Data Analysis for Disease Accounts

Dr Ravi P. Rannan-Eliya

Estimating Expenditure by Disease Training Workshop

WHO, Geneva

August 20-24, 2007

Outline

- Health accounts and databases
- Data analysis for DSA
  - Classifications and mapping tables
  - Matrix manipulation
  - Weighting and over-sampling
- Post-analysis routines and validation

Goals of WHO GFATM Evaluation Project

- To estimate full health accounts
- To classify all disease-relevant expenditures by ICD-10
- To identify HIV/TB/Malaria-specific expenditure

General Health Accounts Databases

Tables

- Tables are the tools used to present the analysis and estimates
- Standard tables are primarily to allow international reporting or comparison
  - Actual tables (or charts) for policy-makers or for local readers will differ
- Primary objective of the SHA ICHA is to cross-classify expenditures by financing agent, function and provider

Databases for Health Accounts

- What is in a table?
  - Every cell in a table is simply the sum of all expenditures whose classes match the relevant year, row and column (financing agent, provider, function, disease, age, etc)
- Emerging best practice for health accounts
  - Shift from manually filling in cells in tables, to compiling databases with required information
  - -> Automated table production

Producing tables from databases

- Analysis phase: Disaggregate all expenditures to single elements defined by single year, source, provider, function, etc

Example:

YEAR	SOURCE	PROVIDER	FUNCTION	AMOUNT
2001	S1.3	P2.4	F1.1	$1020
2001	S1.3	P2.4	F1.2	$2090
2001	S1.4	P1.1	F1.2	$ 85
2001	S1.4	P3.1	F5.2	$ 231

Producing tables from databases

- Reporting phase: Use software (database applications, statistical software) to sum the expenditures and generate each table

Example:

Database

YEAR	SOURCE	PROVIDER	FUNCTION	AMOUNT
2001	S1.3	P2.4	F1.2	$ 40
2001	S1.1	P2.4	F1.2	$2090
2001	S1.3	P1.1	F1.2	$ 45
2001	S1.3	P3.1	F5.2	$ 231
2001	S1.3	P3.1	F5.2	$ 231
2001	S2	P3.1	F5.2	$ 231

Source / Function Table

S / F	F1.1	F1.2	F2	F3
S1.1	$1,000	$500	$3,000	$0
S1.2	$,3090	$200	$2,000	$5,030
S1.3	$ 100	$ 85	$ 2,00	$ 0
S2	$ 200	$ 231	$ 0	$ 34

Producing tables from databases

- Using multiple classifications
  - E.g., Country-specific classification for local use, plus ICHA classification for international reporting
  - Implement using mapping tables/dictionaries

Example:

YEAR	SOURCE	ICHA-HF	FUNCTION	AMOUNT
2001	S1.3	HF1.3	F1.1	$1020
2001	S1.3	HF1.3	F1.2	$2090
2001	S1.4	HF2.1	F1.2	$ 85
2001	S1.4	HF2.1	F5.2	$ 231

Producing tables from databases

- **Advantages**
  - Reduces time required for table construction
  - Ensures internal consistency between tables
  - Facilitates use of parallel classifications
  - Facilitates creation of extended analyses

- **GFATM Evaluation Project**
  - Essential for preparing tables with disease, age and sex breakdowns

Data Analysis for Disease Accounts

Classifications

- Develop classifications early in process
  - Record in tables/spreadsheets
  - Map to standard classifications if necessary
- Classifications
  - Disease (ICD-10, GBD, GFATM)
  - Age, sex
  - Source (ICHA, GFATM)
  - Functions (ICHA, GFATM)
  - Providers (ICHA, GFATM)

