Life Expectancy in China and the Contribution from Regional Dynamics

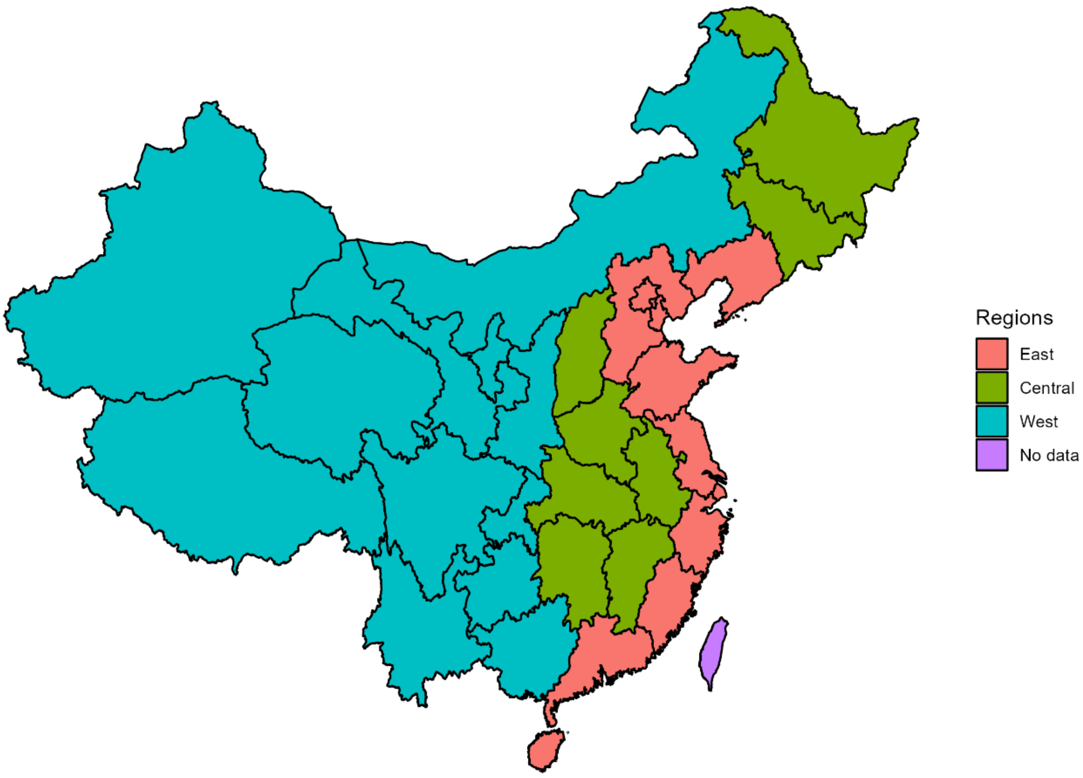
Appendix

Table of contents

Appendix 1. Classification of Regions

Regions	Provinces
East	Hebei, Liaoning, Fujian, Jiangsu, Zhejiang, Shandong, Guangdong, Hainan, Shanghai, Beijing, Tianjin
Central	Heilongjiang, Jilin, Shanxi, Anhui, Jiangxi, Henan, Hubei, Hunan
West	Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang

China is a vast and varied nation with substantial regional social and economic differences, especially across the East, Central, and West parts of the country. East China, which includes the provinces of Hebei, Liaoning, Fujian, Jiangsu, Zhejiang, Shandong, Guangdong, Hainan, and the municipalities of Shanghai, Beijing, and Tianjin, is more developed and urbanized than the rest of the nation. Central China consists of the eight provinces of Heilongjiang, Jilin, Shanxi, Anhui, Jiangxi, Henan, Hubei, and Hunan. West China includes Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang. The classification is illustrated in the map below. Note that the map below excludes the Sino-India disputed areas and the South China Sea due to lack of data and technical constraints.



According to the National Bureau of Statistics of China, Urban Population refers to the total population of districts under the jurisdiction of a city with a district establishment, the population of street committees under the jurisdiction of a city without a district establishment, the population of resident committees of towns under the jurisdiction of a city without district establishment, and the population of resident committees of towns under the jurisdiction of a county. Statistically, urban areas are classified as the entire jurisdiction of a city with a population density above 1,500 people per square kilometer of land; the jurisdiction of the seat of city government, county government, and street committees in cities that have a population density below 1,500 people per square kilometer of land; the entire jurisdiction of a town or township if the construction of the urban area in the seat of city government, county government has been extended to the surrounding establishment of the town or township. The Rural Population refers to the total population in China except for the urban population.

For China, the different regions of East, Central, and West, as well as different subpopulations of Urban-Rural creates a significant divide in terms of economic development and therefore disposable income for each individual (see the table A1), which has been shown to greatly influence individual health and lifespan. This regional heterogeneity serves as a proxy for our assessment of the divide in health outcomes (life expectancy) across different subpopulations of China.

Table A1. Chinese regions and their population and disposable income.

Regions	Population (million) and % in the total population	Average Disposable Income (Chinese Yuan)
Urban-West	219.32 (15.6%)	37,548
Rural-West	163.76 (11.6%)	14,111
Urban-Central	250.80 (17.8%)	37,658
Rural-Central	169.35 (12.0%)	16,213
Urban-East	430.74 (30.5%)	52,027
Rural-East	176.16 (12.5%)	21,286

Data source: China Statistical Yearbook 2021. National Bureau of Statistics of China.
http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm

Appendix 2. Data adjustment measures and age-specific mortality rates before and after adjustment

Data adjustment measures

The main data source used in this study was the Disease Surveillance Points (DSP) system in China (Chinese Center for Disease Control and Prevention, 2010, 2020). To enhance the data quality, under-five mortality rates were adjusted for underreporting using data from the Maternal and Child Health Surveillance (MCHS) system (National Health Commission, 2020) as a benchmark. The under-reporting rates were generated by comparing DSP and the Maternal and Child Health Surveillance (MCHS) system data, which has higher data quality for under-five mortality, but they are aggregated at the urban/rural level (National Health Commission 2020). We smoothed the population counts data with the "ungroup" package in R version 4.0.3 (Pascariu et al., 2018) and adjusted the adult mortality rates with the two-dimensional log-quadratic mortality model (Wilmoth et al., 2012) using under-five mortality and life expectancy as the input. Old-age mortality rates were extrapolated from age 85 to 110+ in the log-quadratic model. Specifically, we adjusted the raw data with the following steps.

1) Calculate the under-five mortality underreporting ratio. Using under-five mortality data in 2010 and 2020 from the MCHS system as a reference, the underreporting ratio for each subgroup (by sex and urban/rural sector) was obtained. As there is no regional data in the MCHS system, the regional populations within the same sex and urban/rural sector were adjusted using the same underreporting ratio, calculated as

$$u_x^{i,s,t} = \frac{m_{x\,(MCHS)}^{i,s,t}}{m_{x\,(DSP)}^{i,s,t}},$$

where $m_{x\,(MCHS)}^{i,s,t}$ and $m_{x\,(DSP)}^{i,s,t}$ are the child mortality, x = 0-5, for location i = urban/rural, sex s and time t, arising from MCHS and DSP data respectively.

2) Adjust the under-five mortality on the regional and aggregated level. The original DSP child mortality was adjusted using the underreporting ratio.

