Lecture 2: Direct Standardization of Measures of Disease Occurrence

Dankmar Böhning

Southampton Statistical Sciences Research Institute
University of Southampton, UK

March 2 - 4, 2015

Purpose

The purpose of this material is to provide an introduction to the problems of medical surveillance and associated standardization problems:

- comparing disease (risk factor) occurrence
- standardization methodology
- examples

Medical Surveillance

Example on problems with comparison of rates

The Directly Standardized Rate

How to execute in STATA?

Definition

detection of the occurrence of health-related events or exposures in a target population

Goal

to identify changes in the distributions of diseases in order to prevent or control these diseases within a population

potential specific goals

- identification of pattern of disease occurrence
- detection of disease outbreaks
- development of clues about possible risk factors (ecological study)
- finding of cases for further investigation
- anticipation of health service needs

traditionally

medical surveillance activities were developed to monitor the spread of infectious disease through a population

today

target are all diseases and health related conditions and exposures such as traffic accident morbidity and mortality, smoking, sexual habits, etc

Data Sources

Surveillance of deaths

- mortality statistics

Surveillance of morbidity

- important function of registries such as cancer registries, traffic accident registries, etc.
- legislation on certain transmittable diseases

Surveillance of risk factors

- micro-census
- survey

to detect change

morbidity or mortality needs frequently be compared

- in time (weekly, monthly, yearly, ...)
- in space (county, states, city-areas, ...)

such a comparison - if done without care - can be quite problematic

Comparing Mortality from Lung Cancer in Berlin (West) 1960 and 1989

age-group	deaths 1989	under risk	deaths 1960	under risk
35-39	3	78862	2	44454
40-44	15	74485	5	38932
45-49	49	96516	24	66595
50-54	64	78693	63	83553
55-59	88	48942	145	83353
60-64	83	38789	202	65947
65-69	125	29128	181	50805
70-74	86	19168	160	40282
75-79	126	25109	114	25545
80-84	113	17417	43	12431
85+	54	8821	9	4183
total	**806**	**515930**	**948**	**516080**

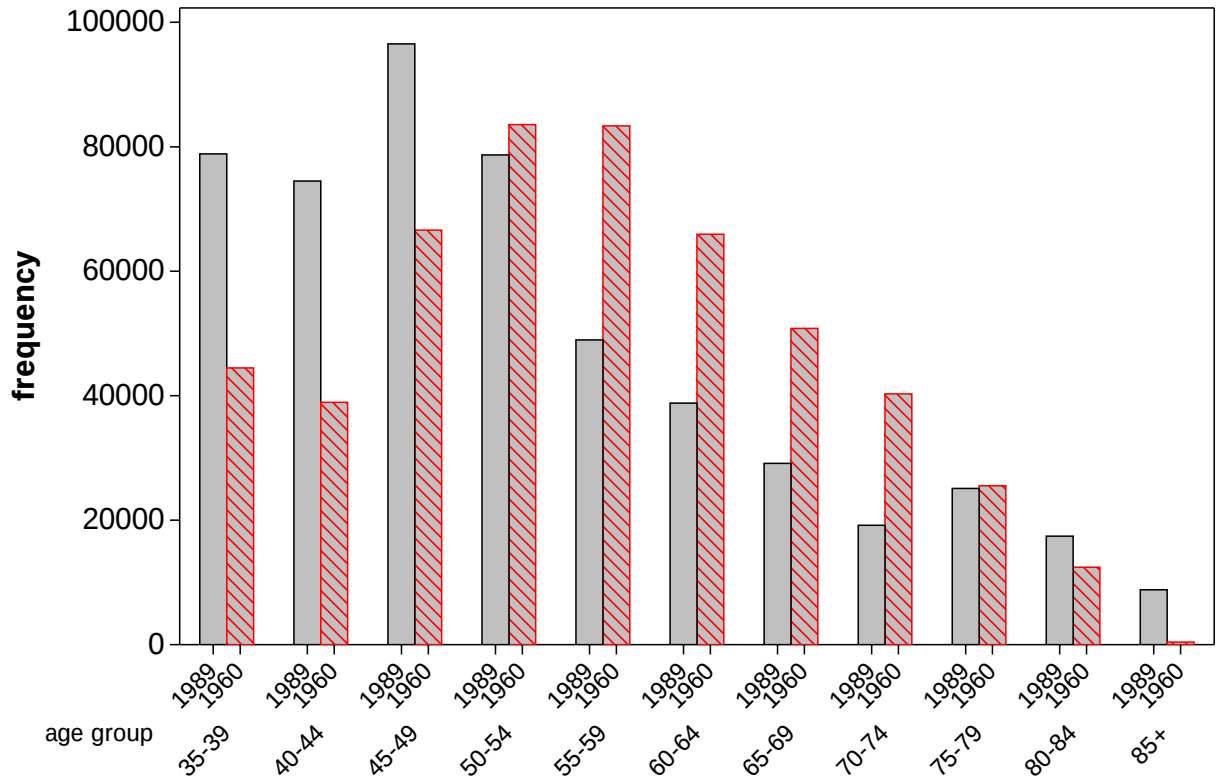
Comparing Mortality from Lung Cancer in Berlin (West) 1960 and 1989

