

Supplementary materials to “The widening rural-urban gap in life expectancy since COVID-19 ”

Shasha Han, PhD ^{1,2,3*#} Binbin Su, PhD ^{1*} Yihao Zhao, PhD ¹ Chen Chen, PhD ¹ Xiaoying Zheng, PhD ^{1#}

1. School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China
2. State Key Laboratory of Respiratory Health and Multimorbidity, Beijing, China
3. Key Laboratory of Pathogen Infection Prevention and Control (Peking Union Medical College), Ministry of Education, Beijing, China

* These authors contributed equally.

Corresponding Authors:

Shasha Han, PhD, Assistant Prof.

School of Population Medicine and Public Health

Chinese Academy of Medical Sciences & Peking Union Medical College

Beijing, P. R. China. 100730

hanshasha@pumc.edu.cn

Xiaoying Zheng, PhD, Prof.

School of Population Medicine and Public Health

Chinese Academy of Medical Sciences & Peking Union Medical College

Beijing, P. R. China. 100730

zhengxiaoying@sph.pumc.edu.cn

Technical details

We calculated life expectancy from 1987 to 2021 using annual period life tables via standard demographic techniques¹ for rural and urban residences separately. All-age rural-urban gaps were estimated as the life expectancies of the urban population minus those of the rural population. Using the Arriaga decomposition technique,² we attributed annual changes in the gaps to changes in age-specific mortality, and attributed annual gaps to all-cause mortality. Additionally, we calculated excess rural-urban gaps for 2003, 2010, 2020, and 2021 defined as observed gap minus expected gap based on a continuation of outbreak trends. We estimated the expected rural-urban gap using the Lee–Carter forecasts³ of age- and sex-specific mortality over the five years before the start of the outbreaks.

We describe the Lee-Carter (LC) method and Arriaga decomposition method below.

Let $m_{x,t}$ be the central death rate at age x in year t . The LC model can be expressed as:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}$$

where a_x is the general shape of mortality by age, and k is a time index of the general level of mortality. The parameter b_x describes the extent to which mortality at age x changes given the overall temporal change in the general level of mortality: greater values of b_x are associated with faster mortality change. The error term $\varepsilon_{x,t}$ reflects residual age-specific temporal influences not captured by the model, and is assumed to have zero mean and variance σ_ε^2 . By modelling logged rates, the fitted and forecast rates (obtained by exponentiation) are constrained to be positive. The logarithmic transformation also serves to counter the heteroscedasticity in the observed rates.

Let $e_0^u - e_0^r$ denote the expectation of life at birth for rural population u and urban population r , e_x^i expectation of life at age x in population i . l_x^i number of survivors to age x in population i , ${}_nL_x^i$ person years of life contributed in the age group $(x, x + n)$ in population i . Arriaga decomposition formulates the differences in the expectation of life at birth $e_0^u - e_0^r$ due to

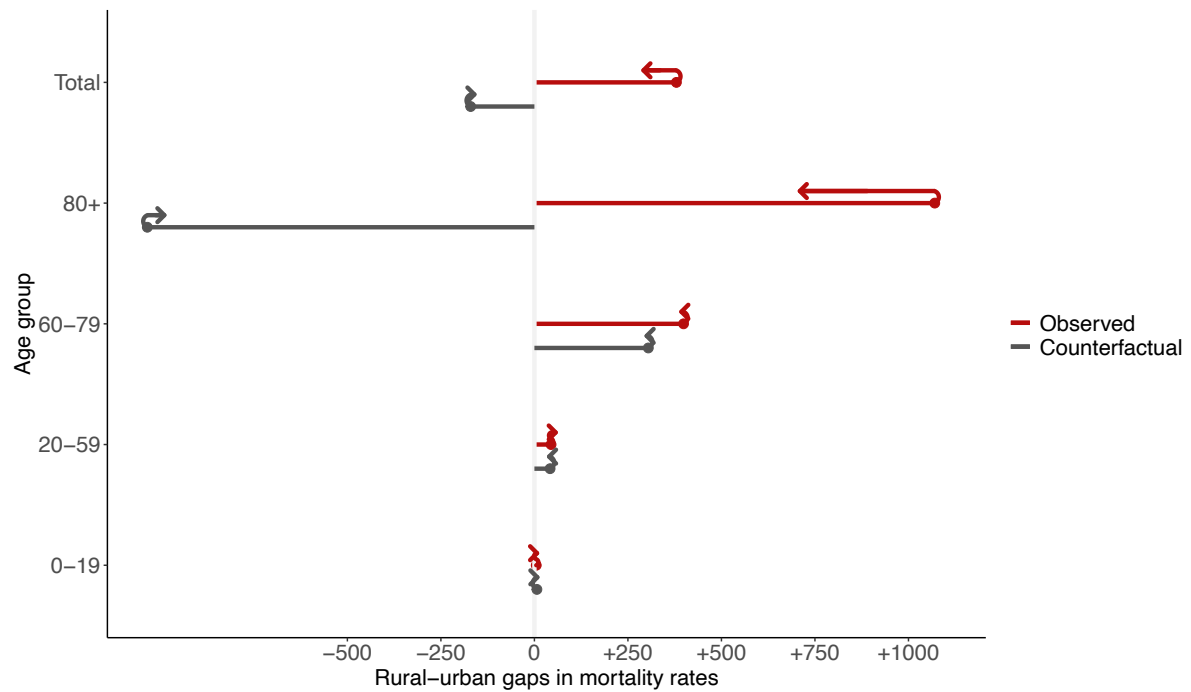
differences in age-specific death rates as the sum of direct effect of age x (DE_x) and the indirect effect of age x , where the direct effect (DE_x) of age group $(x, x + n)$ is

