Applied Epidemiology 1

May 12, 2017

Alessio Crippa

Biostatistics Team

Department of Public Health Sciences

Karolinska Institutet

Recap

Direct standardization

Indirect standardization

Regression approach

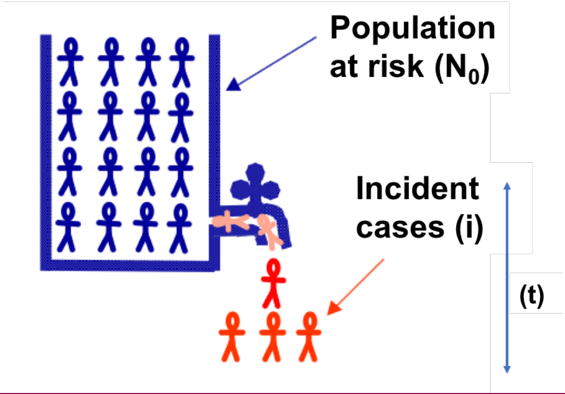
Recap

Measure of occurrence of a disease/mortality

- Risk ($\pi(t)$): what is the probability that a subject have the event within a certain time t?
- Rate (λ): how quickly a disease occurs in a population per unit of time?
- Prevalence (P): what is the probability that a condition is present in the population at a certain time?

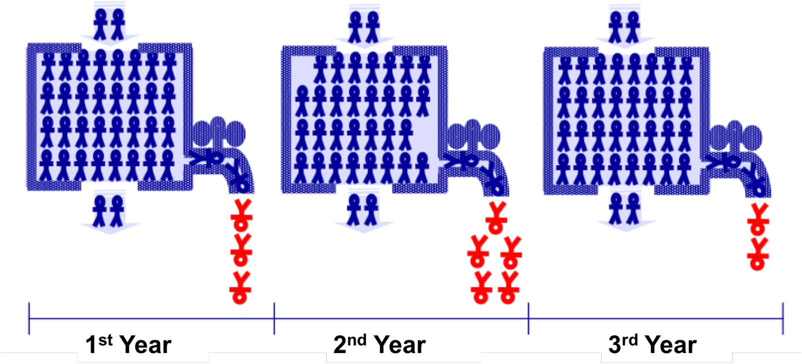
Risk

$$\pi(t) = i(t)/N_0$$



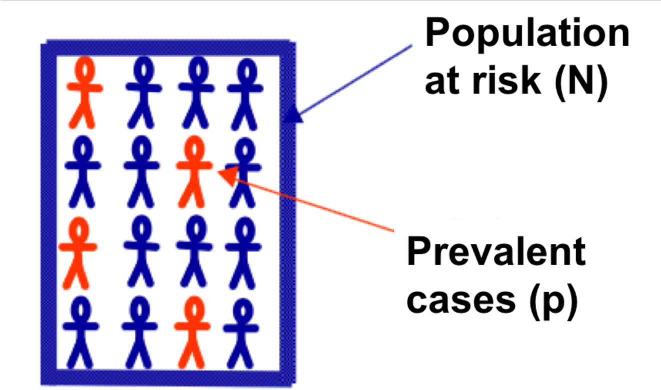
Rate

$$\lambda = i/PT$$

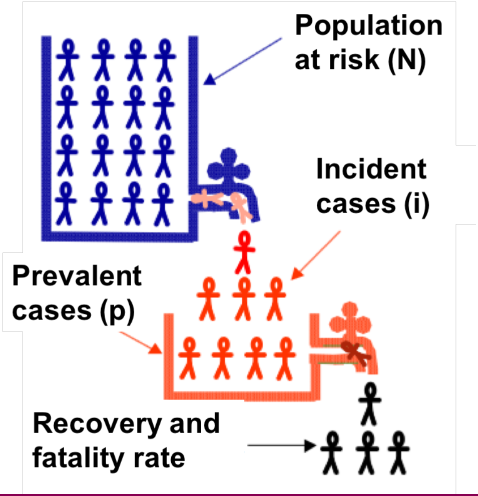


Prevalence

$$P = p/N$$



Prevalence is influenced by incidence rate, disease duration, recovery rate and fatality rate.