Aggregated level (e.g. urban male in 2010):

$$\widehat{m}_x^{i,s,t} = u_x^{i,s,t} \cdot m_{x\,(DSP)}^{i,s,t}$$

Regional level (e.g. urban-east male in 2010):

$$\widehat{m}_x^{i,j,s,t} = u_x^{i,s,t} \cdot m_{x\,(DSP)}^{i,j,s,t},$$

where j denotes region (East/ Central/ West).

3) Adjust death numbers in the under-five age groups to ensure that the sum of regional deaths still equals the deaths at the higher/aggregated level.

Using the counts of individuals by age, sex and location $P_x^{i,s,t}$ as well as region $P_x^{i,j,s,t}$ for each time t, we calculate their share of the population in each region

$$c_x^{i,j,s,t} = \frac{P_x^{i,j,s,t}}{P_x^{i,s,t}},$$

and the ratio of the adjusted under-five mortality over the sum of adjusted mortality by their share of the population

$$R_x^{i,j,s,t} = \frac{\widehat{m}_x^{i,j,s,t}}{\sum_j \widehat{m}_x^{i,j,s,t}\, c_x^{i,j,s,t}},$$

we obtain a new adjusted death rate that ensures that the regions will add (when the $c_x^{i,j,s,t}$ weight is included) to the total mortality, $\widehat{m}_x^{i,s,t}$, as

$$\widetilde{m}_x^{i,j,s,t} = \widehat{m}_x^{i,s,t}\, R_x^{i,j,s,t}.$$

4) Use the log-quadratic model to get a smooth age-pattern of mortality. The equation of the log-quadratic model is

$$\log(m_x) = a_x + b_x h + c_x h^2 + v_x k$$

where h is the logarithm of under-five mortality; the coefficients (a_x, b_x, c_x) are those originally estimated from the model with historical data from countries included in the Human Mortality Database (HMD); and the second parameter v_x corresponds to adult mortality between ages 15 to 60 (${}_{45}q_{15}$), with value k obtained iteratively to reduce the error term.

In our case, the underreporting-adjusted under-five mortality calculated in step (3) and the original adult mortality from DSP are used as the input value for the log-quadratic model to generate smoothed mortality rates at the regional level ($\breve{m}_x^{i,j,s,t}$) and aggregated level ($\breve{m}_x^{i,s,t}$) for ages 0-110+.

5) Adjust the smoothed regional mortality rates with a similar method as that used in step (3). Using the same regional population composition as in step (3), namely $c_x^{i,j,s,t}$ we estimate the ratio of the smoothed death rates $\breve{m}_x^{i,j,s,t}$ over the sum of the product of those rates by the population composition

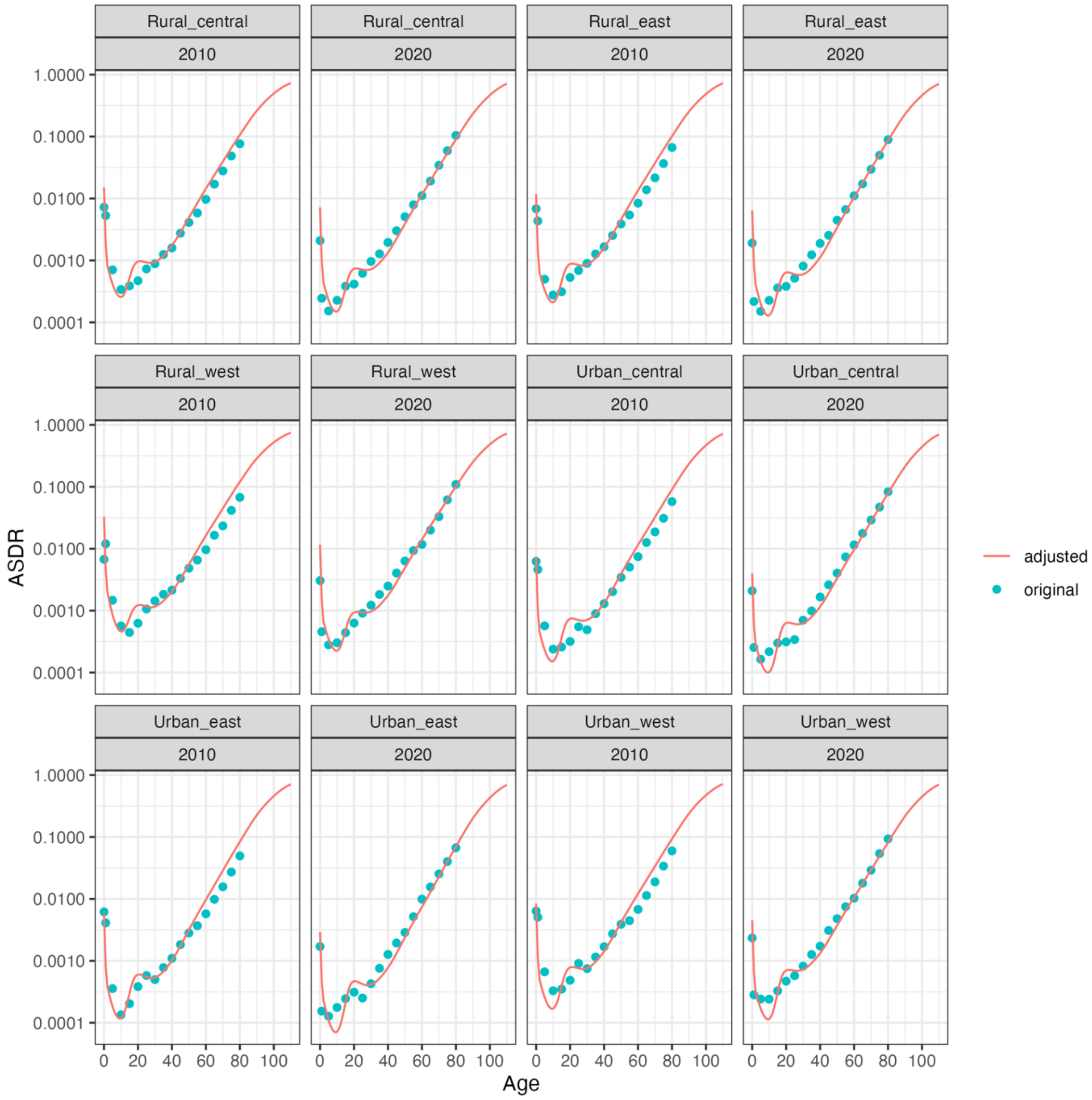
$$r_x^{i,j,s,t} \quad = \frac{\breve{m}_x^{i,j,s,t}}{\sum_j \breve{m}_x^{i,j,s,t} \quad c_x^{i,j,s,t}},$$

we obtain a final death rate that assures that the regions will add (when the $c_x^{i,j,s,t}$ weight is included) to the total mortality, $\breve{m}_x^{i,s,t}$, as

$$\underline{m}_x^{i,j,s,t} \quad = \breve{m}_x^{i,s,t} \ \ r_x^{i,j,s,t} \quad .$$

Data adjustment results

Figure A1. Age-specific mortality rates before and after adjustment for the male subnational populations in China, 2010 and 2020.



Data source: Authors' calculation derived with data from Chinese Disease Surveillance Points (DSP) 2010 and 2020.