$$DE_x = l_x^u \left(\frac{nL_x^r}{l_x^r} - \frac{nL_x^u}{l_x^u} \right)$$

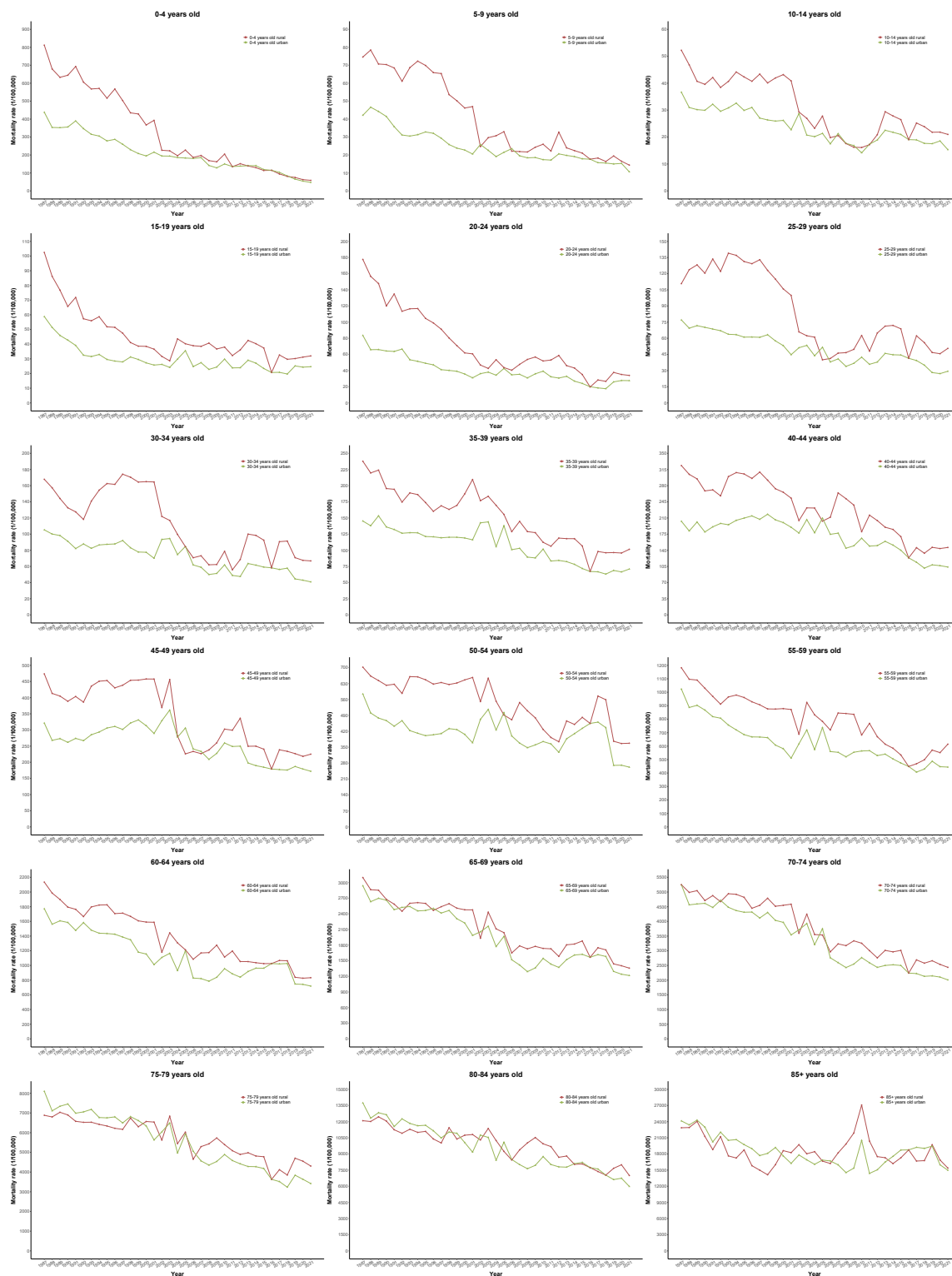
and the indirect effect (IDE_x) of age group $(x, x + n)$ is

