Fundamentals of Program Evaluation Course 380.611

Monitoring Outputs and Outcomes

And Introduction to Study Design

Topics to cover

- RHIS as a tool to monitor outputs and outcomes
- Use of surveys to monitor outcomes
- Break
- Threats to validity
- Intro to “impact”

Purpose of MIS at local level

Management to track:

- Inputs (finances, personnel)
- Process (description of activities)
- Outputs
  - # activities conducted
  - Measures of access (optional)
  - Results (# products distributed, service utilization)

Inputs: Finance, Personnel

- Essential data for management
- Less important for M&E
- Covered in management course, not Evaluation Course
  - Dissenting comment by Duff Gillespie

Process

- Records of training courses
- Records on activities implemented (type, purpose, # participants)
- Evaluation of quality of services

Note: database for determining:

- Was project implemented as planned
- How well was it implemented

Outputs – Two main types

Number of activities conducted

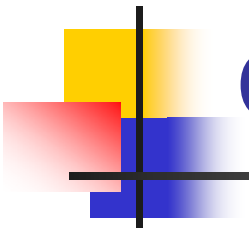
- # persons trained
- # pamphlets, posters produced/distributed

Results:

- # products distributed
- Volume of service utilization
- Profile of clients

Note: all data are program - level

Typical “results” from tracking outputs (service utilization)



Steps in establishing an MIS for service statistics

- Determine indicators to track
- Design forms to collect the data
- Assign responsibility for data collection
- Establish a system of remitting the data back to the central offices
- Train staff in use of the forms
- Ensure timely submission of data

Figure 1. Overlap in Terminology used in Programme Evaluation

Developed by Jane T. Bertrand

What do we measure?

K-A-P, groups affected, barriers, channels, access	Funding, personnel, equipment, supplies, infrastructure	Activities completed, items produced	Quality Assess (Reach)	Distribution/ sales data, service utilisation	Change in:			Change in:
					Psycho-social (knowledge, attitude, self-efficacy, etc.)	Behaviour, practice	Health status	Outputs or outcomes attributable to the intervention

What is the terminology?

Formative	Process evaluation	Summative evaluation

Monitoring	Monitoring outputs		Monitoring outcomes			Impact Assessment
Inputs	(functional) outputs	Service Outputs	Outputs	Initial	Intermediate	Long-term	• RCTs, exp. design • Quasi exp. design • Cross-sectional survey with advanced statistical analysis

Performance monitoring

Monitoring and evaluation (M & E)

Programme evaluation

Monitoring outcomes

- Program-based measurement
  - Available through RHIS
  - Ex: % of TB patients enrolled who complete treatment
- Population-based measurement
  - Not available through RHIS
  - Ex: % contraceptive use over time