- mortality rate 1960 $= \frac{948}{516080} \times 1000 = 1.84$
- mortality rate 1989 $= \frac{806}{515930} \times 1000 = 1.56$

coming to the **perplexing conclusion** that mortality has **dropped** from 1960 to 1989!

Comparing Mortality Rates from Lung Cancer in Berlin (West) 1960 and 1989

age-group	mortality rate 1989	mortality rate 1960
35-39	0.04	0.04
40-44	0.20	0.13
45-49	0.51	0.36
50-54	0.81	0.75
55-59	1.89	1.74
60-64	2.14	3.06
65-69	4.29	3.56
70-74	4.49	3.97
75-79	5.02	4.46
80-84	6.49	3.46
85+	6.12	2.15
total	**1.56**	**1.84**


frequency
100000
80000
60000
40000
20000
0
1989
1960
age group
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85+

Explanation

- age distributions 1960 and 1989 are quite different
- 1989 age distribution puts more weight on younger ages
- 1960 age distribution puts more weight on older ages
- hence **crude rates** are not comparable

Solution

use identical age distribution

- World (Segi's Standard)
- Europe
- national

Two Reference Populations

age-group	World	Europe
...	...	...
35-39	6000	7000
40-44	6000	7000
45-49	6000	7000
50-54	5000	7000
55-59	4000	6000
60-64	4000	5000
65-69	3000	4000
70-74	2000	3000
75-79	1000	2000
80-84	500	1000
85+	500	1000
total	**100000**	**100000**

Construction of Directly Standardized Rate

	study population			reference population
age-group	**deaths**	**at risk**	**rate**	**at risk**
1	d_1	n_1	$p_1 = \frac{d_1}{n_1}$	N_1
2	d_2	n_2	$p_2 = \frac{d_2}{n_2}$	N_2
...	...	...	...	
k	d_k	n_k	$p_k = \frac{d_k}{n_k}$	N_k
total	d	n	$p = \frac{d}{n}$	N

crude rate:

$$p = \sum_{i=1}^{k} \frac{d_i}{n_i} \times \frac{n_i}{n}$$

standardized rate:

$$p_{\mathrm{DS}} = \sum_{i=1}^{k} \frac{d_i}{n_i} \times \frac{N_i}{N}$$

Computing the Standardized Mortality Rate for Lung Cancer in Berlin (West) 1989

age	deaths	under risk	rate	World	Expect.
35-39	3	78862	3/78862=0.00004	6000	0.23
40-44	15	74485	15/74485=0.00020	6000	1.21
45-49	49	96516	49/96516=0.00051	6000	3.05
50-54	64	78693	64/78693=0.00081	5000	4.07
...	...	...	...	...	
85+	54	8821	54/8821=0.00612	500	3.06
total	**806**	**515930**		**38000**	**57.47**

standardized rate (1989):

$$p_{\mathrm{DS}} = \frac{57.47}{38000} \times 1000 = 1.51$$

and, similarly, (1960): $p_{\mathrm{DS}} = \frac{52.08}{38000} \times 1000 = 1.37$

how to execute in STATA?