Table A2.1. Population weights in the DSP data –Males 2010

Age	Total	Urban-east	Urban-central	Urban-west	Rural-east	Rural-central	Rural-west
0-1	540,700	13.4	9.2	6.7	23.4	25.5	21.7
1-4	2,003,711	15.0	9.9	7.5	22.6	24.2	20.8
5-9	2,393,564	15.5	9.9	8.3	21.8	23.4	21.1
10-14	2,380,376	16.9	9.6	9.3	20.6	20.4	23.2
15-19	3,117,868	14.1	10.3	7.2	20.8	24.8	22.8
20-24	3,621,063	13.8	10.1	5.9	22.6	25.9	21.8
25-29	2,950,621	19.1	13.2	8.1	19.0	21.7	18.9
30-34	2,690,338	17.5	12.7	8.7	18.9	22.1	20.1
35-39	3,585,750	16.2	12.0	9.1	19.9	22.7	20.2
40-44	3,660,767	16.3	12.4	8.1	21.3	23.2	18.7
45-49	3,070,288	20.9	13.4	8.1	21.0	20.7	15.9
50-54	2,764,880	23.6	12.3	8.1	21.6	19.8	14.6
55-59	2,543,221	22.3	12.0	8.7	20.8	20.5	15.8
60-64	1,688,841	19.7	11.6	8.5	21.2	21.9	17.1
65-69	1,201,794	18.8	11.9	9.2	20.0	21.4	18.7
70-74	947,098	20.9	12.8	8.9	20.9	20.2	16.4
75-79	634,958	23.4	12.9	8.5	20.2	19.5	15.5
80-84	300,527	24.6	12.5	7.9	20.7	19.5	14.7
85+	80,166	26.4	11.0	6.9	21.2	19.5	15.1
Total	40,176,531	17.8	11.6	8.1	20.9	22.4	19.2

Table A2.2. Population weights in the DSP data –Females 2010

Age	Total	Urban-east	Urban-central	Urban-west	Rural-east	Rural-central	Rural-west
0-1	504,647	13.8	9.4	6.8	23.2	25.3	21.5
1-4	1,849,066	15.0	10.1	7.6	22.5	24.1	20.7
5-9	2,193,500	15.9	10.2	8.5	21.6	23.0	20.8
10-14	2,065,775	18.0	9.9	9.9	20.1	19.1	23.0
15-19	2,795,356	14.4	10.8	7.4	20.9	24.1	22.6
20-24	3,401,328	13.9	10.3	6.3	22.6	25.6	21.3
25-29	2,815,998	19.8	13.3	8.2	19.7	21.4	17.6
30-34	2,627,412	17.3	12.5	8.5	20.0	21.8	19.9
35-39	3,500,627	16.6	12.3	9.1	20.5	21.9	19.6
40-44	3,554,572	16.4	12.4	8.0	21.9	23.1	18.2
45-49	2,960,449	20.9	13.2	7.9	21.5	20.9	15.7
50-54	2,655,362	24.3	12.3	8.0	21.6	19.5	14.3
55-59	2,487,936	22.0	12.0	8.7	21.2	20.2	15.8
60-64	1,657,203	19.9	11.9	8.8	21.5	21.3	16.7
65-69	1,212,474	19.9	12.9	9.3	19.7	20.2	17.9
70-74	1,015,849	22.8	13.1	8.7	20.2	19.1	16.0
75-79	743,785	24.0	12.1	7.8	21.3	19.5	15.4
80-84	414,366	24.5	11.4	6.9	22.7	20.2	14.4
85+	134,390	26.1	10.3	6.0	22.9	20.5	14.2
Total	38,590,095	18.3	11.8	8.2	21.2	21.9	18.7

Table A2.3. Population weights in the DSP data –Males 2020

Age	Total	Urban-east	Urban-central	Urban-west	Rural-east	Rural-central	Rural-west
0-1	1,525,573	13.3	10.0	10.6	24.3	23.7	18.0
1-4	6,803,884	11.7	9.0	9.5	25.5	26.6	17.8
5-9	8,162,735	10.7	8.7	9.4	24.1	28.8	18.3
10-14	8,169,848	10.5	8.7	10.0	23.7	27.7	19.3
15-19	7,559,880	10.5	8.6	11.3	23.3	25.3	21.0
20-24	7,950,535	12.6	10.0	11.3	23.3	24.3	18.6
25-29	9,824,595	17.9	12.4	10.1	23.2	21.9	14.4
30-34	11,110,837	18.4	10.6	9.8	23.6	22.0	15.7
35-39	9,492,454	16.8	10.9	9.7	23.1	22.9	16.5
40-44	9,834,129	14.4	10.7	11.5	22.2	23.4	17.8
45-49	12,543,089	13.9	10.5	10.4	24.4	24.2	16.6
50-54	12,396,040	15.6	11.0	9.9	24.7	23.5	15.3
55-59	10,011,340	17.5	10.4	9.3	26.5	22.3	14.0
60-64	8,351,399	15.8	9.7	9.9	25.8	23.6	15.2

65-69	7,363,317	15.3	9.5	9.6	26.6	24.1	15.0
70-74	4,722,855	15.0	9.5	10.0	26.5	23.5	15.5
75-79	2,873,847	16.1	10.1	9.8	26.4	22.8	14.8
80-84	1,733,620	19.0	10.4	9.4	26.5	21.4	13.3
85+	1,039,574	18.8	10.4	9.5	26.4	21.5	13.3
Total	141,469,551	14.8	10.2	10.1	24.4	24.0	16.5

Table A2.4. Population weights in the DSP data –Females 2020

Age	Total	Urban-east	Urban-central	Urban-west	Rural-east	Rural-central	Rural-west
0-1	1,392,777	13.7	9.6	10.9	24.4	22.9	18.5
1-4	6,014,214	12.1	8.7	9.8	25.5	25.6	18.3
5-9	7,016,731	10.9	8.6	9.8	24.1	27.8	18.8
10-14	6,905,850	10.6	8.5	10.5	23.9	26.5	20.0
15-19	6,439,494	10.4	8.5	11.7	23.0	24.6	21.7
20-24	6,954,762	12.2	10.4	11.5	23.0	24.1	18.8
25-29	9,307,167	16.9	12.4	10.2	23.0	23.3	14.2
30-34	11,201,924	17.9	10.9	9.8	23.6	22.8	15.0
35-39	9,485,680	16.4	11.3	9.9	23.3	23.2	16.0
40-44	9,654,080	14.0	10.8	11.5	22.7	23.8	17.2
45-49	12,312,950	13.5	10.4	10.3	24.9	25.0	15.9
50-54	12,309,040	15.0	10.7	9.8	25.3	24.4	14.8
55-59	9,988,924	17.3	10.3	9.2	26.7	22.7	13.7
60-64	8,282,642	15.8	9.8	10.1	25.7	23.7	15.0
65-69	7,652,968	15.3	9.7	9.6	26.8	23.9	14.7
70-74	4,987,303	15.3	9.9	9.9	26.3	23.0	15.4
75-79	3,273,821	17.0	10.4	9.7	26.0	22.1	14.7
80-84	2,243,142	18.7	9.5	8.6	28.3	21.8	13.0
85+	1,624,840	18.6	9.6	8.7	28.2	21.9	13.0
Total	137,048,309	14.7	10.2	10.2	24.6	24.1	16.2