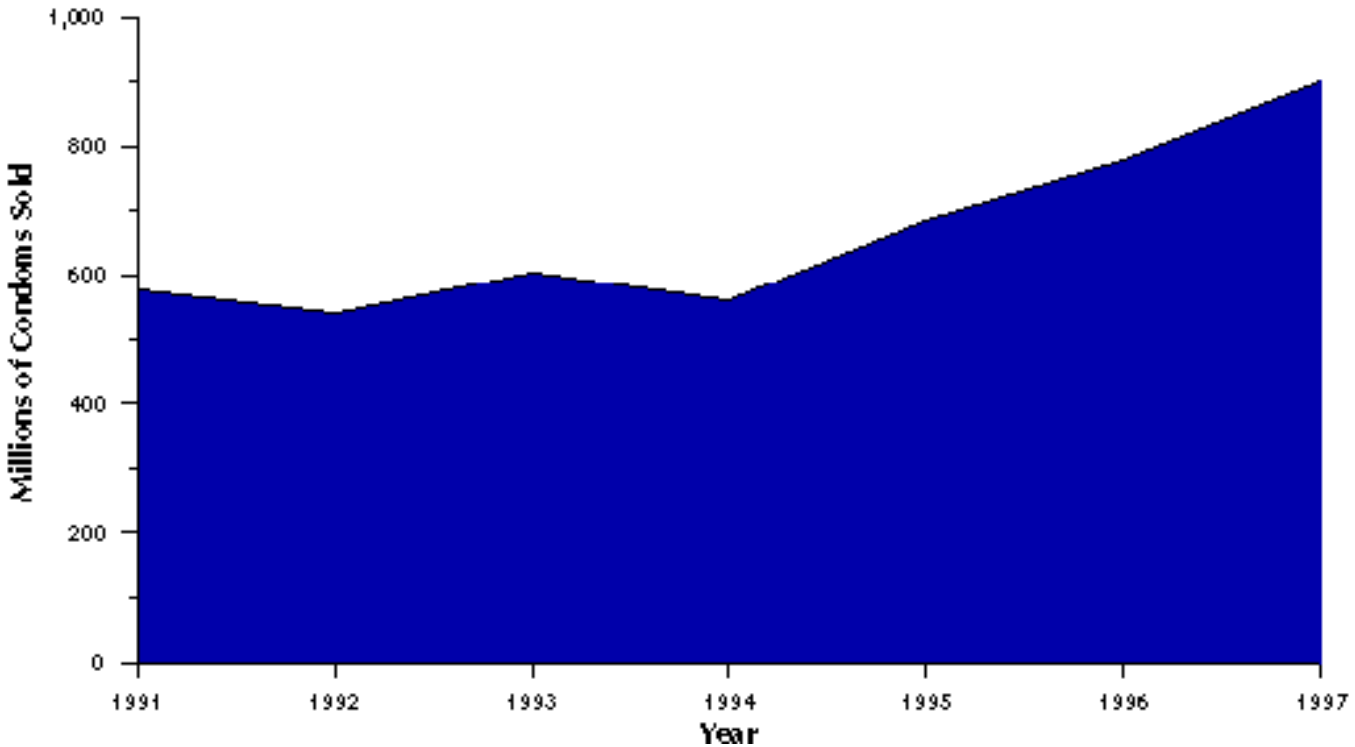
Utility of data from monitoring

- To track levels in a given service over time
- To compare regions/sites on key indicators
- To compare program performance to specified standards

Example of monitoring outputs: # condoms sold

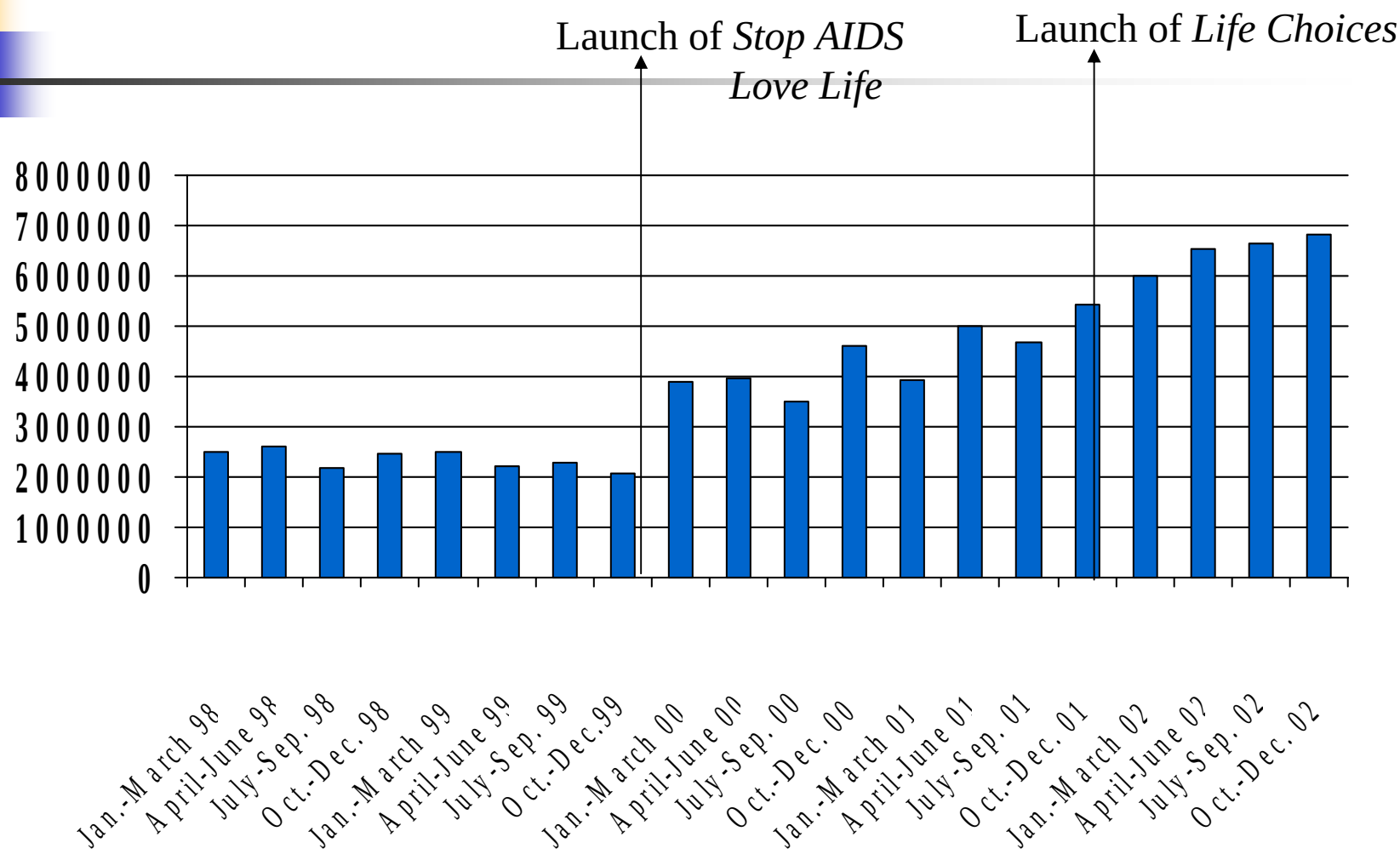
Figure 1. Condoms Sold Through Social Marketing, 1991–1997*