Let's compare the mortality rate between Swedish and Kazakhstani female in 1992 (Table 2.1 from Preston et al.)

country	deaths	population	rate (x 1000)
Kazakhstan	64572	8698860	7.42
Sweden	46256	4385469	10.55

Is the mortality in Kazakhstan lower than in Sweden?

The rate of occurrence of a disease or mortality may vary substantially with age, sex, or other important demographic characteristics.

Can the different age composition between the countries alter the crude results?

ageg	country	pop	deaths	rates	prop	country	pop	deaths	rates	prop
0	Kazakhstan	174078	3720	21.37	0.02	Sweden	59727	279	4.67	0.01
01-4	Kazakhstan	754758	1220	1.62	0.09	Sweden	229775	42	0.18	0.05
05-9	Kazakhstan	879129	396	0.45	0.10	Sweden	245172	31	0.13	0.06
10-14	Kazakhstan	808510	298	0.37	0.09	Sweden	240110	33	0.14	0.05
15-19	Kazakhstan	720161	561	0.78	0.08	Sweden	264957	61	0.23	0.06
20-24	Kazakhstan	622988	673	1.08	0.07	Sweden	287176	87	0.30	0.07
25-29	Kazakhstan	733057	752	1.03	0.08	Sweden	311111	98	0.32	0.07
30-34	Kazakhstan	732312	965	1.32	0.08	Sweden	280991	140	0.50	0.06
35-39	Kazakhstan	612825	1113	1.82	0.07	Sweden	286899	197	0.69	0.07
40-44	Kazakhstan	487996	1405	2.88	0.06	Sweden	308238	362	1.17	0.07
45-49	Kazakhstan	284799	1226	4.30	0.03	Sweden	320172	643	2.01	0.07
50-54	Kazakhstan	503608	2878	5.71	0.06	Sweden	242230	738	3.05	0.06
55-59	Kazakhstan	301879	3266	10.82	0.03	Sweden	210785	972	4.61	0.05
60-64	Kazakhstan	374317	5212	13.92	0.04	Sweden	216058	1640	7.59	0.05
65-69	Kazakhstan	256247	6866	26.79	0.03	Sweden	224479	2752	12.26	0.05
70-74	Kazakhstan	154623	6182	39.98	0.02	Sweden	222578	4509	20.26	0.05
75-79	Kazakhstan	149917	8199	54.69	0.02	Sweden	184102	6745	36.64	0.04
80-84	Kazakhstan	88716	9013	101.59	0.01	Sweden	140667	9587	68.15	0.03
85+	Kazakhstan	58940	10627	180.30	0.01	Sweden	110242	17340	157.29	0.03

How can we compare the mortality rates of two populations whose age structure (for example) substantially differ?

One possible solution is to *standardize* the populations, i.e. to make them comparable.

Standardization is a common technique in epidemiology and demography to correct/adjust measures of occurrence of a disease for major confounding variables.

Direct standardization

Direct standardization

1. calculate the number of events (deaths) expected in a standard population if this was to experience the observed (stratum-specific) rates
2. calculate the standardized rate as the number of expected events divided by the size of the standard population

Let's consider the average of the two population compositions as the standard

ageg	prop	population	std_rates
0	0.02	168.15	13.02
01-4	0.07	695.80	0.90
05-9	0.08	784.84	0.29
10-14	0.07	738.48	0.25
⋮	⋮	⋮	⋮
75-79	0.03	296.07	45.66
80-84	0.02	211.37	84.87
85+	0.02	159.57	168.80

The expected events can be calculated as

$$\text{exp deaths}_i = \frac{\text{obs deaths}_i \times \text{std pop}_i}{\text{obs pop}_i}$$

ageg	obs deaths	obs pop	std pop	exp deaths
0	279	59727	168.15	279*168.154/59727=0.785
01-4	42	229775	695.80	42*695.799/229775=0.127
05-9	31	245172	784.84	31*784.84/245172=0.099
10-14	33	240110	738.48	33*738.478/240110=0.101
15-19	61	264957	716.03	61*716.025/264957=0.165