$$IDE_x = \left(l_x^u \frac{l_{x+n}^r}{l_x^r} - l_{x+n}^u \right) e_{x+n}^r.$$

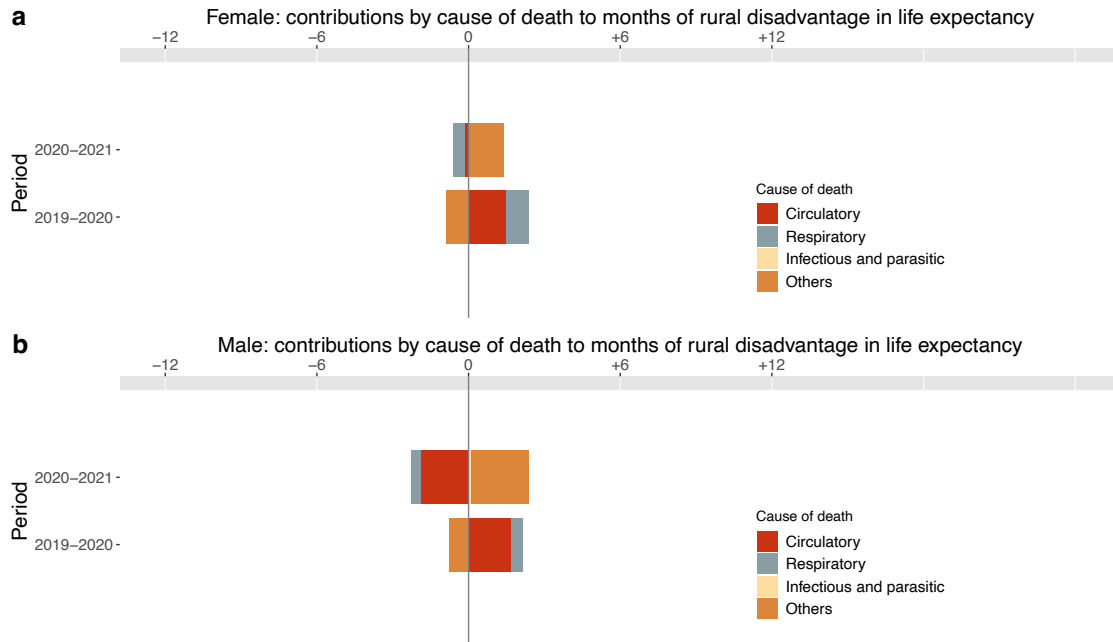
We further decompose the age-specific effects can into the age- and cause-specific effects using the weights, which are calculated as the proportion of the difference in the cause-specific mortality rates as a share of the total mortality difference in the specific age interval.



eFigure 1 Rural-urban gap in mortality rates during the pandemic. The position of the point indicates the rural-urban gap in mortality rates in a given age group, defined as the mortality rates of rural populations minus that of urban populations, in 2020. The position of the arrowhead indicates the gap in 2021. Red lines represent the gap in the real-world scenario, and the grey lines the gap in the counterfactual scenarios without pandemics. Negative values suggest a diminished rural-urban gap in the age group.



eFigure 2 Mortality rates in China from 1987 to 2021, broken down by five-year age group.



eFigure 3 Contributions by cause of death to changes in months of rural-urban gaps in life expectancy since 1987 by sex, using the contour method.