Asia & Near East, Africa, Latin America & Caribbean, Eastern Europe & Central Asia



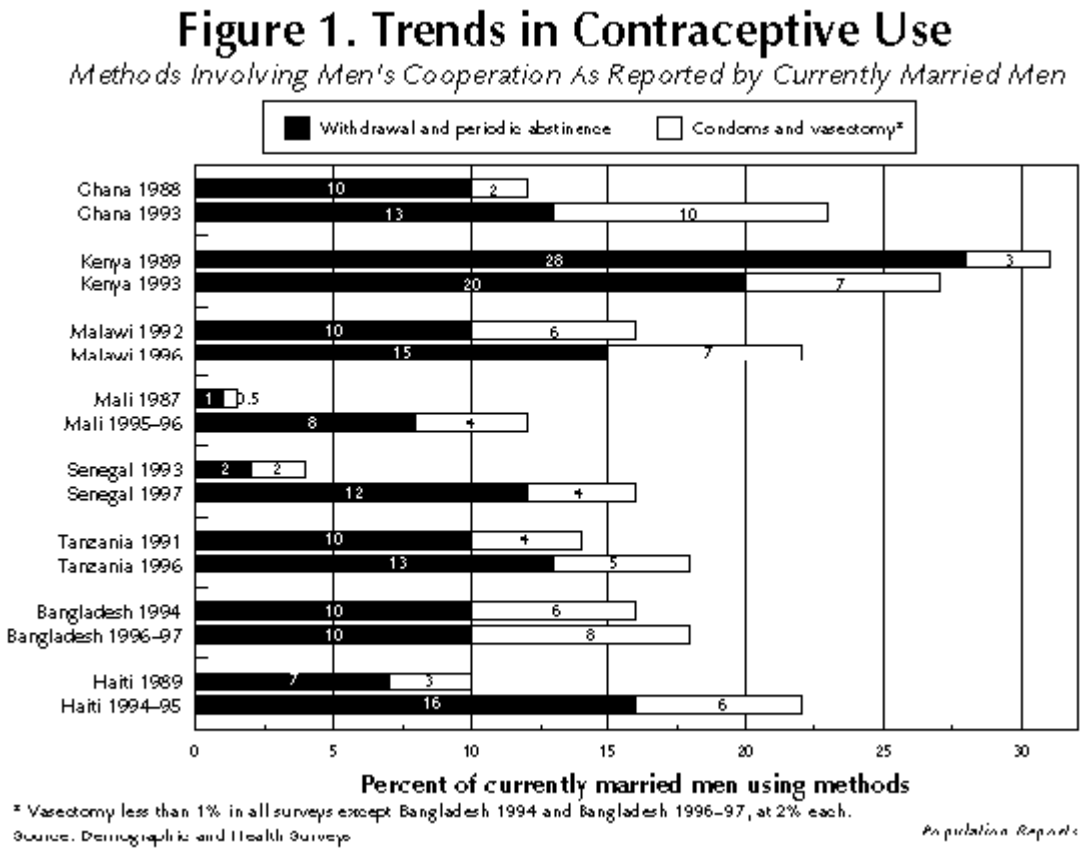
From Gardner, R., Blackburn, R.D., and Upadhyay, U.D. *Closing the Condom Gap*. *Population Reports*, Series H, No. 9. Baltimore, Johns Hopkins University School of Public Health, Population Information Program, April 1999.

