JOURNAL OF SOCIAL COMPUTING
ISSN 2688-5255 05/06 pp250−261
Volume 3, Number 3, September 2022
DOI: 10.23919/JSC.2022.0009

Why Hong Kong Ranks Highest in Life Expectancy: Looking for Answers from Data Science and Social Sciences

Ting Xu, Ming-sum Tsui, and Dah Ming Chiu*

Abstract: In trying to explain why Hong Kong of China ranks highest in life expectancy in the world, we review what various experts are hypothesizing, and how data science methods may be used to provide more evidence-based conclusions. While more data become available, we find some data analysis studies were too simplistic, while others too overwhelming in answering this challenging question. We find the approach that analyzes life expectancy related data (mortality causes and rate for different cohorts) inspiring, and use this approach to study a carefully selected set of targets for comparison. In discussing the factors that matter, we argue that it is more reasonable to try to identify a set of factors that *together* explain the phenomenon.

Key words: life expectancy; data science; social science; population

1 Introduction

Life expectancy is a well-accepted standard and simple metric to measure the public health of a society. Based on the ranking and trend, it is always interesting to ask how well a given society has done. This is particularly valid for those societies in top (or bottom) positions. Hong Kong as a Special Administration Region of China often enjoys the privilege of being ranked and compared to different countries and regions in the world. After fast improvements in the past 50 years, Hong Kong's life expectancy has managed to top the world ranking (85.16 in 2022)[1].

But why is Hong Kong of China doing so well? This is an interesting question that prompted some coverage and commentary in the news[2–4], social media[5, 6], and more serious discussions and studies by various stakeholders, such as academics conducting research on government policies and population aging[7–9], government departments[10], and even insurance industry[11].

Of course, there are a variety of possible reasons and factors. Some are directly medical in nature, such as obesity (BMI) level and smoking rate; some are public health policy related, such as expenditures of health care and elderly care. Other factors can be environmental, social, and cultural. The authors of the CNN Health Story report[2] interviewed two academics (in health and medicine) and an expert from WHO‡. They offered the following reasons to explain the good performance: (1) easy to get around; (2) green environment; (3) efficient public hospital care; (4) physical and psychological fitness (as 70% of elderly are immigrants from Mainland of China who had to overcome various difficulties); (5) good subtropical climate; (6) healthy diet (similar to Mediterranean countries); (7) intensive family care and respect (culturally Asians are closer to their parents). While these factors are interesting and

• Ting Xu is with the Department of Information Engineering, Chinese University of Hong Kong, Hong Kong 999077, China. E-mail: 1155162643@link.cuhk.edu.hk.
• Ming-sum Tsui is with the Felizberta Lo Padilla Tong School of Social Sciences, Caritas Institute of Higher Education, Hong Kong 999077, China. E-mail: mstsui@cihe.edu.hk.
• Dah Ming Chiu is with the Department of Information Engineering, Chinese University of Hong Kong, and also with Data Science Research Center, Caritas Institute of Higher Education, Hong Kong 999077, China. E-mail: dmchiu@cihe.edu.hk.
∗ To whom correspondence should be addressed.
Manuscript received: 2022-04-30; revised: 2022-06-21; accepted: 2022-08-13

‡Professor Timothy Kwok of Chinese University of Hong Kong and director of Jockey Club for positive ageing, Professor Gabriel Leung of University of Hong Kong and dean of the Li Ka Shing Faculty of Medicine, and Dr. John Beard, director of the Department of Aging and Life course at the World Health Organization.

quite plausible, and subsequently reported and discussed by others[5, 6], they can only be taken as hypothetical without more evidence, preferably based on data analysis.

A more recent academic commentary article[7] addressed the same question. The authors discussed the factors suggested in the CNN report and gave some counter arguments about their significance. They suggested their own list of potential factors correlated to life expectancy: (1) GDP per capita; (2) total health care expenditure (as a percentage of GDP); (3) Gini coefficient; (4) infant mortality rate; (5) adolescent birth rate; (6) percentage of youth not in education/job; (7) homicide rate; (8) incarceration rate. Most of these factors come from previous academic studies on life expectancy. In discussing the likely contribution of these factors, the authors compared the performance of Hong Kong of China, with that of UK and USA, two familiar societies with comparable per capita GDP. They found factors (4)−(8) to have consistent correlation to life expectancy in this small comparison group. However, Hong Kong of China has achieved the highest life expectancy with significantly less expenditure on (2) health care expenditure and (3) a significantly higher Gini coefficient (meaning higher income disparity), which is puzzling. They thus concluded that the question is worthy of further studies.