Extended Data Table 1 Years of change in life expectancy by geographical area

Year	Rural mean	Rural 95% CI	Urban mean	Urban 95% CI
2021	+8.4	(+5.6, +11.9)	+9.7	(+6.1, +12.8)
2020	+10.5	(+7.8, +13.6)	+14.8	(+11.3, +17.8)
2019	-6.3	(-9.1, -3.4)	+5.3	(+2.3, +8.0)
2018	+3.9	(+1.2, +7.0)	+5.3	(+2.4, +8.4)
2017	-5.5	(-8.8, -3.2)	+0.6	(-2.8, +3.6)
2016	+15.8	(+12.8, +18.7)	+5.4	(+3.0, +8.1)
2015	-1.4	(-4.2, +1.0)	-1.2	(-3.5, +2.0)
2014	+7.0	(+4.3, +10.1)	-4.4	(-7.0, -1.8)
2013	-2.0	(-4.4, +0.0)	-8.9	(-12.4, -6.2)
2012	+10.7	(+8.0, +13.0)	+1.0	(-2.0, +5.2)
2011	+14.1	(+12.4, +16.7)	+26.0	(+23.0, +28.3)
2010	-1.6	(-3.6, +0.4)	-25.0	(-27.9, -22.5)
2009	-6.3	(-8.6, -3.6)	-7.8	(-11.0, -4.1)
2008	-1.7	(-4.2, +0.3)	+15.3	(+12.3, +18.5)
2007	-17.6	(-20.5, -14.5)	+7.2	(+4.8, +10.0)
2006	+18.7	(+15.6, +20.8)	+26.8	(+24.0, +29.3)
2005	+8.4	(+6.2, +10.7)	-21.3	(-23.3, -18.6)
2004	+17.4	(+15.8, +19.2)	+26.5	(+24.4, +29.0)
2003	-15.8	(-18.2, -14.0)	-3.9	(-6.4, -1.7)
2002	+29.9	(+28.2, +32.3)	-14.4	(-17.0, -12.2)
2001	-0.6	(-2.6, +1.9)	+14.1	(+12.0, +16.3)
2000	-3.2	(-5.1, -1.3)	+7.9	(+5.8, +10.2)
1999	+1.4	(-1.0, +3.9)	+4.8	(+2.5, +6.9)
1998	+3.0	(+0.4, +5.0)	-2.4	(-5.3, -0.7)
1997	+3.9	(+1.8, +6.5)	+7.5	(+5.8, +9.7)
1996	+9.5	(+7.8, +11.4)	+1.1	(-0.8, +3.2)
1995	+1.7	(-0.1, +4.2)	+3.3	(+1.7, +5.3)
1994	-0.2	(-2.3, +1.9)	+2.8	(+1.1, +5.0)
1993	-1.5	(-3.4, +0.8)	+6.5	(+4.8, +8.2)

1992	+7.5	(+5.2, +9.5)	-3.1	(-5.0, -1.2)
1991	+2.1	(-0.2, +4.5)	+7.3	(+5.6, +9.3)
1990	+9.5	(+7.5, +11.3)	+3.3	(+1.7, +4.9)
1989	+2.4	(+0.6, +4.3)	-2.9	(-4.5, -1.6)
1988	+11.3	(-7.1, -3.2)	+18.2	(+14.9, +18.4)

Extended Data Table 2 Contributions by age to months of excess life expectancy in recent epidemics by geographical area

Year	Age				
	group	Rural mean	Rural 95% CI	Urban mean	Urban 95% CI
2021	0-19	+0.5	(+0.3, +0.7)	+0.3	(+0.1, +0.5)
2021	20-59	-1.3	(-1.9, -0.4)	+1.1	(+0.5, +1.7)
2021	60-79	-0.2	(-1.9, +1.7)	+2.8	(+1.4, +4.4)
2021	80+	+8.2	(+5.2, +11.3)	+20.6	(+16.3, +24.9)
2021	Total	+7.1	(+3.6, +11.3)	+24.9	(+20.4, +29.1)
2020	0-19	+0.4	(+0.3, +0.6)	+0.1	(-0.1, +0.3)
2020	20-59	+1.0	(+0.5, +1.7)	+1.4	(+1.0, +1.9)
2020	60-79	-0.4	(-1.7, +1.3)	+2.1	(+0.9, +3.4)
2020	80+	-0.7	(-4.0, +1.8)	+12.9	(+8.8, +15.7)
2020	Total	+0.3	(-3.4, +4.1)	+16.5	(+12.4, +20.6)
2010	0-19	-2.3	(-2.5, -1.9)	-1.0	(-1.2, -0.6)
2010	20-59	+2.4	(+1.6, +3.2)	-6.1	(-6.7, -5.6)
2010	60-79	-2.4	(-3.8, -1.0)	-12.8	(-14.3, -11.8)
2010	80+	-10.0	(-11.9, -8.1)	-26.9	(-30.9, -23.1)
2010	Total	-12.2	(-14.6, -9.7)	-46.9	(-51.3, -43.1)
2003	0-19	+8.7	(-0.7, +12.6)	+0.1	(-0.4, +0.7)
2003	20-59	+3.2	(+1.3, +4.4)	-10.1	(-10.7, -9.5)
2003	60-79	+2.1	(-0.7, +3.8)	-12.5	(-13.9, -10.7)
2003	80+	-0.2	(-6.4, +13.7)	-8.0	(-10.7, -5.1)
2003	Total	+13.8	(+11.4, +17.5)	-30.4	(-34.3, -26.3)