Table A3.1. Percentage Points Change in Population Share, Males, 2010-2020

Age	Urban-east	Urban-central	Urban-west	Rural-east	Rural-central	Rural-west
0-1	–0.1	0.8	3.9	0.9	–1.8	–3.7
1-4	–3.3	–0.9	2.0	2.9	2.4	–3.0
5-9	–4.8	–1.2	1.1	2.3	5.4	–2.8
10-14	–6.4	–0.9	0.7	3.1	7.3	–3.9
15-19	–3.6	–1.7	4.1	2.5	0.5	–1.8
20-24	–1.2	–0.1	5.4	0.7	–1.6	–3.2
25-29	–1.2	–0.8	2.0	4.2	0.2	–4.5
30-34	0.9	–2.1	1.1	4.7	–0.1	–4.4
35-39	0.6	–1.1	0.6	3.2	0.2	–3.7
40-44	–1.9	–1.7	3.4	0.9	0.2	–0.9
45-49	–7.0	–2.9	2.3	3.4	3.5	0.7
50-54	–8.0	–1.3	1.8	3.1	3.7	0.7
55-59	–4.8	–1.6	0.6	5.7	1.8	–1.8
60-64	–3.9	–1.9	1.4	4.6	1.7	–1.9
65-69	–3.5	–2.4	0.4	6.6	2.7	–3.7
70-74	–5.9	–3.3	1.1	5.6	3.3	–0.9
75-79	–7.3	–2.8	1.3	6.2	3.3	–0.7
80-84	–5.6	–2.1	1.5	5.8	1.9	–1.4
85+	–7.6	–0.6	2.6	5.2	2.0	–1.8
Total	–3.0	–1.4	2.0	3.5	1.6	–2.7

Table A3.2. Percentage Points Change in Population Share, Females, 2010-2020

Age	Urban-east	Urban-central	Urban-west	Rural-east	Rural-central	Rural-west
0-1	–0.1	0.2	4.1	1.2	–2.4	–3.0
1-4	–2.9	–1.4	2.2	3.0	1.5	–2.4
5-9	–5.0	–1.6	1.3	2.5	4.8	–2.0
10-14	–7.4	–1.4	0.6	3.8	7.4	–3.0
15-19	–4.0	–2.3	4.3	2.1	0.5	–0.9

20-24	–1.7	0.1	5.2	0.4	–1.5	–2.5
25-29	–2.9	–0.9	2.0	3.3	1.9	–3.4
30-34	0.6	–1.6	1.3	3.6	1.0	–4.9
35-39	–0.2	–1.0	0.8	2.8	1.3	–3.6
40-44	–2.4	–1.6	3.5	0.8	0.7	–1.0
45-49	–7.4	–2.8	2.4	3.4	4.1	0.2
50-54	–9.3	–1.6	1.8	3.7	4.9	0.5
55-59	–4.7	–1.7	0.5	5.5	2.5	–2.1
60-64	–4.1	–2.1	1.3	4.2	2.4	–1.7
65-69	–4.6	–3.2	0.3	7.1	3.7	–3.2
70-74	–7.5	–3.2	1.2	6.1	3.9	–0.6
75-79	–7.0	–1.7	1.9	4.7	2.6	–0.7
80-84	–5.8	–1.9	1.7	5.6	1.6	–1.4
85+	–7.5	–0.7	2.7	5.3	1.4	–1.2
Total	–3.5	–1.6	2.0	3.5	2.1	–2.5

Change of population sizes and age structure

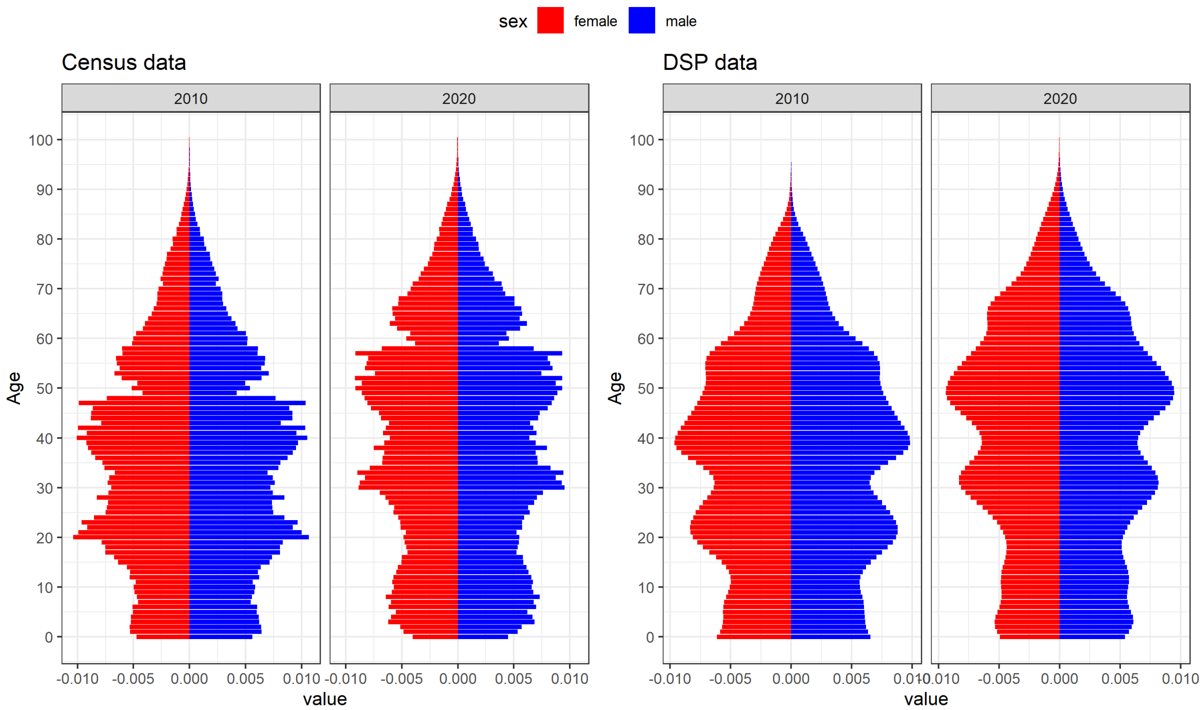
Figure A2 shows the comparisons of population size and age structure between Disease Surveillance Point (DSP 2010 & 2020) data and the data reported by the Chinese Bureau of Statistics (CBS 2010 & 2020). The age structure in the DSP dataset is younger than the age structure reported by the Chinese Bureau of Statistics (see table below). However, most of the changes in life expectancy in our analysis can be explained by ages 60 and up, which is fairly similar between the two data sources. In the decomposition results, for both females and males, the total contributions from ages 60 and up share a significant proportion of the total differences in life expectancy between 2010 and 2020 (72.8% for females and 61.7% for males).

Table A4. Summary of population pyramid by age and sex between DSP and CBS (Shown in Figure A2)

Data Source	**Age group**	**Female 2010**	**Male 2010**	**Female 2020**	**Male 2020**
CBS	<25	16.0%	17.7%	13.3%	15.1%
DSP	<25	15.7%	17.1%	12.1%	14.1%
CBS	25~59	26.0%	27.0%	25.8%	27.0%
DSP	25~59	26.7%	27.4%	27.0%	27.1%
CBS	60+	6.8%	6.5%	10.0%	9.0%
DSP	60+	6.8%	6.3%	10.2%	9.5%

Data source: Chinese Bureau of Statistics, 2010, 2020*;* Chinese Center for Disease Control and Prevention, 2010, 2020

Figure A2. Comparison between population counts given by the Chinese Bureau of Statistics (CBS) and the DSP data from the Chinese Center for Disease Control and Prevention (CCDCP).



Data source: Chinese Bureau of Statistics, 2010, 2020*;* Chinese Center for Disease Control and Prevention, 2010, 2020.

Appendix 3. Comparison of male life expectancy in 2020 with open age interval set at 90+ and 110+.