One of the co-authors of our paper conducted a pilot study by carefully interviewing four categories of 20 stakeholders in Hong Kong of China, including (1) five elderly service users, (2) five informal caregivers, (3) five helping professionals, and (4) five research experts in different areas. Twelve possible factors were identified: (1) natural selection: many old adults were refugees with super physical strength and will power, (2) effective emergency rescue service within a short physical distance, (3) no serious natural disaster in the last 60 years, (4) healthy diet and hard working attitude, (5) integrative use of modern western medical technology and traditional Chinese medicines, (6) good care for pregnant women, (7) high level of income per capita, (8) respect and care for the senior adults, (9) use of domestic helpers, (10) advantage in comparing a metropolitan city with other countries and regions with rural villages, (11) good post-natal care, and (12) upbeat life style. Although the number of interviews is relatively small, this is a step towards collecting data methodically.

In contrast to the above three studies that explored the social, environmental, and government policy related factors that may have contributed to Hong Kong's life longevity, another recent paper[8] did a very throughout study of the cause-of-death statistics in Hong Kong of China and its correlation to life expectancy. Based on their analysis, they concluded that lower smoking rate is the single most significant contributing factor to Hong Kong's high life expectancy ranking. This study has been covered by popular press[4] as well. A couple of other articles[9, 10] also did data analysis related to life expectancy in Hong Kong of China, towards better understanding of potential reasons, trends, and issues. This is also our approach, and we will explain their method and discuss their results in more detail in the next section when we review various data analytical methods used to study life expectancy.

An important objective of this paper is to explore how data analytical methods can be adopted to identify important factors that contribute to a society's life expectancy ranking, particularly in Hong Kong, China. Data analytical methods are usually considered evidence-based, hence more credible. There are still serious challenges since some of the factors are hard to quantify and related data may not be easily measurable or available. In this paper, we first review the methods based on previous works in Section 2. We then explain our own approach of selected comparisons and decomposition; and discuss some results in Section 3. Finally, we discuss the implications of the findings and conclude in Section 4.

2 Review of Data Analytic Methods for Life Expectancy Studies

2.1 Preston Curve

There has been plentiful research on life expectancy and its contributing factors. Most serious studies tried to draw evidence based on data. One of the classic results is the Preston Curve[12] as mentioned by Ref. [7]. The original version[12] was published in 1975. The famous plot in the paper, reproduced here as Fig. 1, shows two inverse hockey-stick curves relating life expectancy (LE) to per capita Gross Domestic Product (pGDP) of a nation or region; one curve is for 1930 and

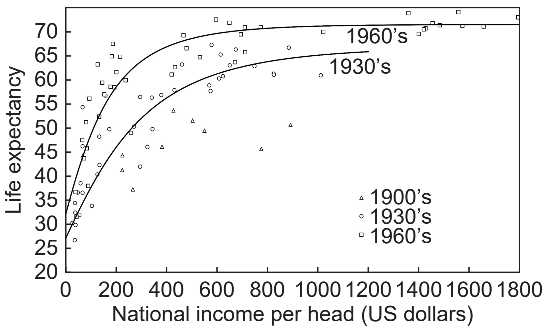


Fig. 1 Preston Curve (regenerated using the data in the original (1975) Preston paper), Scatter-diagram of relations between life expectancy at birth and national income per head for nations in the 1900's, 1930's, and 1960's.

the other for 1960. The curves are basically of the same shape, with the 1960 curve shifted higher, indicating significantly increased LE around the world over those 30 years. In fact, a few data points are also included for 1900, but a curve was not drawn due to lack of more data points. The simple message of this plot is to show that LE is correlated to pGDP, but the correlation is not linear. When pGDP is higher than some threshold, the correlation diminishes. While this finding is clear and frequently cited, a major part of this paper is spent on addressing another question, that is, to what extent pGDP contributes to explain LE. This is the most challenging key issue.

From the point of view of predicting the LE of a society given its pGDP at a given year, the fitted curves for 1930 and 1960 give performance of 80% and 84.7%, respectively. But this is not the same as the contribution of pGDP towards LE. Historically, it is known that the advances in medical sciences and technology, especially the successfully eradication of various epidemic diseases, have contributed greatly to LE. While pGDP helps medical research and deployment of medical resources, it seems far-fetched to let pGDP represent the contribution from medicine towards LE. The paper mentioned Hong Kong of China as an example of a population achieving high LE (in 1966) without correspondingly high standing in pGDP at that time. The paper then tried a logical analysis of using the 1930 model to predict 1960's LE, for the limited cases there are data, and found that pGDP can only explain 16% of the improvement of LE between 1930 and 1960.