ageg	obs deaths	obs pop	std pop	exp deaths
0	279.00	59727.00	168.15	0.79
01-4	42.00	229775.00	695.80	0.13
05-9	31.00	245172.00	784.84	0.10
10-14	33.00	240110.00	738.48	0.10
15-19	61.00	264957.00	716.03	0.16
20-24	87.00	287176.00	685.50	0.21
25-29	98.00	311111.00	776.06	0.24
30-34	140.00	280991.00	741.29	0.37
35-39	197.00	286899.00	679.35	0.47
40-44	362.00	308238.00	631.93	0.74
45-49	643.00	320172.00	528.74	1.06
50-54	738.00	242230.00	565.64	1.72
55-59	972.00	210785.00	413.84	1.91
60-64	1640.00	216058.00	461.49	3.50
65-69	2752.00	224479.00	403.22	4.94
70-74	4509.00	222578.00	342.64	6.94
75-79	6745.00	184102.00	296.07	10.85
80-84	9587.00	140667.00	211.37	14.41
85+	17340.00	110242.00	159.57	25.10
	46256.00	4385469.00	10000.00	73.74

The standardized mortality rate for Sweden can be calculated as the ratio between the expected deaths and the total of the standard population

$1000*73.741/10000 = 7.374$

The same can be done for the Kazakhstani female

country	std
Kazakhstan	11.88
Sweden	7.37

Standard population

https://seer.cancer.gov/stdpopulations/

Standard Populations (Millions) for Age-Adjustment

Standard populations, often referred to as standard millions, are the age distributions used as weights to create age-adjusted statistics. Files containing standard population data for use in statistical software are available below. These contain the same data distributed with the SEER*Stat software. SEER also provides US Population Data which can be used for analyses with SEER*Stat or other software.

Starting with the November 2004 SEER submission of data (diagnoses through 2002), the SEER Program age-adjusts using the 2000 US standard population based on single years of age from the Census P25-1130 (PDF)[1] series estimates of the 2000 US population. For the 5-year age groups, the single year of age populations are summarized from the five single-year of age populations. See 2000 US Standard Population vs. Standard Million for more discussion.

Standard Population Files

The standard population data files contain the following data:

- US Standards (1940, 1950, 1960, 1970, 1980, 1990, 2000)
- World (Segi 1960) Standard[2]
- Canadian Standards (1991, 1996)

Indirect standardization

It is particular useful when we don't have age-specific rates, but can easily obtain the age distribution.

ageg	country	pop	deaths	country	pop	deaths	rates
0	Kaz	174078		Sw	59727	279	4.67
01-4	Kaz	754758		Sw	229775	42	0.18
05-9	Kaz	879129		Sw	245172	31	0.13
⋮	⋮	⋮		⋮	⋮	⋮	⋮
80-84	Kaz	88716		Sw	140667	9587	68.15
85+	Kaz	58940		Sw	110242	17340	157.29
Tot	Kaz	8698860	64572	Sw	4385469	46256	

Indirect standardization

1. calculate the expected number of events in the population under study by applying age-specific rates from the standard population;
2. calculate the standardized mortality ratios (SMR) by dividing the observed deaths to the expected ones.

Let consider the population of Swedish female 1992 as the standard population.

The expected deaths in Kazakhstan using the Swedish mortality rates are:

$$\text{exp deaths}_i = \text{std rate}_i \times \text{obs pop}_i = \frac{\text{std deaths}_i \times \text{obs pop}_i}{\text{std pop}_i}$$

ageg	obs pop	std deaths	std pop	std rate	exp deaths
0	174078	279	59727	4.67	279*174078/59727=813.163
01-4	754758	42	229775	0.18	42*754758/229775=137.96
05-9	879129	31	245172	0.13	31*879129/245172=111.159
10-14	808510	33	240110	0.14	33*808510/240110=111.119
15-19	720161	61	264957	0.23	61*720161/264957=165.8

ageg	obs pop	std deaths	std pop	std rate	exp deaths
0	174078.00	279.00	59727.00	4.67	813.16
01-4	754758.00	42.00	229775.00	0.18	137.96
05-9	879129.00	31.00	245172.00	0.13	111.16
10-14	808510.00	33.00	240110.00	0.14	111.12
15-19	720161.00	61.00	264957.00	0.23	165.80
20-24	622988.00	87.00	287176.00	0.30	188.73
25-29	733057.00	98.00	311111.00	0.32	230.91
30-34	732312.00	140.00	280991.00	0.50	364.86
35-39	612825.00	197.00	286899.00	0.69	420.80
40-44	487996.00	362.00	308238.00	1.17	573.11
45-49	284799.00	643.00	320172.00	2.01	571.96
50-54	503608.00	738.00	242230.00	3.05	1534.34
55-59	301879.00	972.00	210785.00	4.61	1392.06
60-64	374317.00	1640.00	216058.00	7.59	2841.27
65-69	256247.00	2752.00	224479.00	12.26	3141.46
70-74	154623.00	4509.00	222578.00	20.26	3132.36
75-79	149917.00	6745.00	184102.00	36.64	5492.55
80-84	88716.00	9587.00	140667.00	68.15	6046.34
85+	58940.00	17340.00	110242.00	157.29	9270.69
					36540.67