Table A5. Comparison of male life expectancy in 2020 with open age interval set at 85+, 90+ and 110+.

Region	Life expectancy from DSP raw data (85+ open age interval)	Life expectancy from adjusted data (90+ open age interval)	Life expectancy from adjusted data (110+ open age interval)
Urban-east	80.01	79.07	79.04
Rural-east	77.91	76.10	76.09
Urban-central	78.25	77.06	77.05
Rural-central	76.72	75.20	75.19
Urban-west	77.51	76.28	76.27
Rural-west	75.46	72.98	72.98

The results show a negligible difference between life expectancies with 90+ vs 110+ open age interval.

Appendix 4.

Table A6.1. Comparison between age-specific mortality rates from UN, China Population Census, DSP, and our models based on DSP, Males, 2010.

Year	Age	UN	China Population Census	DSP raw data	DSP log-quadratic model
2010	0-1	0.01382	0.00373	0.00629	0.00765
2010	1-4	0.00082	0.00069	0.00075	0.00041
2010	5-9	0.00062	0.00036	0.00033	0.00019
2010	10-14	0.00071	0.00037	0.00034	0.00020
2010	15-19	0.00071	0.00052	0.00049	0.00064
2010	20-24	0.00088	0.0007	0.00076	0.00087
2010	25-29	0.00101	0.00084	0.00085	0.00083
2010	30-34	0.00123	0.00111	0.00124	0.00093
2010	35-39	0.00168	0.00159	0.00161	0.00127
2010	40-44	0.00274	0.00237	0.00257	0.00195
2010	45-49	0.00388	0.0035	0.00380	0.00325
2010	50-54	0.00624	0.00548	0.00513	0.00550
2010	55-59	0.00888	0.00804	0.00801	0.00923
2010	60-64	0.01572	0.01302	0.01384	0.01531
2010	65-69	0.02504	0.02126	0.02156	0.02526
2010	70-74	0.04119	0.03702	0.03692	0.04161
2010	75-79	0.06521	0.05913	0.06288	0.06838
2010	80-84	0.10874	0.09856	0.11949	0.11118
2010	85+	0.22360	0.17683	0.37617	0.21418

Table A6.2. Comparison between age-specific mortality rates from UN, China Population Census, DSP, and our models based on DSP, Females, 2010.

Year	Age	UN	China Population Census	DSP raw data	DSP log-quadratic model
2010	0-1	0.01177	0.00392	0.00468	0.00556
2010	1-4	0.00081	0.00059	0.00053	0.00031
2010	5-9	0.00056	0.00023	0.00019	0.00010
2010	10-14	0.00048	0.00022	0.00018	0.00009
2010	15-19	0.00045	0.00025	0.00023	0.00016
2010	20-24	0.00055	0.0003	0.00033	0.00019
2010	25-29	0.00059	0.00037	0.00039	0.00022
2010	30-34	0.00072	0.0005	0.00051	0.00033
2010	35-39	0.00099	0.00071	0.00068	0.00055
2010	40-44	0.00148	0.00111	0.00110	0.00097
2010	45-49	0.00207	0.00168	0.00170	0.00169
2010	50-54	0.00314	0.00281	0.00242	0.00280
2010	55-59	0.00428	0.00429	0.00401	0.00446
2010	60-64	0.00833	0.00749	0.00730	0.00745
2010	65-69	0.01458	0.01306	0.01187	0.01257
2010	70-74	0.02565	0.02436	0.02180	0.02263
2010	75-79	0.0439	0.04089	0.03906	0.04151
2010	80-84	0.07879	0.07398	0.07756	0.07650
2010	85+	0.17960	0.15340	0.29075	0.17764

Table A6.3. Comparison between age-specific mortality rates from UN, China Population Census, DSP, and our models based on DSP, Males, 2020.

Year	Age	UN	China Population Census	DSP raw data	DSP log-quadratic model
2020	0-1	0.00652	0.00166	0.00629	0.00251
2020	1-4	0.00038	0.00028	0.00075	0.00019
2020	5-9	0.00029	0.00016	0.00033	0.00010
2020	10-14	0.00037	0.00021	0.00034	0.00012
2020	15-19	0.00046	0.00036	0.00049	0.00051
2020	20-24	0.00071	0.00044	0.00076	0.00078
2020	25-29	0.00092	0.00054	0.00085	0.00073
2020	30-34	0.00104	0.00071	0.00124	0.00082
2020	35-39	0.00128	0.0011	0.00161	0.00111
2020	40-44	0.00212	0.00178	0.00257	0.00173
2020	45-49	0.00336	0.00271	0.00380	0.00277
2020	50-54	0.00577	0.00414	0.00513	0.00449
2020	55-59	0.00827	0.00628	0.00801	0.00703
2020	60-64	0.01465	0.01023	0.01384	0.01146
2020	65-69	0.02287	0.01557	0.02156	0.01874
2020	70-74	0.03627	0.02579	0.03692	0.03166
2020	75-79	0.05776	0.04366	0.06288	0.05223
2020	80-84	0.10096	0.07422	0.11949	0.08742
2020	85+	0.21380	0.14729	0.37617	0.18458

Table A6.4. Comparison between age-specific mortality rates from UN, China Population Census, DSP, and our models based on DSP, Females, 2020.

Year	Age	UN	China Population Census	DSP raw data	DSP log-quadratic model
2020	0-1	0.0056	0.00137	0.00184	0.00209
2020	1-4	0.00039	0.00022	0.00022	0.00016
2020	5-9	0.00026	0.00012	0.00014	0.00006
2020	10-14	0.00024	0.00015	0.00018	0.00006
2020	15-19	0.00028	0.0002	0.00021	0.00012
2020	20-24	0.00045	0.00021	0.00021	0.00013
2020	25-29	0.00055	0.00023	0.00023	0.00015
2020	30-34	0.0006	0.0003	0.00033	0.00023
2020	35-39	0.00071	0.00044	0.00046	0.00039
2020	40-44	0.00105	0.00069	0.00072	0.00072
2020	45-49	0.00155	0.00114	0.00119	0.00125
2020	50-54	0.00257	0.00183	0.002	0.00211
2020	55-59	0.00373	0.00274	0.00307	0.00332
2020	60-64	0.00677	0.00469	0.00493	0.00543
2020	65-69	0.01123	0.008	0.00917	0.00872
2020	70-74	0.02007	0.01499	0.01755	0.01537
2020	75-79	0.03497	0.0284	0.03326	0.02792
2020	80-84	0.06699	0.05366	0.06338	0.05441
2020	85+	0.16716	0.1334	0.15463	0.15091

Appendix 5. Details of decomposition and discrete approximation

To disentangle the change over time in the national mortality into sub-group contributions, the age-specific death rate, or $m(x,t)$, is disaggregated into sub-national age-specific death rates weighted by the age-specific subnational composition:

$$m(x,t) = \frac{\sum_i D(x,i,t)}{P(x,t)} = \sum_i m(x,i,t)c(x,i,t) \, , \tag{A1}$$

where $P(x,t)$, $D(x,i,t)$ and $c(x,i,t) = \frac{P(x,i,t)}{P(x,t)}$ correspond to the person-years, deaths and proportion of the population at age *x* and time *t*, with i referring to the subnational population by location (urban or rural) and region (west, central or east).