Extended Data Table 3 Contributions by age to months of the excess rural-urban gap in recent epidemics

Year	Age group	Female mean	Female 95% CI	Male mean	Male 95% CI
2003	0-19	-9.9	(-11.8, -3.7)	-3.8	(-12.6, -10.2)
2003	20-59	-12.7	(-13.8, -11.3)	-14.7	(-15.8, -13.7)
2003	60-79	-13.1	(-17.2, -13.4)	-14.5	(-16.4, -12.2)
2003	80+	-12.0	(-25.8, -8.4)	-2.8	(-5.4, -0.8)
2003	Total	-47.7	(-57.4, -47.4)	-35.8	(-46.6, -39.8)
2010	0-19	+0.4	(+0.2, +1.0)	+1.6	(+1.4, +2.3)
2010	20-59	-7.8	(-8.5, -7.2)	-9.9	(-11.1, -8.9)
2010	60-79	-6.9	(-8.5, -5.0)	-17.2	(-19.6, -15.1)
2010	80+	-12.6	(-17.7, -8.3)	-7.3	(-9.2, -5.0)
2010	Total	-27.0	(-32.4, -21.6)	-32.8	(-36.4, -28.7)
2020	0-19	+0.3	(-0.0, +0.4)	+0.3	(+0.0, +0.5)
2020	20-59	-0.8	(-1.4, -0.2)	-1.3	(-2.4, -0.0)
2020	60-79	+1.9	(-0.3, +3.3)	+3.3	(+1.2, +5.5)
2020	80+	+18.9	(+12.3, +24.2)	+10.8	(+7.0, +14.9)
2020	Total	+20.2	(+12.9, +26.3)	+13.1	(+8.4, +17.8)
2021	0-19	+0.5	(+0.2, +0.7)	+0.8	(+0.4, +1.0)
2021	20-59	+0.1	(-0.6, +0.7)	+0.2	(-1.1, +1.4)
2021	60-79	+2.7	(+0.4, +4.6)	+3.4	(+1.0, +6.1)
2021	80+	+18.8	(+12.1, +27.1)	+9.6	(+5.0, +13.5)
2021	Total	+22.1	(+15.3, +29.5)	+14.0	(+8.2, +18.8)