Number of condoms distributed, Ghana 1998-2002



Source: Ghana's *Stop AIDS Love Life* Program. Phase 1: Evaluation Report. February 2000 to June 2001

Use of male methods: comparisons by country and across time



Source: Drennan, M. *Reproductive Health: New Perspectives on Men's Participation. Population Reports*, Series J, No. 46. Baltimore, Johns Hopkins School of Public Health, Population Information Program, October 1998.

Advantages of a health MIS

- Routine:
  - Doesn’t require a special study
  - Existing staff collect the info
- Local:
  - Provides info relevant to local program
- Timely:
  - Data are available on a monthly or quarterly basis

Shortcomings of Health MIS

- Data are often of poor quality
- Data collection detracts from clinical service delivery (I.e., takes time)
- Data are often sent to the central level without being used at local level
- Staff at the local level may not feel comfortable “analyzing” the data

…As a result…

- Service statistics often provide unreliable data on health interventions
  - Inflated results
  - Staff disinterested in process
  - One-way flow of info (no local benefit)
- Result: widespread use of DHS-type surveys instead of routine service statistics

So what is "program monitoring?"

Can include tracking activities:

- # training events, # materials produced

Also tracks program-level results:

- Levels of service delivery
- (behavior-related) products distributed

May track outcomes:

- Percent of target population performing behavior

Figure 1. Overlap in Terminology used in Programme Evaluation

Developed by Jane T. Bertrand

What do we measure?

K-A-P, groups affected, barriers, channels, access	Funding, personnel, equipment, supplies, infrastructure	Activities completed, items produced	Quality Assess (Reach)	Distribution/ sales data, service utilisation	Change in:			Change in:
					Psycho-social (knowledge, attitude, self-efficacy, etc.)	Behaviour, practice	Health status	Outputs or outcomes attributable to the intervention

What is the terminology?

Formative	Process evaluation	Summative evaluation

Monitoring	Monitoring outputs		Monitoring outcomes			Impact Assessment
Inputs	(functional) outputs	Service Outputs	Outputs	Initial	Intermediate	Long-term	• RCTs, exp. design • Quasi exp. design • Cross-sectional survey with advanced statistical analysis