Classifications

Code	Disease Category	IMMR codes	ICD-10 codes
D	**Neonatal Causes**		
1	Birth Trauma & Asphyxia	200, 201	P10-P15, P20-P28
2	Low Birth Weight	199	P05-P07
3	Other Neonatal Causes	198, 202,203	P00-P04, P08, P29, P35-P39, P50-P61, P70-P78, P80-P83, P90-96
E	**Nutritional Deficiencies**		
1	Iron Deficiency Anaemia	084a	D50
2	Protein-Energy Malnutrition	091	E40-E46
3	Other Nutritional Deficiencies	092	E50-E56,E90
F	**Malignant Neoplasms**		
1	Breast Cancer	064	C50
2	Lung Cancer	059	C33-C34
3	Cervix Cancer	065	C53
4	Prostate Cancer	069	C61
5	Other Malignant Neoplasms	050-058, 060-063, 066-068, 070-081, 083	C00-C26, C30-C32, C37-C41, C43-C49, C51, C52, C54-C58, C60, C62-C85 C88-C97, D00-D09, D37-D48

Disease analysis

- Intermediate stage
  - Health Accounts estimates of target spending areas for analysis by disease
  - Datasets containing the percentage (%) distribution of expenditures by ICD-10, age and sex for different spending areas
- Final stage
  - Combining the health accounts totals with the information on distribution of expenditures by disease

