}_{t2} - \bar{X}_{t1}) = \frac{\text{var}(X_{t2})}{n} + \frac{\text{var}(X_{t1})}{n} - \frac{2\text{cov}(X_{t2}, X_{t1})}{n}$$

SO, VARIANCE OF CHANGES IS REDUCED.

BUT, VARIANCE OF MEAN LEVEL (OR TREND) IS LARGER, SINCE

$$\text{var}\frac{\bar{X}_{t1} + \bar{X}_{t2}}{2} = \frac{\text{var}(X_{t1})}{4n} + \frac{\text{var}(X_{t2})}{4n} + \frac{\text{cov}(X_{t1}, X_{t2})}{2n}$$

3. ALLOCATION PROBLEM (PANEL SURVEY)

Q: HOW MANY INDIVIDUALS TO REPLACE IN SECOND SURVEY?

MAXIMUM PRECISION FOR ESTIMATING CHANGES/TRENDS: NONE

MAXIMUM PRECISION FOR ESTIMATING LEVEL: ALL

“BEST” SOLUTION MAY FALL SOMEWHERE IN BETWEEN, E.G., REPLACE 1/3 OF PANEL EACH TIME

4. BAYESIAN ESTIMATION

PRIOR DISTRIBUTION OF THE PARAMETER: $f(\mu) \sim N(\mu^*, \sigma^2)$

SAMPLING DISTRIBUTION: $f(X | \mu)$

POSTERIOR DISTRIBUTION OF THE PARAMETER:

$$f(\mu | X_1, X_2, \dots, X_n) \sim N(\mu', \sigma'^2)$$

$$f(\mu | \underline{X}) = \frac{f(\mu)f(\underline{X} | \mu)}{\int_{\mu} f(\mu)f(\underline{X} | \mu)}$$

BAYES ESTIMATE: MEAN OF POSTERIOR DISTRIBUTION

5. TIME SERIES DESIGN

- MEASUREMENT OF POLLUTION LEVELS
- SPACING OF MEASUREMENTS
- TIME SERIES MODELS (ARIMA, BOX-JENKINS), INTERVENTION ANALYSIS
- OPTIMAL DESIGN OF CONTROL VARIABLE
- SPECTRAL ANALYSIS OF RESIDUALS

6. PROBLEMS IN LONGITUDINAL SURVEYS

- NONRESPONSE – CAN DESTROY THE VALUE OF OBSERVATIONS AT SEVERAL TIME PERIODS
- NONRESPONSE INCREASES IN TIME (LOCATION PROBLEMS)
- SAMPLE UNITS MAY CHANGE (E.G., SCHOOL DISTRICTS, CONGRESSIONAL DISTRICTS)

V. SAMPLE FRAME PROBLEMS

- INCOMPLETE
- DUPLICATION
- EXTRANEIOUS ELEMENTS IN THE FRAME
- SAMPLING FROM MULTIPLE OVERLAPPING FRAMES

VI. SAMPLING FOR RARE ELEMENTS

- SCREENING
- CANVASS PRIMARY SAMPLING UNITS COMPLETELY
- NETWORK SAMPLING
- “SNOWBALL” SAMPLING OR REFERRAL SAMPLING (NONPROBABILITY)

VII. TREATMENT OF NONRESPONSE

A. DESIGN FOR NONRESPONSE

- NONRESPONSE BIAS
 - NONRESPONDENTS MAY DIFFER FROM RESPONDENTS WITH RESPECT TO VARIABLES OF INTEREST, INTRODUCING A BIAS.
- SOURCES OF NONRESPONSE
 - NOT AT HOME
 - REFUSAL
 - ILLNESS / INCAPACITY
 - NOT LOCATED
 - LOST INTERVIEW FORMS
- DESIGN PROCEDURES TO CONTROL FOR NONRESPONSE
 - CALL-BACKS
 - NONRESPONDENT SUBSAMPLING (E.G., TELEPHONE CALLS TO A SUBSAMPLE OF MAIL NONRESPONDENTS) (USE TWO-PHASE SAMPLING TO ADJUST RESULTS)
 - POLITZ-SIMMONS' PROCEDURE (ESTIMATE PROBABILITY OF BEING AT HOME)
 - SUBSTITUTION: GENERALLY NOT RECOMMENDED (SUBSTITUTES RESEMBLE THE RESPONDENTS, NOT THE NONRESPONDENTS.)
 - OVERSAMPLE TO COMPENSATE FOR LOSS OF SAMPLE (AND ASSOCIATED LOSS IN PRECISION)