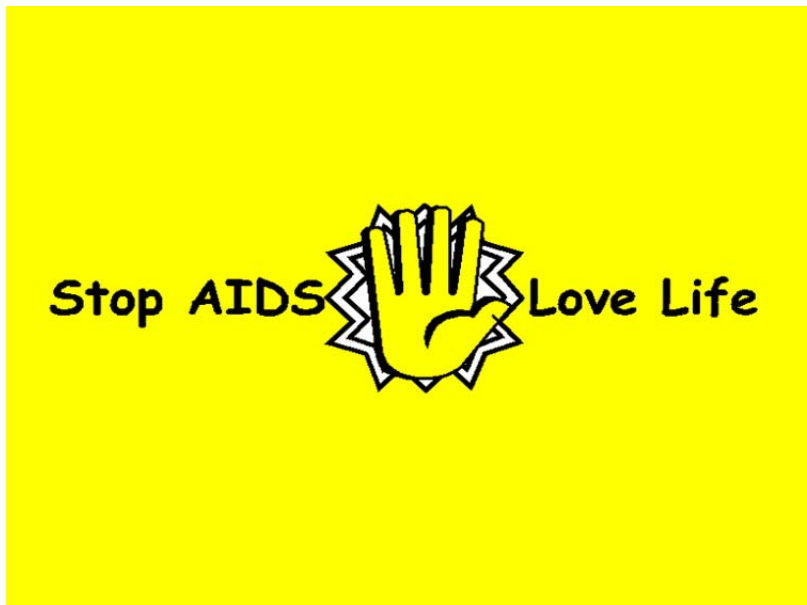
Performance monitoring

Monitoring and evaluation (M & E)

