Demographic shift disproportionately increases the cancer burden in an aging nation – current and expected incidence and mortality in Hungary up to 2030

Supplementary materials

Performing the age-period-model for generating projections

Methods used for projection generation

Trends in incidence and mortality were predicted by the R software package Nordpred (https://www.kreftregisteret.no/en/Research/Projects/Nordpred/Nordpred-software/). Predictions were generated in a two-step process, including:

1. Fitting a model to the observed data
2. Based on the fitted model generating a projection.

In the first step, Nordpred fitted an age-period-cohort regression model to the observed data.

Incidence/Mortality cases = (Age) + (Period) + (Cohort) + (Drift)

The variables implemented in our calculations:

Age (categorical variable): Coded age groups: 0-4 years =>1; 5-9 years =>2; … 85+ years=>18

Period (categorical variable): Coded periods – 2001-2005 => 1; 2006-2010 =>2; 2011-2015=>3

Cohort (categorical variable): Derived variable from the maximum of age – age + period.

Drift (continuous variable): The linear effects of period and cohort cannot be estimated simultaneously as a result of collinearity between age, period and cohort, therefore a common linear trend (Drift) is estimated.

Estimation of the model is based on assumptions that observed cohort specific patterns and period effects will be similar in the future, therefore cohort and period effects are equal to the last estimated effects in the fitted model (additional details discussed in references [1, 2]).

Description of the applied model

Inputs for model predictions are the new cancer cases diagnosed in 2001-2014, cancer mortality data in 1996-2015, observed population data in 1996-2014 and projected population size and structure up to 2030 in Hungary.

Data were aggregated into 5-year periods (mortality: 1996–2000; 2001-2005; 2006-2010; 2011–2015; incidence: 2001-2005; 2006-2010; 2011-2015) and 5-year age groups (0–4; 5–9; …80–84; 85+).

A lowest age group for each cancer site was determined by considering the number of cases in each cell. Calculations for age groups below this age limit were based on average rates in the last 10 observed years. For fitting the regression model, we have used one or two age groups below the age limit in order to get an improved estimation for the younger cohorts.

Although the package is capable to do a goodness-of-fit test to select the best fitted model from the observation intervals, we used fixed periods in our regression models due to limited observation periods (three 5-year periods for incidence and four 5-years periods for mortality).

The Drift parameter was cut by 0%, 25% and 50% in the first, second and third 5-year period, respectively, to decrease the effect of current trends, because it is unlikely trends continue at the same rate. Furthermore, if there was a statistically significant change between observed periods, we had used the most recent 2 periods (last 10 years) to make projections for drift.

Applied age limits, observed periods and drift trends for males are listed below.

	Incidence			Mortality		
	Age	Base	Trend	Age	Base	Trend
Male total	0-4	3	T	0-4	4	T
Thyroid (C73)	20-24	3	A	-	-	-
Lung and Bronchus (C33-C34)	25-29	3	T	35-39	4	T
Colorectal (C18–C21)	25-29	3	T	35-39	4	T
Prostate (C61)	25-29	3	T	55-59	4	T
Oral cavity and pharynx (C00–C14)	25-29	3	T	40-44	4	T
Bladder (C67)	25-29	3	A	45-49	4	A
Stomach (C16)	25-29	3	T	40-44	4	T
Kidney (C64)	25-29	3	T	40-44	4	T
Pancreas (C25)	25-29	3	A	35-39	4	A
Larynx (C32)	25-29	3	A	45-49	4	T
Brain (C71)	0-4	3	A	0-4	4	A
Leukemia (C91-C95)	0-4	3	A	0-4	4	A
Melanoma of skin (C43)	25-29	3	A	35-39	4	T
Liver (C22)	0-4	3	T	45-49	4	A
Non-Hodgin Lymphoma (C82-C85, C88, C96)	0-4	3	A	30-34	4	A
Hodgkin lymphoma (C81)	-	-	-	40-44	4	T
Multiple myeloma (C90)	-	-	-	50-54	4	T
Oesophagus (C15)	40-44	3	A	45-49	4	T
Testis (C62)	25-29	3	T	25-29	4	T
Gallbladder and billiary tract (C23-C24)	45-49	3	A	45-49	4	T
Bones and joints (C40-C41)	0-4	3	A	20-24	4	T
Soft tissue (C47, C49)	0-4	3	T	-	-	-

Age: Lower limit of the age group., Base: Number of periods used, Trend: Average or recent trend was projected

Example of performing the age-period-cohort model for male thyroid cancer incidence

Fitting a model to the observed data:

A. Input: male thyroid cancer incidence data

		Five-year periods		
		2001-2005	2006-2010	2011-2015
Five-year age groups	00-04	0	0	0
	05-09	3	0	6
	10-14	5	1	5
	15-19	8	10	13
	20-24	16	12	9
	25-29	33	22	38
	30-34	31	47	58
	35-39	31	54	75
	40-44	33	45	81
	45-49	53	63	82
	50-54	76	55	78
	55-59	86	79	91
	60-64	62	88	93
	65-69	46	63	80
	70-74	56	62	60
	75-79	54	48	41
	80-84	13	24	29
	85+	11	11	16