Let a dot on top of a variable denote the change in a variable with respect to time t (Vaupel & Canudas-Romo, 2002). The age-specific population growth rate is the relative derivative of the population counts with respect to time, denoted as $r(x,t) = \frac{\dot{P}(x,t)}{P(x,t)}$. The derivative of $c(x,i,t)$ with respect to time can also be expressed as:

$$\dot{c}(x,i,t) = \left[\frac{\dot{P}(x,i,t)}{P(x,i,t)} - \frac{\dot{P}(x,t)}{P(x,t)}\right]\frac{P(x,i,t)}{P(x,t)} = [r(x,i,t) - r(x,t)]c(x,i,t), \tag{A2}$$

with $r(x,i,t)$ the region *i* age-specific growth rate at age *x* and time *t*.

The life expectancy at time t, denoted as $e_0(t)$, can be expressed in terms of age-specific death rates. This is under the assumption that mortality is constant within each age interval (or piecewise hazard statistically), the force of mortality is equal to the age-specific death rates, so the survival function is $\ell(a) = \mathrm{e}^{-\int_0^a m(x,t)\,dx}$ and in terms of the subnational variables as:

$$e_0(t) = \int_0^\omega \mathrm{e}^{-\int_0^a \sum_i m(x,i,t)c(x,i,t)\,dx}\, da. \tag{A3}$$

To study the changes in national life expectancy, we derive it with respect to time. The changes in national life expectancy are decomposed into the contribution attributable to mortality changes, denoted as $\Delta\, mortality$, and to regional population composition changes, or $\Delta\, composition$, as:

$$\dot{e}_0(t) = \int_0^\omega \dot{\ell}(x,t)dx \tag{A4}$$

$$= -\int_0^{\omega} \ell(x,t) \int_0^{x} \sum_i [\dot{m}(a,i,t)c(a,i,t)] \; dadx$$
$$-\int_0^{\omega} \ell(x,t) \int_0^{x} \sum_i [m(a,i,t)[r(a,i,t)-r(a,t)]c(a,i,t)] \; dadx$$
$$= \Delta\, mortality + \Delta\, composition. \qquad \text{(A5)}$$

While $\Delta\, mortality = -\int_0^{\omega} \ell(x,t) \int_0^{x} \sum_i [\dot{m}(a,i,t)c(a,i,t)] \; dadx$ can be interpreted as changes in life expectancy due to the changes in the risk of death, or $\dot{m}(x,i)$, $\Delta\, composition = -\int_0^{\omega} \ell(x,t) \int_0^{x} \sum_i [m(a,i,t)[r(a,i,t)-r(a,t)]c(a,i,t)] \; dadx$ can be interpreted as the contribution to changes in life expectancy from the changes in exposures. $\Delta\, composition$ contrasts the values between national and subnational population growth rates ($[r(x,i) - r(x)]$); with positive or negative contributions to the change in the national life expectancy, depending on lower growth rate than national ($[r(x,i) - r(x)] < 0$) or higher ($[r(x,i) - r(x)] > 0$), and a corresponding change in the population exposure to mortality.

The terms in Eq. (A5) are analogous to the derivations presented by Torres et al. (2019). However, the latter group apply the reversal of integration in Eq. (A4) and use the definition of life expectancy at age a, or $e_a(t) = \frac{\int_a^{\omega} \ell(x,t)\, dx}{\ell(a,t)}$, to obtain the following equation:

$$\dot{e}_0(t)$$
$$= -\int_0^{\omega} \dot{m}(a,t) \int_a^{\omega} \ell(x,t)\, dx\, da$$
$$= -\int_0^{\omega} \dot{m}(a,t)\ell(a,t)e_a(t)da.$$
$$\text{(A6)}$$

Similarly to our substitution in Eq. (A5), further substituting the term $\dot{m}(x,t)$ in Eq. (A6) aids Torres et al. (2019) at separating the two components of $\Delta\, mortality$ and $\Delta\, composition$. This clearly shows the analogy between the two approaches in Eq. (A5 & A6). We use our derivations in Eq. (A5), since they include one term less compared to Torres et al. (2019) terms, namely life expectancy.

We use discrete approximations for the mathematical equations presented (see e.g. Vaupel & Canudas-Romo, 2002; Vaupel & Canudas-Romo, 2003). If data are available between time t and time $t + h$, with the assumption of the constant change over a certain time interval, the formula writes:

$$\dot{v}(t+\tfrac{h}{2}) \approx \frac{\ln\left[\frac{v(t+h)}{v(t)}\right]}{h} v\left(t+\frac{h}{2}\right).$$

The mid-point function is expressed as:

$$v\left(t+\frac{h}{2}\right) \approx [v(t)v(t+h)]^{\frac{1}{2}}.$$

In some cases it seems more appropriate to assume that the change in the time interval was linear (such as $l(x)$), then we use:

$$v\left(t+\frac{h}{2}\right) \approx \frac{v(t+h)+v(t)}{2}.$$

Table A7. Illustration of discrete calculations of contributions of ages 0-5 for urban east male.

\	$\dot{m}(a,i,t)$	c(a,i,t)	m(a,i,t)	r(a,i,t)−r(a,t)	ℓ(x,t)	*Δ mortality*	*Δ composition*	Age Contributions
0	–0.0029	0.1338	0.0044	–0.0052	0.989	0.00039	0.00000	0.00039
1	–0.0003	0.1343	0.0006	–0.1263	0.987	0.00043	0.00001	0.00045
2	–0.0001	0.1343	0.0003	–0.2258	0.987	0.00045	0.00002	0.00048
3	–0.0001	0.1338	0.0002	–0.2953	0.986	0.00047	0.00003	0.00050
4	–0.0001	0.1326	0.0002	–0.3313	0.986	0.00048	0.00004	0.00052
5	–0.0001	0.1312	0.0001	–0.3450	0.986	0.00049	0.00005	0.00053
					…			
85	–0.0195	0.2269	0.1307	–0.2735	0.204	0.01021	0.01542	0.02563
86	–0.0210	0.2278	0.1450	–0.2975	0.173	0.00948	0.01478	0.02427
87	–0.0224	0.2282	0.1604	–0.3244	0.145	0.00865	0.01405	0.02270
88	–0.0236	0.2281	0.1770	–0.3517	0.119	0.00773	0.01321	0.02094
89	–0.0247	0.2276	0.1946	–0.3776	0.096	0.00676	0.01223	0.01900
					…			
110+	–0.0172	0.1806	0.6993	–0.1291	0	0	0	0

Notes:

1. $\dot{m}(a,i,t)$ = change in death rate at age *a*, region *i* and time *t*

2. $c(a,i,t)$ = population composition at age *a*, region *i* and time *t*

3. $m(a,i,t)$ = death rate at age *a*, region *i* and time *t*

4. $r(a,i,t) - r(a,t)$ = difference in growth rates between region *i* and national growth, at age *a*, and time *t*