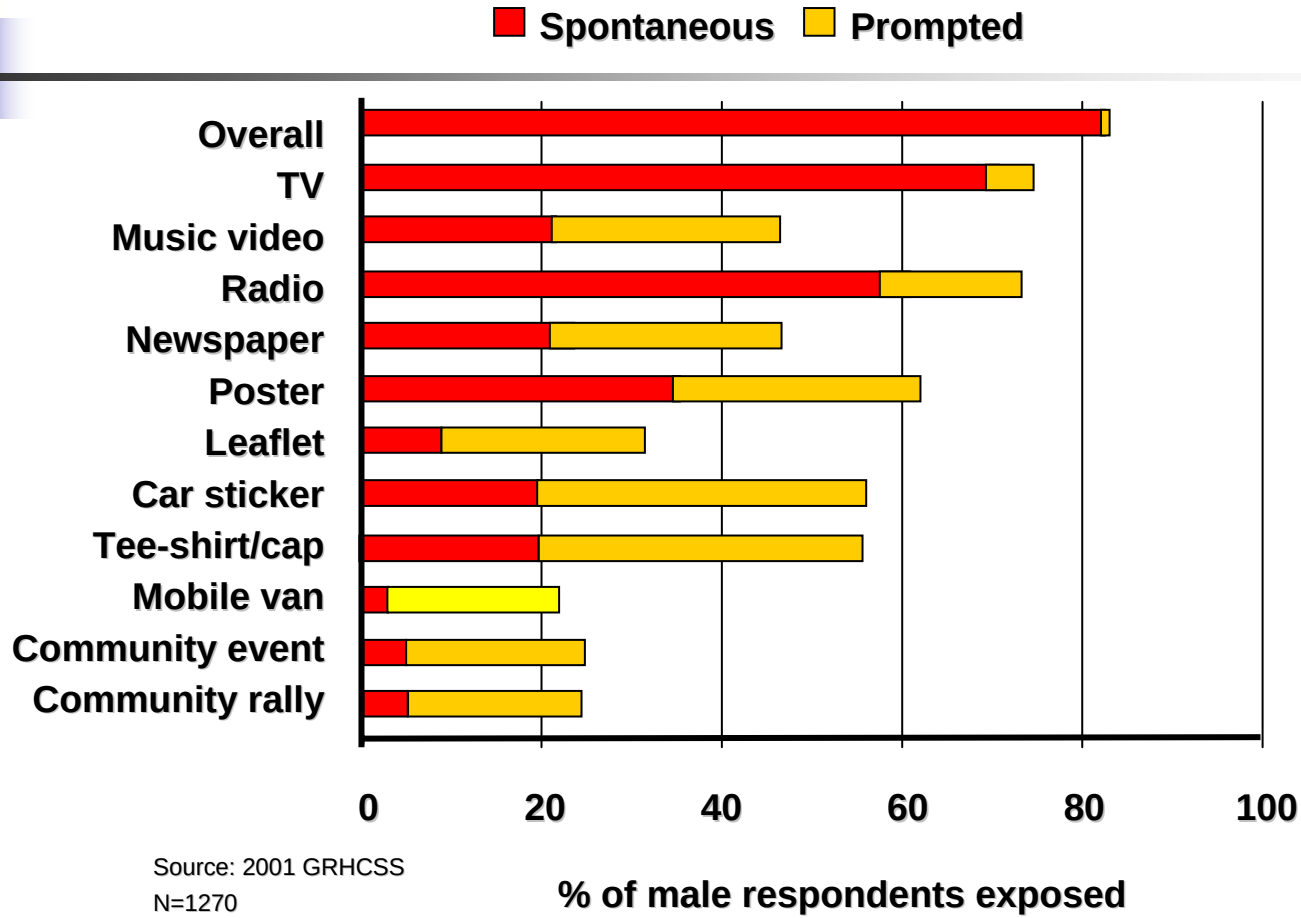
Programme evaluation

Measuring "reach" (using survey data)

- Example from Stop AIDS Love Life



Exposure to the *Stop AIDS Love Life* campaign components, Males, 2001



Source: 2001 GRHCSS
N=1270

Monitoring outcomes:

- Initial outcomes
  - Useful if topic is new, difficult (e.g., FGM)
- Intermediate outcomes
  - Focus of much evaluation
- Long-term outcomes
  - Important but they change slowly; other factors contribute

Initial outcomes

- Knowledge
- Attitudes
- Self-efficacy

Question: why do we care?

(Example: knowledge of HIV/AIDS in the U.S. in mid-1980s)

Intermediate outcomes

Health behavior, practice: examples

- % employees who exercise 3 times/week
- % 15-19 years olds who don't smoke
- % of infants fully immunized at 12 months
- % pregnant women who deliver with skilled birth attendant

Long-term outcomes

- Fertility
  - Total fertility rate
- Morbidity
  - HIV prevalence
  - % diarrheal disease (past month)
- Mortality
  - Infant mortality, maternal mortality
- Quality of life

Source of data (if population-based)

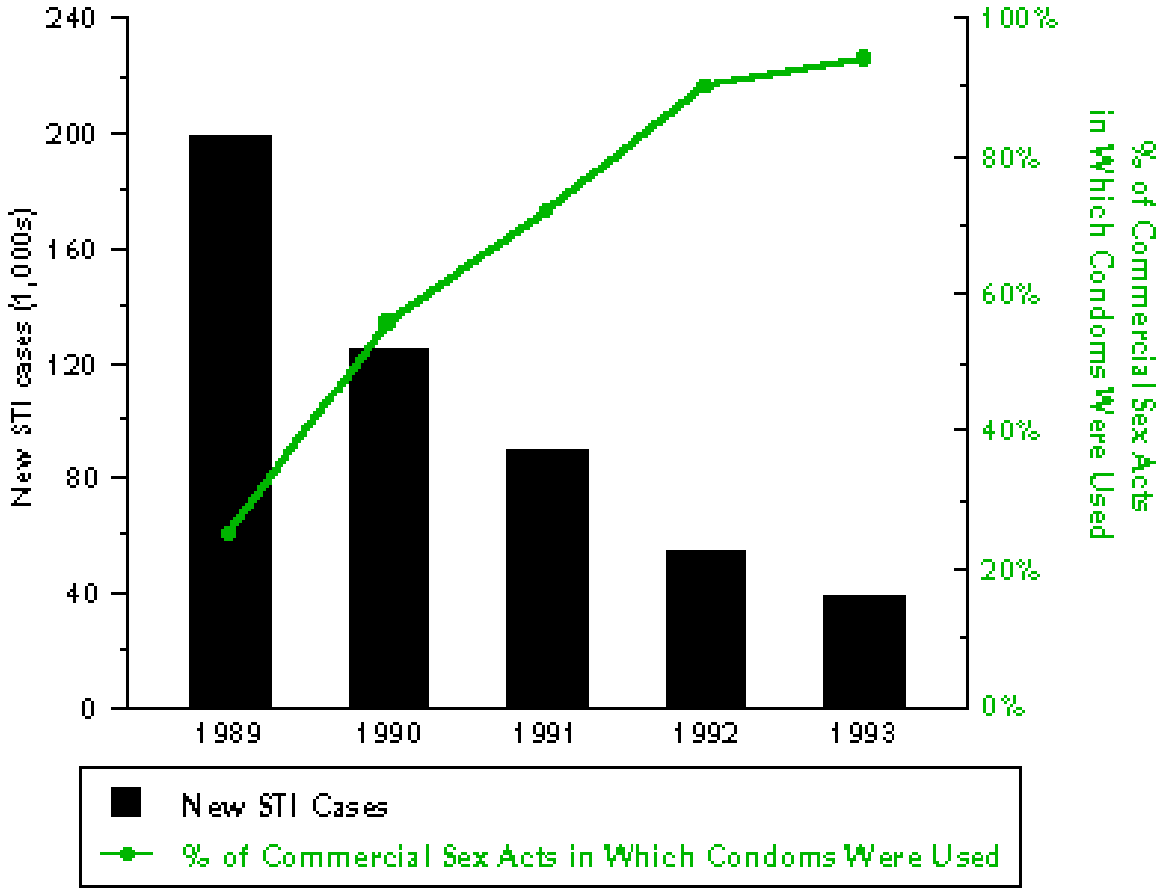
- Representative survey of the general population (or relevant subsection)
- Census (e.g., fertility, mortality rates)
- National or regional surveys (e.g., DHS)