Cost distribution keys

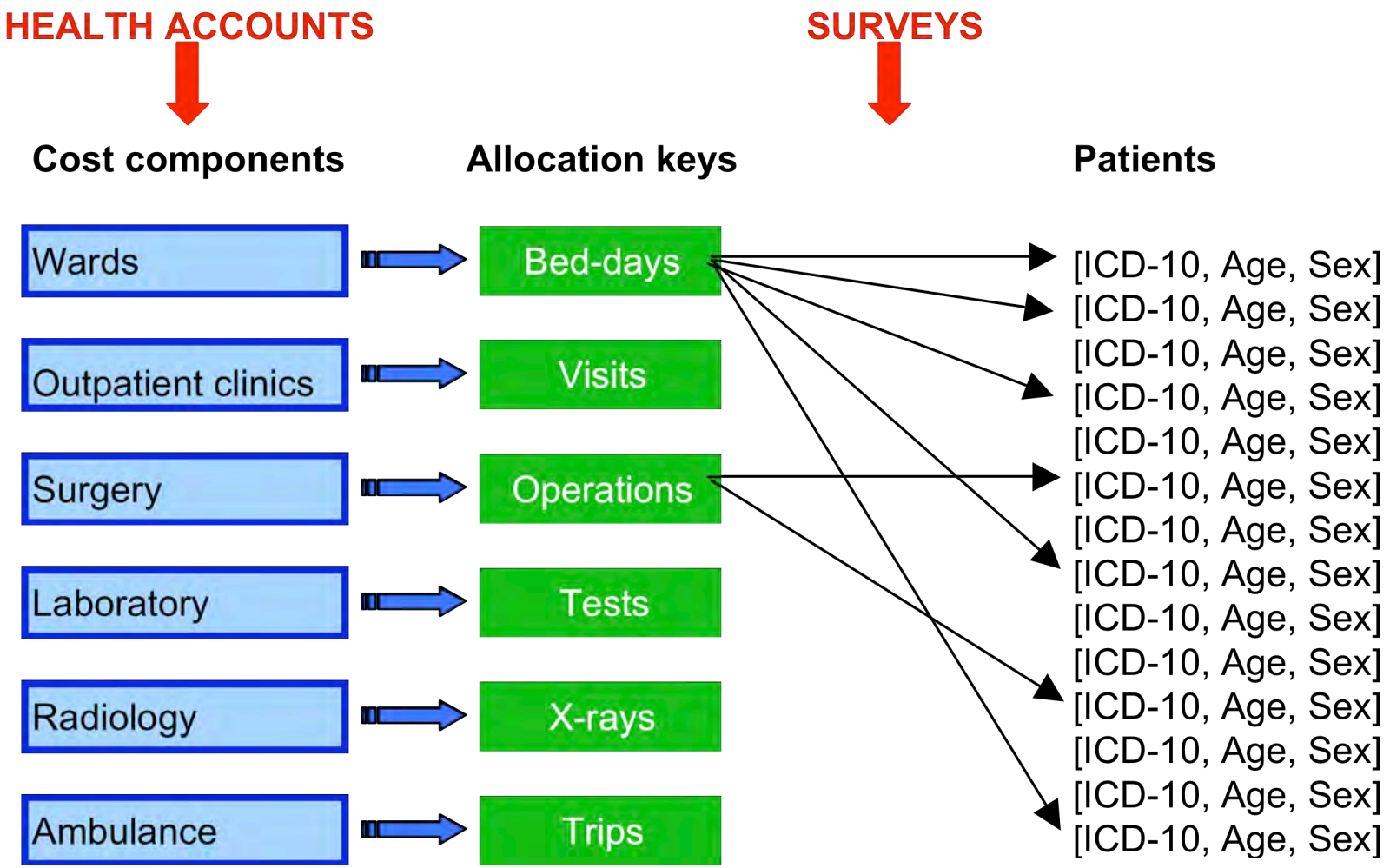
Basic method

$$\text{Disease A expenditures} = \text{Total expenditures} \times \frac{\text{No. of patients with Disease A}}{\text{No. of all patients}}$$

Keys

- Patient outpatient visits, admissions
- Bed-days, operations, lab tests, prescriptions
- Drug costs, staff costs, lab costs
- Doctors' time, nursing staff time, other staff time
- Utilities costs, administrative costs

Inpatient analysis



Outpatient analysis 1

* *Assuming approach using household health survey + ICPC-based patient survey*

- Map conditions/problems in household survey to ICD codes or ICPC RFE codes
- Use patient survey to map ICPC RFE codes to actual ICD diagnoses
  - Save analysis as a dataset containing:
    - ICPC code, ICD-10 code, Age, Sex

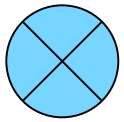
Outpatient analysis 2

Survey code	Illness term	ICD code	%	ICPC code
010	Measles	B05		
011	Chickenpox	B01		
012	Smallpox	B03		
013	Infectious Jaundice	B15	0.80	
013	" "	B16	0.20	
014	Yellow Fever	A95		
019	Fever with other and unspecified skin manifestations			A03
020	Encephalitis	A83		
021	Dengue	A90		
022	Plague	A202		
023	Poliomyelitis	A80		
024	Meningitis	A87	0.40	
024	" "	G00	0.40	
024	" "	A17	0.20	
029	Fever with other and unspecified neurological manife			A03
030	Malaria	B50	0.29	
030	" "	B51	0.71	
031	Malaria, Confirmed	B50	0.29	
031	" "	B51	0.71	

Outpatient analysis 3

Age category	Sex	Problem
3	female	1
2	male	4
3	female	5
1	female	1
2	male	2
2	male	9

Household survey



Patient survey

Sex	Age category	Problem	ICD-10	Fraction
female	1	1	B25	45%
female	1	1	R05	55%
female	2	1	B25	65%
female	2	1	R05	35%
female	3	1	B25	10%
female	3	1	R05	90%

Multiple classifications

Example: HIV

- Disease accounts
  - Code using ICD-10 codes
- GFATM tables
  - Combine HIV plus OIs
  - HIV-TB to be coded separately
  - GFATM spending for TB under HIV grant is coded as HIV

Multiple classifications

ICD Code	ICD Term	Note	GFATM Term
B20	HIV		HIV
B59	Pneumocystis		HIV
A15	TB	*TB with HIV*	HIV-TB
A15	TB	*Other TB*	TB

- Must estimate splitting of HIV into HIV and HIV-TB
- Map HIV OI codes to HIV

Over-sampling for GFATM

- Survey samples may not pick up representative samples of HIV/TB/Malaria
- Solution:
  - Over-sample to increase number of records
    - Eg Hospitals treating GFATM diseases
    - Eg Over-sample records at individual facilities if patient has TB/HIV/Malaria
  - Reweight in analysis phase

Post-analysis cleaning

Final processing routines

- Missing values - age, sex, disease
  - Generally handle at end of each step
  - Use information on distribution from known cases to distribute unknown cases
  - If a record has a missing value, split the record to reflect the likely distribution of values, weighting each new record

- Check and verify final distributions by age, sex and disease for face validity