Extended Data Table 4 Months of rural-urban gaps in life expectancy since 1987

Year	Population		Female			
	mean	Population 95% CI	mean	Female 95% CI	Male mean	Male 95% CI
2021	+23.7	(+19.6, +26.7)	+21.2	(+17.4, +26.1)	+24.9	(+21.9, +27.0)
2020	+22.7	(+19.6, +25.8)	+20.4	(+16.1, +24.4)	+24.5	(+21.9, +26.8)
2019	+19.2	(+16.4, +22.2)	+16.5	(+12.7, +19.3)	+21.4	(+19.7, +23.9)
2018	+7.6	(+5.0, +10.5)	+2.0	(-1.3, +5.3)	+12.5	(+10.3, +14.9)
2017	+6.4	(+3.5, +9.1)	+0.3	(-2.8, +4.3)	+11.5	(+8.6, +13.5)
2016	-0.0	(-2.6, +3.2)	+0.6	(-3.9, +4.1)	+11.5	(+9.1, +14.2)
2015	+10.3	(+7.8, +13.1)	+7.1	(+3.1, +9.9)	+12.6	(+10.3, +15.3)
2014	+10.0	(+6.4, +13.0)	+4.4	(+1.6, +7.5)	+14.3	(+11.9, +16.4)
2013	+21.2	(+18.9, +24.4)	+18.0	(+14.5, +21.2)	+23.2	(+21.0, +25.2)
2012	+27.8	(+25.5, +31.3)	+24.3	(+20.0, +27.4)	+30.3	(+28.3, +32.7)
2011	+36.9	(+33.8, +39.8)	+32.6	(+29.1, +36.2)	+38.7	(+36.3, +40.9)
2010	+25.9	(+23.8, +27.9)	+21.8	(+19.4, +24.9)	+29.0	(+27.2, +30.6)
2009	+47.1	(+44.7, +49.8)	+42.2	(+39.0, +45.3)	+49.5	(+47.0, +51.2)
2008	+48.5	(+46.1, +50.9)	+42.0	(+39.2, +45.5)	+52.7	(+50.4, +54.7)
2007	+32.2	(+29.6, +34.6)	+29.9	(+27.6, +33.2)	+33.0	(+31.1, +34.9)
2006	+7.8	(+5.6, +10.6)	+3.9	(+1.1, +7.8)	+11.7	(+9.5, +13.7)
2005	-0.2	(-2.4, +2.6)	+8.5	(+7.7, +13.5)	-13.5	(-14.0, -9.9)
2004	+29.4	(+27.1, +31.7)	+26.5	(+23.5, +29.5)	+30.8	(+28.9, +32.5)
2003	+20.0	(+17.9, +21.9)	+20.5	(+13.5, +18.8)	+19.8	(+18.0, +21.3)
2002	+8.3	(+5.8, +10.4)	+8.9	(+5.8, +11.1)	+8.4	(+21.9, +25.4)
2001	+51.7	(+49.8, +54.6)	+55.2	(+52.9, +57.7)	+46.1	(+49.9, +53.6)
2000	+37.3	(+35.4, +39.3)	+36.8	(+34.0, +39.3)	+37.0	(+35.3, +38.8)
1999	+26.4	(+24.4, +28.7)	+25.1	(+23.0, +27.8)	+27.0	(+24.8, +28.6)
1998	+23.3	(+21.0, +25.3)	+5.6	(+3.1, +8.6)	+37.7	(+36.7, +40.2)
1997	+28.1	(+26.2, +30.4)	+28.6	(+25.8, +31.8)	+27.4	(+24.9, +29.5)
1996	+24.6	(+22.6, +26.9)	+20.4	(+17.5, +22.0)	+28.6	(+26.5, +30.4)
1995	+33.2	(+31.2, +35.1)	+29.2	(+26.5, +30.9)	+36.4	(+34.4, +38.3)
1994	+31.7	(+29.7, +33.4)	+27.9	(+25.8, +30.1)	+35.2	(+46.9, +50.7)

1993	+28.7	(+27.0, +30.2)	+24.4	(+22.1, +26.7)	+32.2	(+30.1, +33.8)
1992	+20.5	(+17.9, +22.4)	+19.8	(+17.8, +21.6)	+32.3	(+30.2, +33.6)
1991	+31.0	(+28.8, +33.0)	+29.3	(+27.0, +31.8)	+32.8	(+31.0, +34.8)
1990	+25.9	(+23.7, +27.8)	+22.5	(+20.7, +24.5)	+28.6	(+24.6, +28.0)
1989	+32.2	(+30.1, +34.1)	+30.6	(+28.9, +33.3)	+34.0	(+32.8, +35.8)
1988	+37.5	(+35.9, +39.3)	+37.6	(+35.2, +39.5)	+38.4	(+38.1, +41.8)
1987	+30.6	(+14.1, +17.9)	+27.0	(+25.2, +29.1)	+52.6	(+21.9, +26.2)

Extended Data Table 5 Regression results on the rural-urban gap in life expectancy by the gap in public health expenditure

	Estimate	Standard error	t value	Pr(> t)
Age < 60				
(Intercept)	16.726	0.911	18.370	< 2e-16
Rural-urban gaps in public health expenditure	1.549	0.512	3.028	0.003
Year	-0.501	0.048	-10.338	1.64e-15
Multiple R-squared: 0.616,		Adjusted R-squared: 0.604		
Age ≥ 60				
(Intercept)	0.325	1.483	0.219	0.827
Rural-urban gaps in public health expenditure	0.836	2.248	0.373	0.711
Year	0.199	0.076	2.570	0.012
Multiple R-squared: 0.107,		Adjusted R-squared: 0.080		

References:

1. Chiang CL. *Life Table and Mortality Analysis*. World Health Organization; 1978.
2. Arriaga EE. Measuring and explaining the change in life expectancies. *Demography*. 1984;21(1):83-96.
3. Lee RD, Carter LR. Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*. 2012;87(419):659-671.