B. ANALYSIS OF NONRESPONSE

- TYPES OF NONRESPONSE:
 - MISSING OBSERVATIONS (OBSERVATION NONRESPONSE)
 - MISSING VALUES (ITEM NONRESPONSE)
- FOR MISSING OBSERVATIONS:
 - IF NONRESPONDENT SUBSAMPLING WAS USED, USE TWO-PHASE ESTIMATION PROCEDURE TO OBTAIN UNBIASED ESTIMATE
 - FOR HARD-CORE NONRESPONSE, DETERMINE NONRESPONSE STRATA, RESPONSE RATES, MODIFY WEIGHTS
- FOR ITEM NONRESPONSE:
 - DELETE INCOMPLETE ITEMS (NOT SATISFACTORY, UNLESS MOST OF ITEMS ARE MISSING)
 - SUBSTITUTE MEAN VALUES (NOT RECOMMENDED)
 - REGRESSION METHODS (ESTIMATE MISSING VALUE BASED ON OBSERVED RELATIONSHIPS TO USUALLY-AVAILABLE ITEMS)
 - HOT-DECK PROCEDURES: REPLACE THE MISSING VALUE BY THE VALUE FROM A PREVIOUS RECORD (DISTRIBUTIONAL PROPERTIES PRESERVED)

NOTE: IN AN ANALYTICAL MODEL, COEFFICIENT ESTIMATES ARE UNBIASED IF NONRESPONSE EVENT IS INDEPENDENT OF MODEL ERROR TERM,

VIII. NONSAMPLING ERRORS

- TRUE VALUE
- OBSERVATIONAL ERROR
- REPORTING ERROR: RANDOM, SYSTEMATIC
- INTERVIEWER BIAS
- RELATIVE MAGNITUDE OF SAMPLING AND NONSAMPLING ERROR
- DIFFERENT MODELS FOR ESTIMATING AND TESTING FOR MEASUREMENT ERRORS

IX. RANDOMIZED RESPONSE

USED FOR SENSITIVE QUESTIONS. THE INTERVIEWER DOES NOT KNOW IF THE RESPONDER IS ANSWERING THE SENSITIVE QUESTION OR ANSWERING AN INNOCUOUS QUESTION. THE QUESTION IS CHOSEN AT RANDOM.

X. RANDOM DIGIT DIALING

- UNRESTRICTED
- PLUS 1 METHOD
- WORKING BANK METHOD
- PPS METHODS

XI. MAJOR NATIONAL AND INTERNATIONAL SURVEYS

A. CURRENT POPULATION SURVEY (CPS)

THE CURRENT POPULATION SURVEY (CPS) IS A MONTHLY SURVEY OF HOUSEHOLDS CONDUCTED BY THE BUREAU OF CENSUS FOR THE BUREAU OF LABOR STATISTICS. IT PROVIDES A COMPREHENSIVE BODY OF DATA ON THE LABOR FORCE, EMPLOYMENT, UNEMPLOYMENT, PERSONS NOT IN THE LABOR FORCE, HOUSEHOLD CHARACTERISTICS.

THE CPS USES A SAMPLE OF 60,000 HOUSEHOLDS; DATA ARE COLLECTED BY PERSONAL AND TELEPHONE INTERVIEWS. BASIC LABOR FORCE DATA ARE GATHERED MONTHLY; DATA ON SPECIAL TOPICS ARE GATHERED IN PERIODIC SUPPLEMENTS.