organization of data

first a data file needs to be constructed containing

- the stratums variable (age)
- the event variable (cases or deaths)
- the population size variable (population)
- the group variable containing information on the groups to be compared (year)

an example is given as follows:

	age	death	population	Year
1.	35-39	3	78862	1989
2.	40-44	15	74485	1989
3.	45-49	49	96516	1989
4.	50-54	64	78693	1989
5.	55-59	88	48942	1989
6.	60-64	83	38789	1989
7.	65-69	125	29128	1989
8.	70-74	86	19168	1989
9.	75-79	126	25109	1989
10.	80-84	113	17417	1989

	age	death	population	Year
11.	85+	54	8821	1989
12.	35-39	2	44454	1960
13.	40-44	5	38932	1960
14.	45-49	24	66595	1960
15.	50-54	63	83553	1960
16.	55-59	145	83353	1960
17.	60-64	202	65947	1960
18.	65-69	181	50805	1960
19.	70-74	160	40282	1960
20.	75-79	114	25545	1960
21.	80-84	43	12431	1960
22.	85+	9	4183	1960

how to execute in STATA?

organization of data

a second data file needs to be constructed containing

- the stratums variable (age) matching with **exactly the same name**
- the population size variable containing the **reference population** carrying the same name as the study population variable

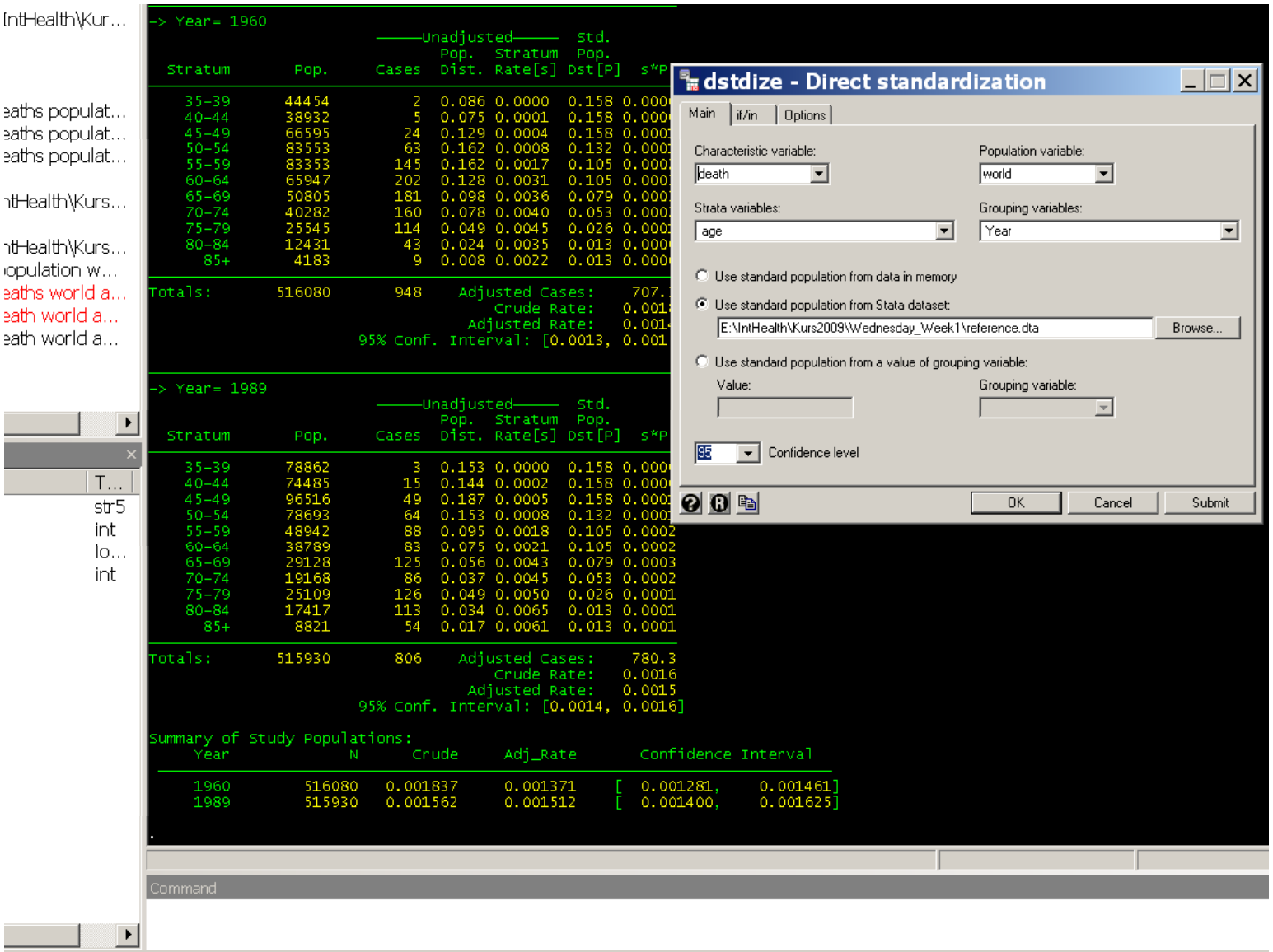
an example is given as follows in which `population` contains now the distribution of the world standard