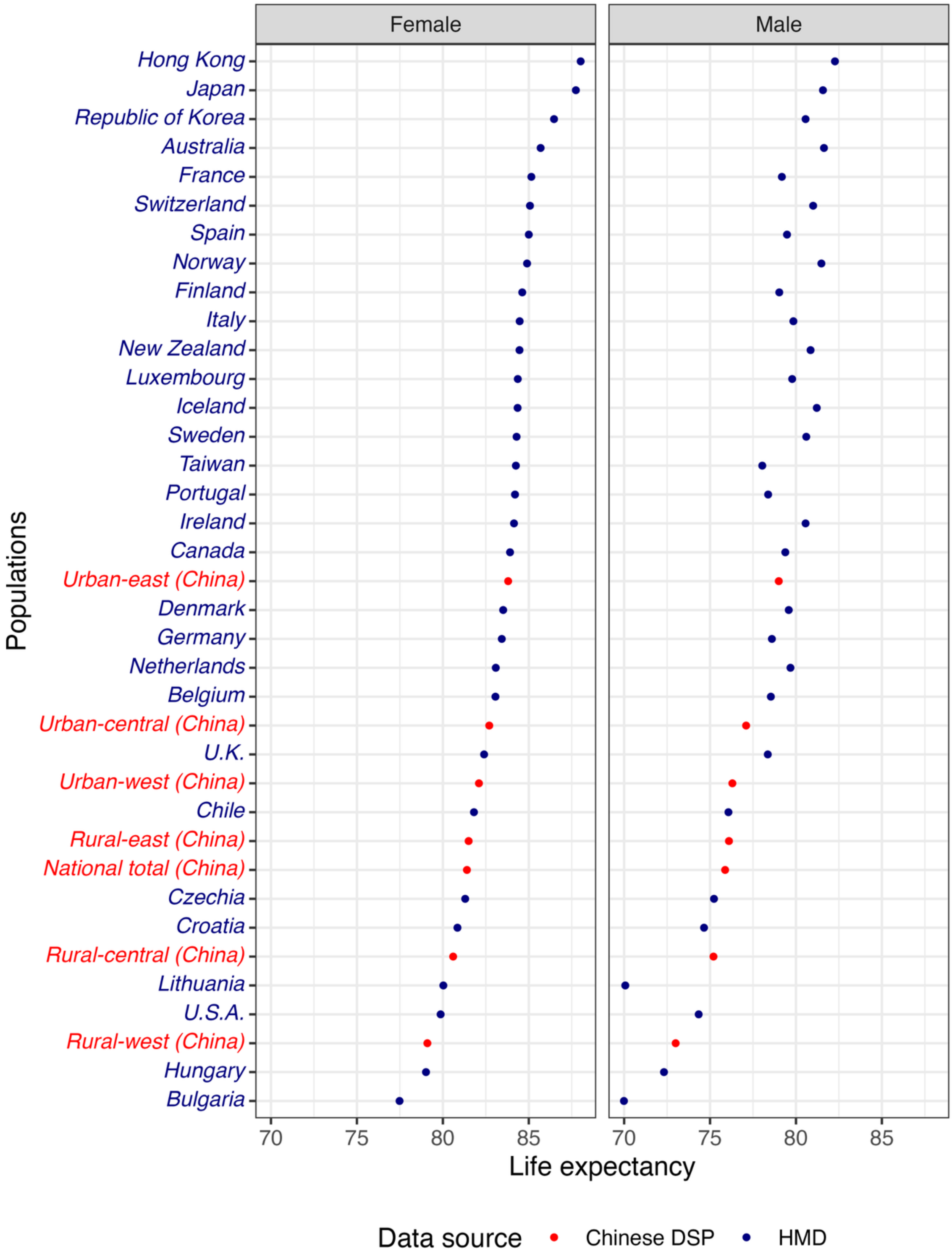
5. $\ell(x,t)$ = life table survival function at age *a* and time *t*

6. $\Delta\ mortality$ = mortality component

7. $\Delta\ composition$ = composition component

8. Age Contributions

Appendix 6. Regional life expectancies in China in comparison with populations included in the HMD, 2020.



Source: *Authors' calculation based on 2020 DSP data and the HMD 2023.*

Appendix 7. Confidence interval for the decomposition results

Table A8. Decomposition results with confidence intervals, females 2010-2020.

Decomposition of the Chinese female life expectancy (e_0) from 2010 (78.8 with 95% CI: 78.8, 78.9), to 2020 (81.4 CI 81.4, 81.4), into mortality and composition components, by region and rural/urban.			
Region	$\Delta\, mortality$	$\Delta\, composition$	$\Delta\, total$
urban–west	0.16 (0.16, 0.16)	–0.16 (–0.16, –0.16)	–0.001 (–0.005, 0.003)
rural-west	0.88 (0.87, 0.90)	0.20 (0.20, 0.21)	1.09 (1.07, 1.10)
urban-central	0.15 (0.15, 0.16)	0.14 (0.14, 0.14)	0.29 (0.29, 0.30)
rural-central	0.53 (0.52, 0.54)	–0.23 (–0.23, –0.23)	0.30 (0.29, 0.32)
urban-east	0.32 (0.31, 0.32)	0.43 (0.43, 0.43)	0.75 (0.74, 0.76)
rural-east	0.57 (0.56, 0.58)	–0.47 (–0.47. –0.47)	0.10 (0.09, 0.12)
Column sums	2.62 (2.56, 2.67)	–0.08 (–0.09, –0.08)	2.54 (2.48, 2.59)

Table A8. Decomposition results with confidence intervals, males 2010-2020.

Decomposition of the Chinese male life expectancy (e_0) from 2010 (73.2 with 95% CI: 73.1, 73.2), to 2020 (75.9 CI 75.9, 75.9), into mortality and composition components, by region and rural/urban.			
Region	$\Delta\, mortality$	$\Delta\, composition$	$\Delta\, total$
urban-west	0.14 (0.13, 0.14)	–0.17 (–0.17, –0.17)	–0.03 (0.03, 0.03)
rural-west	0.74 (0.73, 0.76)	0.29 (0.29, 0.29)	1.03 (1.02, 1.05)
urban-central	0.18 (0.18, 0.19)	0.19 (0.18, 0.19)	0.37 (0.36, 0.37)
rural-central	0.67 (0.65, 0.68)	–0.25 (–0.25, –0.25)	0.42 (0.41, 0.43)
urban-east	0.36 (0.35, 0.36)	0.42 (0.42, 0.42)	0.77 (0.77, 0.78)
rural-east	0.64 (0.62, 0.65)	–0.53 (–0.53, –0.53)	0.11 (0.09, 0.12)
Column sums	2.73 (2.67, 2.78)	–0.05 (–0.06, –0.05)	2.67 (2.62, 2.73)

*Source: Authors' calculation based on Chinese diseases surveillance points data, 2010 & 2020. Results are rounded to two decimals. The 95% confidence intervals are calculated using Monte Carlo life table bootstrap methods with 1,000 iterations ((*Silcocks et al., 2001)*). That is, we generated 1000 mortality age schedules using the binomial distribution and used those results to calculate decomposition results for 1000 times.*

Appendix 8. Regional contribution curves

Figure A8a. Differences in growth rates (national - regional), by region and age.

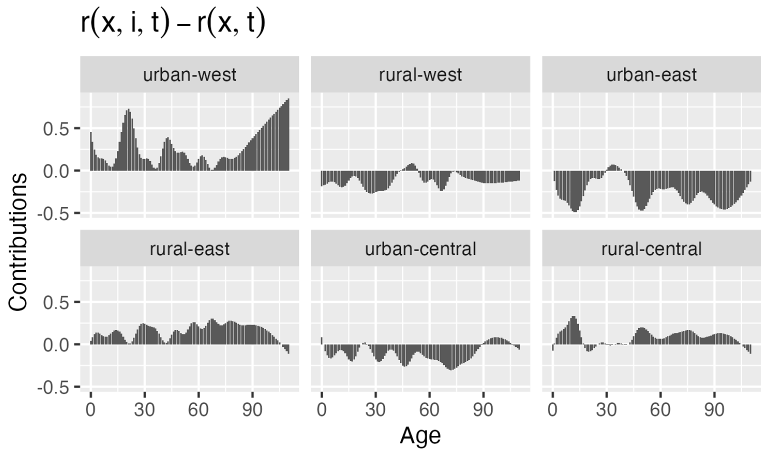


Figure A8b. Product of mortality by population composition, by region and age.