CPS USES THE ROTATING PANEL METHOD OF INTERVIEWS. IT IS A MULTISTAGE CLUSTER SAMPLE.

B. NATIONAL HEALTH INTERVIEW SURVEY (NHIS)

OBTAINS COMPREHENSIVE STATISTICS ON DISEASES, INJURIES AND IMPAIRMENTS IN THE GENERAL POPULATION, DATA OBTAINED FROM HOUSEHOLD INTERVIEWS

THE NATIONAL HEALTH INTERVIEW SURVEY (NHIS) IS THE PRINCIPAL SOURCE OF INFORMATION ON THE HEALTH OF THE CIVILIAN NONINSTITUTIONALIZED POPULATION OF THE UNITED STATES AND IS ONE OF THE MAJOR DATA COLLECTION PROGRAMS OF THE NATIONAL CENTER FOR HEALTH STATISTICS (NCHS). THE NATIONAL HEALTH SURVEY ACT OF 1956 PROVIDED FOR A CONTINUING SURVEY AND SPECIAL STUDIES TO SECURE ACCURATE AND CURRENT STATISTICAL INFORMATION ON THE AMOUNT, DISTRIBUTION, AND EFFECTS OF ILLNESS AND DISABILITY IN THE UNITED STATES AND THE SERVICES RENDERED FOR OR BECAUSE OF SUCH CONDITIONS. THE SURVEY REFERRED TO IN THE ACT, NOW CALLED THE NATIONAL HEALTH INTERVIEW SURVEY, WAS INITIATED IN JULY 1957. SINCE 1960, THE SURVEY HAS BEEN CONDUCTED BY NCHS, WHICH WAS FORMED WHEN THE NATIONAL HEALTH SURVEY AND THE NATIONAL VITAL STATISTICS DIVISION WERE COMBINED.

WHILE THE NHIS HAS BEEN CONDUCTED CONTINUOUSLY SINCE 1957, THE CONTENT OF THE SURVEY HAS BEEN UPDATED ABOUT EVERY 10-15 YEARS. IN 1996 A SUBSTANTIALLY REVISED NHIS CONTENT BEGAN FIELD TESTING. THIS NEW QUESTIONNAIRE, DESCRIBED IN DETAIL BELOW, BEGAN IN 1997 AND IMPROVES THE ABILITY OF THE NHIS TO PROVIDE IMPORTANT HEALTH INFORMATION.

SAMPLE DESIGN. THE NATIONAL HEALTH INTERVIEW SURVEY IS A CROSS-SECTIONAL HOUSEHOLD INTERVIEW SURVEY. SAMPLING AND INTERVIEWING ARE CONTINUOUS THROUGHOUT EACH YEAR. THE SAMPLING PLAN FOLLOWS A MULTISTAGE AREA PROBABILITY DESIGN THAT PERMITS THE REPRESENTATIVE SAMPLING OF HOUSEHOLDS. THE SAMPLING PLAN WAS REDESIGNED IN 1995. INFORMATION ABOUT THE SAMPLING PLAN GIVEN HERE

COVERS THE DESIGN YEARS OF 1995-2004. THE FIRST STAGE CONSISTS OF A SAMPLE OF 358 PRIMARY SAMPLING UNITS (PSU'S) DRAWN FROM APPROXIMATELY 1,900 GEOGRAPHICALLY DEFINED PSU'S THAT COVER THE 50 STATES AND THE DISTRICT OF COLUMBIA. A PSU CONSISTS OF A COUNTY, A SMALL GROUP OF CONTIGUOUS COUNTIES, OR A METROPOLITAN STATISTICAL AREA.