B. Input: observed and forecasted male population data

		Five-year periods					
		Observed population			Forecasted population		
		2001-2005	2006-2010	2011-2015	2016-2020	2021-2025	2026-2030
Five-year age groups	00-04	1 229 755	1 243 299	1 192 325	1 163 251	1 142 915	1 063 556
	05-09	1 388 377	1 240 958	1 250 715	1 187 056	1 161 167	1 141 236
	10-14	1 567 036	1 394 279	1 243 729	1 244 565	1 185 721	1 160 074
	15-19	1 651 950	1 577 272	1 441 807	1 243 234	1 242 675	1 184 155
	20-24	1 916 791	1 674 322	1 616 688	1 434 847	1 231 798	1 234 177
	25-29	2 115 357	1 936 376	1 602 246	1 557 512	1 391 030	1 199 480
	30-34	1 832 465	2 127 467	1 834 502	1 514 594	1 496 721	1 344 404
	35-39	1 600 198	1 834 730	2 102 931	1 761 663	1 459 530	1 453 596
	40-44	1 609 462	1 583 572	1 821 784	2 046 071	1 711 477	1 421 402
	45-49	1 944 914	1 558 840	1 563 638	1 759 944	1 984 903	1 664 434
	50-54	1 750 608	1 833 908	1 512 557	1 489 757	1 683 118	1 909 886

	55-59	1 425 160	1 609 570	1 746 258	1 410 628	1 393 899	1 587 520
	60-64	1 198 293	1 271 236	1 490 213	1 589 917	1 287 824	1 286 228
	65-69	978 907	1 026 718	1 132 130	1 311 410	1 411 636	1 157 546
	70-74	835 626	781 871	851 942	947 442	1 114 576	1 218 729
	75-79	586 898	600 664	577 198	652 575	745 812	900 067
	80-84	309 130	355 291	373 516	382 961	455 790	540 104
	85+	162 749	202 796	229 511	276 396	327 595	414 713

C. Calculating a general linear model fitted to the observed data (Supplementary File Table 4)

Call: glm(formula = Cases ~ as.factor(Age) + Period + as.factor(Period) + as.factor(Cohort) - 1, family = power5link, data = apcdata)

Results of the general linear model

as.factor(Age)2	as.factor(Age)3	as.factor(Age)4	as.factor(Age)5	as.factor(Age)6
0.074446	0.083792	0.095697	0.095689	0.111298
as.factor(Age)7	as.factor(Age)8	as.factor(Age)9	as.factor(Age)10	as.factor(Age)11
0.120059	0.127060	0.133373	0.145449	0.147640
as.factor(Age)12	as.factor(Age)13	as.factor(Age)14	as.factor(Age)15	as.factor(Age)16
0.152421	0.152282	0.149782	0.154532	0.153997
as.factor(Age)17	as.factor(Age)18	Period	as.factor(Period)2	as.factor(Period)3
0.145489	0.142583	0.003964	-0.001553	NA
as.factor(Cohort)2	as.factor(Cohort)3	as.factor(Cohort)4	as.factor(Cohort)5	as.factor(Cohort)6
-0.013043	-0.003896	-0.009885	-0.014663	-0.016122
as.factor(Cohort)7	as.factor(Cohort)8	as.factor(Cohort)9	as.factor(Cohort)10	as.factor(Cohort)11
-0.012759	-0.019604	-0.027103	-0.020608	-0.017287
as.factor(Cohort)12	as.factor(Cohort)13	as.factor(Cohort)14	as.factor(Cohort)15	as.factor(Cohort)16
-0.010707	-0.008739	-0.008142	-0.007419	-0.013209
as.factor(Cohort)17	as.factor(Cohort)18	as.factor(Cohort)19		
-0.012473	-0.018058	NA		

Degrees of Freedom: *51 Total (i.e. Null); 15 Residual*
Null Deviance: *168400*
Residual Deviance: *23.73*
AIC (Akaike Information Criterion): *358.3*

Generating a projection based on the fitted model

		Five-year periods					
		Observed cases			Predicted cases		
		2001-2005	2006-2010	2011-2015	2016-2020	2021-2025	2026-2030
Five-year age groups	00-04	0	0	0	0	0	0
	05-09	3	0	6	3	3	3
	10-14	5	1	5	3	3	3
	15-19	8	10	13	12	14	14
	20-24	16	12	9	14	14	15
	25-29	33	22	38	30	31	29
	30-34	31	47	58	53	47	47
	35-39	31	54	75	78	74	64
	40-44	33	45	81	112	106	97
	45-49	53	63	82	136	181	165
	50-54	76	55	78	100	154	199
	55-59	86	79	91	99	121	180
	60-64	62	88	93	89	99	118
	65-69	46	63	80	87	80	88
	70-74	56	62	60	92	96	87
	75-79	54	48	41	56	78	81
	80-84	13	24	29	26	33	46
	85+	11	11	16	20	22	29

Prediction done with:

Number of periods predicted (nopred):	3
Trend used in predictions (cuttrend):	0, 0.25, 0.5
Number of periods used in estimate (noperiod):	3
P-value for goodness of fit:	0.0698
Used recent (recent):	FALSE
P-value for recent:	0.4062
First age group used (startuseage):	4
First age group estimated (startestage):	2

References:

1. Moller, B., et al., *Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches.* Stat Med, 2003. **22**(17): p. 2751-66.

2. Moller, B., H. Weedon-Fekjaer, and T. Haldorsen, *Empirical evaluation of prediction intervals for cancer incidence.* BMC Med Res Methodol, 2005. **5**: p. 21.