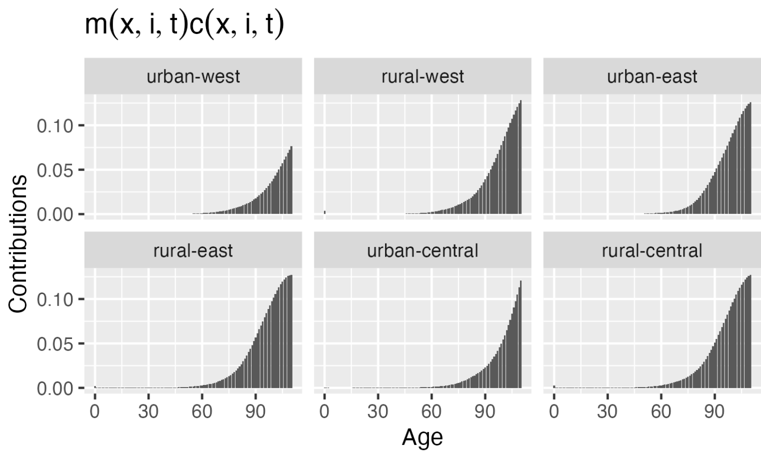


Figure A8c. Product of the difference in growth rates, death rate and population composition (two plots above), by region and age.

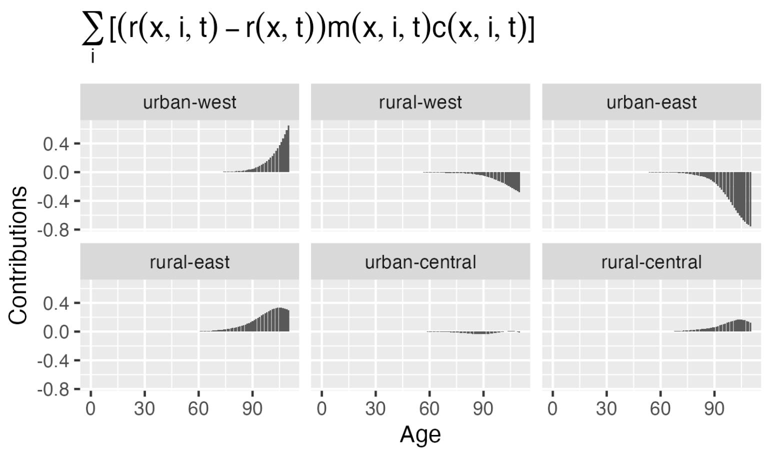
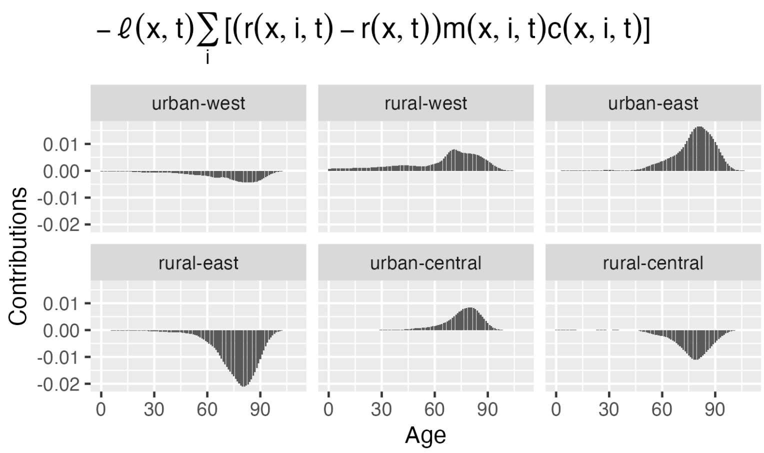


Figure A8d. Regional composition component, by region and age.



Note: This final figure highlights that the important values are those between ages 60 and 90.

Appendix 9. Comparison of contributions to life expectancy changes for males during 2010-2020 with a hypothetical 10% population change from rural-east to urban-east.

Table A9. Comparing contributions to life expectancy changes for males during 2010-2020 with a hypothetical 10% population change from rural-east to urban-east.

Components	Regions	Hypothetical Scenario	Observed Values
Mortality	urban-east	0.38	0.36
	rural-east	0.60	0.64
	urban-central	0.18	0.18
	rural-central	0.67	0.67
	urban-west	0.14	0.14
	rural-west	0.74	0.74
Composition	urban-east	0.21	0.42
	rural-east	–0.25	–0.53
	urban-central	0.19	0.19
	rural-central	–0.25	–0.25
	urban-west	–0.17	–0.17
	rural-west	0.29	0.29
Total Regional	urban-east	0.59	0.78
	rural-east	0.35	0.11
	urban-central	0.37	0.37
	rural-central	0.42	0.42
	urban-west	–0.03	–0.03
	rural-west	1.03	1.03
Total National	total contributions	2.74	2.70

As observed in the table in the hypothetical scenario, the regional contribution to male life expectancy from the urban-east decreased from the observed 0.78 to 0.59 and the regional contribution from the rural-east increased from 0.11 to 0.35 after a move of 10% male population from rural-east to urban-east. This is consistent with our argument that an increased exposure to mortality will lead to a reduction to life expectancy contribution and also signifies the importance of population movement to life expectancy increase.

References

Chinese Bureau of Statistics. (2010). *Chinese Statistical Yearbook*. http://www.stats.gov.cn/tjsj/ndsj/2007/indexch.htm

Chinese Bureau of Statistics. (2020). *China Statistical Yearbook*. http:/www.stats.gov.cn/tjsj/ndsj/2017/indexch.htm

Chinese Center for Disease Control and Prevention. (2010). *National disease surveillance points – cause of death surveillance database.* Military Medical Science Press.

Chinese Center for Disease Control and Prevention. (2020). *National disease surveillance points – cause of death surveillance database.* Military Medical Science Press.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de

National Health Commission. (2020). *China National Maternal and Child Health Surveillance (MCHS): Child mortality.*

Pascariu, M. D., Dańko, M. J., Schöley, J., & Rizzi, S. (2018). "ungroup": An R package for efficient estimation of smooth distributions from coarsely binned data. *Journal of Open Source Software*, *3*(29), 937. https://doi.org/10.21105/joss.00937

Silcocks, P. B. S., Jenner, D. A., & Reza, R. (2001). Life expectancy as a summary of mortality in a population: Statistical considerations and suitability for use by health authorities. *Journal of Epidemiology & Community Health*, *55*(1), 38–43. https://doi.org/10.1136/jech.55.1.38

Torres, C., Canudas-Romo, V., & Oeppen, J. (2019). The contribution of urbanization to changes in life expectancy in Scotland, 1861–1910. *Population Studies*, *0*(0), 1–18. https://doi.org/10.1080/00324728.2018.1549746

Vaupel, J. W., & Canudas-Romo, V. (2002). Decomposing demographic change into direct vs. Compositional components. *Demographic Research*, *7*, 1–14. https://doi.org/10.4054/DemRes.2002.7.1

Wilmoth, J., Zureick, S., Canudas-Romo, V., Inoue, M., & Sawyer, C. (2012). A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies*, *66*(1), 1–28. https://doi.org/10.1080/00324728.2011.611411