The standardized mortality ratios for Kazakhstan can be calculated as the ratio between the observed deaths and the expected deaths

$\text{SMR} = 64572/36540.665 = 1.767$

The standardized mortality ratios for female in Kazakhstan 1992 is 77% higher as compared to Sweden.

country	crude	std	smr
Kazakhstan	7.42	4.2	1.77

Choice of method depends on available data.

If age-specific rates are available for the studied population, direct standardization is preferable.

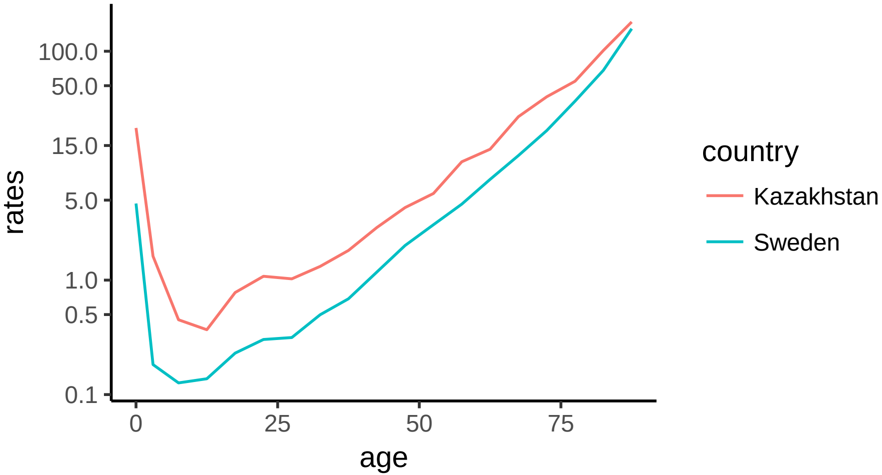
Contrary, if only age-specific rates for the standard population are available, indirect standardization should be the alternative choice.

Standardization is useful to

- compare changes over time (one or more groups/populations at several different time points);
- compare more than one group/populations at one point in time.

Standardization is not appropriate when exact numbers are needed (i.e. for planning health care interventions),

Age-Patterns of Mortality



Regression approach

Regression models are the standard tool for adjust for confounding.

We can analyze the same data using a Poisson regression (appropriate for count data)

$$\log(\lambda|X,Z) = \log(E[Y|X,Z]) - \log(PT) = \beta_0 + \beta_1 X + \beta_2 Z$$

```
              exp(Est.)    2.5%   97.5%
(Intercept)      0.0117  0.0114  0.0121
kz               1.6151  1.5954  1.6351
ageg01-4         0.0743  0.0697  0.0791
ageg05-9         0.0219  0.0198  0.0242
ageg10-14        0.0183  0.0163  0.0204
ageg15-19        0.0371  0.0341  0.0404
ageg20-24        0.0501  0.0464  0.0541
ageg25-29        0.0485  0.0450  0.0522
ageg30-34        0.0643  0.0602  0.0688
ageg35-39        0.0875  0.0822  0.0931
ageg40-44        0.1374  0.1299  0.1453
ageg45-49        0.2042  0.1933  0.2158
ageg50-54        0.2920  0.2792  0.3054
ageg55-59        0.5173  0.4954  0.5402
ageg60-64        0.7118  0.6845  0.7401
ageg65-69        1.2843  1.2378  1.3327
ageg70-74        1.9295  1.8603  2.0013
ageg75-79        2.9886  2.8857  3.0952
ageg80-84        5.5837  5.3949  5.7791
ageg85+         11.6044 11.2229 11.9989
```

Exponentiating the β_1 we see that age-adjusted mortality rate is $100*(1.615 - 1) = 61.5\%$ higher than in Sweden.