Differentiating Outputs versus Outcomes

- Output: measured at program level:
  - Activities conducted:
    - # brochures printed
  - Service utilization:
    - # client visits, # condoms sold, # HIV tests performed
- Outcomes: program or population level:
  - % of TB patients successfully treated
  - % of youth using condom at last sex

Figure 2. Condom Use and Sexually Transmitted Infections (STIs)

Condom Use Among Sex Workers and New Male STI Cases at Government Clinics, Thailand, 1989–1993



Source: Hannenberg et al. 1994 (227)

Population Reports

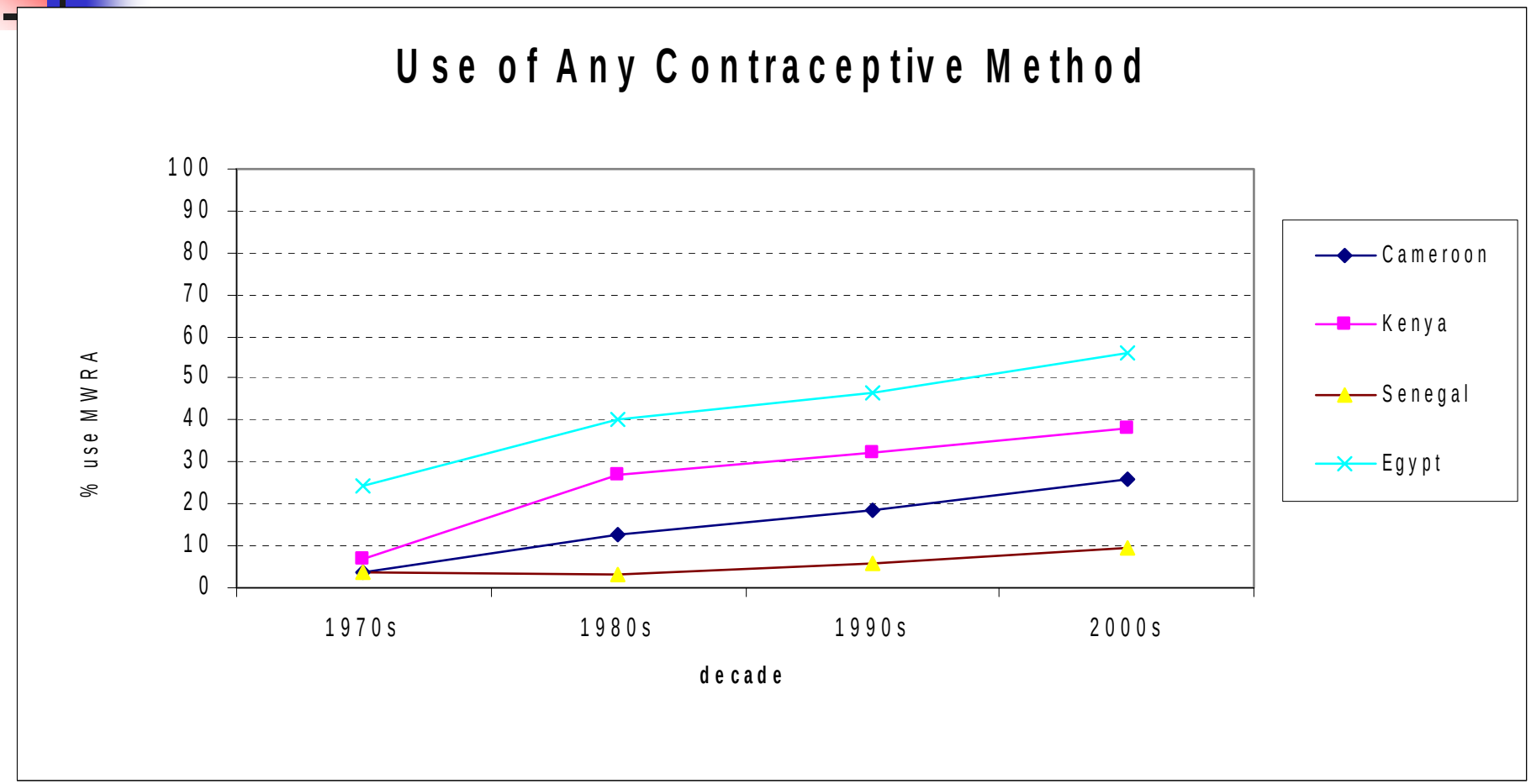
OUTPUTS OR OUTCOMES?