	age	world	europe
1.	35-39	6000	7000
2.	40-44	6000	7000
3.	45-49	6000	7000
4.	50-54	5000	7000
5.	55-59	4000	6000
6.	60-64	4000	5000
7.	65-69	3000	4000
8.	70-74	2000	3000
9.	75-79	1000	2000
10.	80-84	500	1000
11.	85+	500	1000

how to execute in STATA?

execution of standardization

a very practical way to accomplish this is to choose in the first file the population name as the name of the reference standard, in this example `world`


-> Year= 1960
Stratum Pop. Cases Pop. Dist. Stratum Rate[s] Std. Pop. Dst[P]
35-39 44454 2 0.086 0.0000 0.158
40-44 38932 5 0.075 0.0001 0.158
45-49 66595 24 0.129 0.0004 0.158
50-54 83553 63 0.162 0.0008 0.132
55-59 83353 145 0.162 0.0017 0.105
60-64 65947 202 0.128 0.0031 0.105
65-69 50805 181 0.098 0.0036 0.079
70-74 40282 160 0.078 0.0040 0.053
75-79 25545 114 0.049 0.0045 0.026
80-84 12431 43 0.024 0.0035 0.013
85+ 4183 9 0.008 0.0022 0.013
Totals: 516080 948
Adjusted Cases: 707.
Crude Rate: 0.001
Adjusted Rate: 0.001
95% Conf. Interval: [0.0013, 0.001
-> Year= 1989
Stratum Pop. Cases Pop. Dist. Stratum Rate[s] Std. Pop. Dst[P]
35-39 78862 3 0.153 0.0000 0.158
40-44 74485 15 0.144 0.0002 0.158
45-49 96516 49 0.187 0.0005 0.158
50-54 78693 64 0.153 0.0008 0.132
55-59 48942 88 0.095 0.0018 0.105 0.0002
60-64 38789 83 0.075 0.0021 0.105 0.0002
65-69 29128 125 0.056 0.0043 0.079 0.0003
70-74 19168 86 0.037 0.0045 0.053 0.0002
75-79 25109 126 0.049 0.0050 0.026 0.0001
80-84 17417 113 0.034 0.0065 0.013 0.0001
85+ 8821 54 0.017 0.0061 0.013 0.0001
Totals: 515930 806
Adjusted Cases: 780.3
Crude Rate: 0.0016
Adjusted Rate: 0.0015
95% Conf. Interval: [0.0014, 0.0016]
Summary of Study Populations:
Year N Crude Adj_Rate Confidence Interval
1960 516080 0.001837 0.001371 [0.001281, 0.001461]
1989 515930 0.001562 0.001512 [0.001400, 0.001625]
dstdize - Direct standardization
Main
if/in
Options
Characteristic variable:
death
Population variable:
world
Strata variables:
age
Grouping variables:
Year
Use standard population from data in memory
Use standard population from Stata dataset:
E:\IntHealth\Kurs2009\Wednesday_Week1\reference.dta
Browse...
Use standard population from a value of grouping variable:
Value:
Grouping variable:
95
Confidence level
OK
Cancel
Submit
Command