WITHIN A PSU, TWO TYPES OF SECOND-STAGE UNITS ARE USED: AREA SEGMENTS AND PERMIT AREA SEGMENTS. AREA SEGMENTS ARE DEFINED GEOGRAPHICALLY AND CONTAIN AN EXPECTED EIGHT OR TWELVE ADDRESSES. PERMIT AREA SEGMENTS COVER GEOGRAPHICAL AREAS CONTAINING HOUSING UNITS BUILT AFTER THE 1990 CENSUS. THE PERMIT AREA SEGMENTS ARE DEFINED USING UPDATED LISTS OF BUILDING PERMITS ISSUED IN THE PSU SINCE 1990 AND CONTAIN AN EXPECTED FOUR ADDRESSES. WITHIN EACH SEGMENT ALL OCCUPIED HOUSEHOLDS AT THE SAMPLE ADDRESSES ARE TARGETED FOR INTERVIEW.

C. DEMOGRAPHIC AND HEALTH SURVEYS (DHS)

PROVIDES DATA FOR A WIDE RANGE OF MONITORING AND IMPACT EVALUATION INDICATORS IN THE AREAS OF POPULATION, HEALTH, AND NUTRITION.

DEMOGRAPHIC AND HEALTH SURVEYS (DHS) ARE NATIONALLY-REPRESENTATIVE HOUSEHOLD SURVEYS WITH LARGE SAMPLE SIZES (USUALLY BETWEEN 5,000 AND 30,000 HOUSEHOLDS). DHS SURVEYS PROVIDE DATA FOR A WIDE RANGE OF MONITORING AND IMPACT EVALUATION INDICATORS IN THE AREAS OF POPULATION, HEALTH, AND NUTRITION.

TYPICALLY, DHS SURVEYS ARE CONDUCTED EVERY 5 YEARS, TO ALLOW COMPARISONS OVER TIME. INTERIM SURVEYS FOCUS ON THE COLLECTION OF INFORMATION ON KEY PERFORMANCE MONITORING INDICATORS BUT MAY NOT INCLUDE DATA FOR ALL IMPACT EVALUATION MEASURES (SUCH AS MORTALITY RATES). THESE SURVEYS ARE CONDUCTED BETWEEN ROUNDS OF DHS SURVEYS AND HAVE SHORTER QUESTIONNAIRES THAN DHS SURVEYS. ALTHOUGH NATIONALLY REPRESENTATIVE, THESE SURVEYS HAVE SMALLER SAMPLES THAN DHS SURVEYS (2,000–3,000 HOUSEHOLDS).

SURVEY TYPES

DHS SUPPORTS A RANGE OF DATA COLLECTION OPTIONS THAT CAN BE TAILORED TO FIT SPECIFIC MONITORING AND EVALUATION NEEDS OF HOST COUNTRIES. LEARN MORE ABOUT THE TYPES OF SURVEYS, SECONDARY DATA ANALYSIS, AND SPECIALIZED STUDIES THAT MEASURE DHS PERFORMS.

DEMOGRAPHIC AND HEALTH SURVEYS (DHS): PROVIDES DATA FOR A WIDE RANGE OF MONITORING AND IMPACT EVALUATION INDICATORS IN THE AREAS OF POPULATION, HEALTH, AND NUTRITION.

AIDS INDICATOR SURVEYS (AIS): PROVIDE COUNTRIES WITH A STANDARDIZED TOOL TO OBTAIN INDICATORS FOR THE EFFECTIVE MONITORING OF NATIONAL HIV/AIDS PROGRAMS.

SERVICE PROVISION ASSESSMENT (SPA) SURVEYS: PROVIDES INFORMATION ABOUT THE CHARACTERISTICS OF HEALTH AND FAMILY PLANNING SERVICES AVAILABLE IN A COUNTRY.

KEY INDICATORS SURVEY (KIS): PROVIDES MONITORING AND EVALUATION DATA FOR POPULATION AND HEALTH ACTIVITIES IN SMALL AREAS—REGIONS, DISTRICTS, CATCHMENT AREAS—THAT MAY BE TARGETED BY AN INDIVIDUAL PROJECT, ALTHOUGH THEY CAN BE USED IN NATIONALLY REPRESENTATIVE SURVEYS AS WELL.