Source: Gardner, R., Blackburn, R.D., and Upadhyay, U.D. *Closing the Condom Gap. Population Reports*, Series H, No. 9. Baltimore, Johns Hopkins University School of Public Health, Population Information Program, April 1999.

Value of monitoring outcomes

- Demonstrates trends at national/regional level
- Useful for policy makers and public health practitioners
- Allows for comparisons: cross-national and over time

Example: tracking contraceptive use over time: 4 countries



Source: Ghana's *Stop AIDS Love Life* Program. Phase 1: Evaluation Report. February 2000 to June 2001

Limitations of monitoring outcomes

- Program-based:
  - What happened to those outside program?
- Population-based:
  - Does not measure causation
  - Is often misunderstood as demonstrating cause and effect
  - Survey data collection is costly, infrequent

Question

- Can we use qualitative methods to measure outcomes?

Question

- When is monitoring outcomes “enough”? (or “appropriate”)

Moving from monitoring change to measuring impact

Donors, program administrator, providers, evaluators:

Want to know:

- Did this program make a difference?
- Did it have an impact?
- Did the program cause the desired/observed change?

To answer this question, need to have:

- Appropriate study design, and/or
- Appropriate statistical analysis (based on advanced multivariate analysis)

Review of two key measurement concepts

- Reliability: The measure repeatedly gives the same results
- Validity: We are measuring what we think we are measuring

Measurement of validity

- Considers systematic error: a true effect may be biased positively or negatively
- How well measures represent underlying concepts
- Objective behavior measures more likely valid
- More difficult to assess than reliability

Study design needs to control for threats to validity

- History
- Maturation
- Testing
- Instrumentation
- Regression
- Selection
- Attrition

Threats to internal validity

- History: discrete events external to units
- Maturation: events occurring within subjects over time
- Testing: pretest influences results

Threats to internal validity, continued

- Instrumentation: changes in instrument or raters influences results
- Regression (to the mean): units chosen on extreme scores will naturally regress
- Selection: compositional differences in treatment and comparison groups

Threats to internal validity, continued

- Attrition: differential dropout of units between treatment, control groups

- Selection-maturation interaction: selection differences cause different maturation effects
  - Olympic athletes versus couch potatoes

Threats to internal validity, continued

- Selection-history interaction: selection differences cause different exposures to historical events

- Ex: effect of Magic Johnson's HIV announcement on blacks and whites

Which study designs (next class) control to threats to validity?

- Experimental (randomized control trials)
  - "gold standard" for assessing impact
  - Controls threat to validity
- Quasi-experimental
  - Controls some threats to validity
- Non experimental
  - Do not control threats to validity
- Observational with statistical controls:
  - Controls some threats to validity