The example of Preston Curve paper (Section 2.1) touched on various ways of applying data science to LE. Based on our understanding, we categorize them into four approaches: data visualization (Section 2.2), correlation and multivariate regression analysis (Section 2.3), data analysis leading to hypothesis (Section 2.4), and causal inference and model building (Section 2.5). We will briefly explain and discuss each and give some references.

2.2 Data visualization

The Preston Curve illustrates how a figure can capture and summarize important messages from data, though sparse. At this time, data have become much more abundant. You can go to the relevant international organizations, such as UN[13] and WHO[14] and find related data. Another excellent source is Our World in Data[15]. For example, you can easily compare the LE of Hong Kong, China with that of Japan (JP), Switzerland (SW), Singapore (SP), UK, and US, from 1950 to 2019 as shown in Fig. 2. While SW, UK, and US ranked far higher in 1950, Hong Kong of China, JP, SW, and SP dominated by 2019. The trend and differences are clearly illustrated for a researcher to think about possible reasons.

For the case of Hong Kong of China, there are also some websites that keep and help you explore life expectancy related data and metrics, such as Macrotrends[16], AgeWatch[17], and HK Elder Quality of Life Index[18]. Besides providing data, these sites usually provide an interface that allow you to visualize how life expectancy changed over time, how it compares between any combination of countries and regions, and how it correlates to various factors. An important approach in data science is to explore various ideas and insights through visualization. Data visualization is also often the step leading to deeper analysis and modeling.

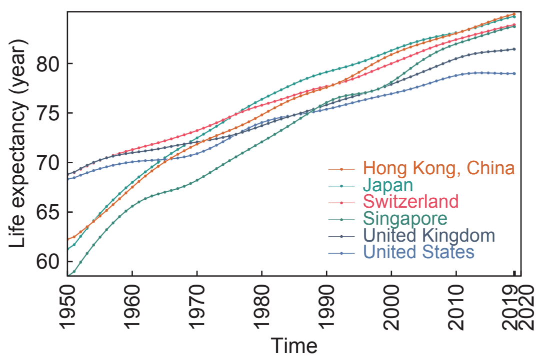


Fig. 2 Data visualization of life expectancy (1950 to 2019) from Our World in Data.

2.3 Correlation and multivariate regression analysis

Probably the majority of publications, especially in medical science and social sciences studying factors related to life expectancy use this approach. The field of data science (whether statistics and computer science) provides abundant tools for studying the correlation between one (dependent) variable and one or more other (independent) variables. Conceptually, there is no difference whether there is a single factor (independent variable), or a multiple of them. As demonstrated by the Preston Curve study, the correlation between pGDP and LE can be established by curve fitting, given multiple case samples relating pGDP and LE. If, besides pGDP, there are other factors X, Y, and Z in the dataset, the fitted curve would be a multi-dimensional function of LE $= f(\text{pGDP}, X, Y, Z)$. Such curve-fitting can be achieved by a machine learning algorithm, arriving at a classification mechanism, which is in fact a process that gives you a prediction of LE given input pGDP, X, Y, and Z.

Technically, there are lots of machinery to establish the goodness of the fitted curve, in terms of the accuracy of prediction for different possible values of the factors, and how much confidence you should have with the fitting based on only a limited sample of the space of factor values. A review of these methods is beyond the scope of our discussion here. But all these methods are within the realm of data science, without the need to involve any domain experts, such as medical scientists or social researchers.

In fact, applying multivariate regression analysis to find out the correlation between various factors to LE is used as an example to learn data science. There is a dataset derived from WHO's LE data[14] at Kaggle[19], a popular website for people to learn and share related information for data science projects. It gives the LE of 193 countries and regions over 16 years (2000 to 2015), each comes with corresponding values of a set of factors of different varieties: medical (e.g., BMI), social (e.g., population), and policy (e.g., schooling). Various tools (open-source code) for analyzing such data can also be found on this and related websites. Based on such a dataset and standard multivariate linear regression analysis, results are reported in blogs such as Ref. [20] on Medium (a social media publishing platform). While such studies can help illustrate how easily standard tools can be applied to perform multivariate linear regression analysis, it is hard to draw convincing conclusions based on the findings. There are several problems with this correlation-only approach: (1) the factors included in the dataset may not be comprehensive. The resultant relative importance for these factors can change drastically when other important factors are added to the study. (2) As discussed in the Preston paper, the accuracy for such multivariate regression only characterizes the predictability of LE (under some statistical assumptions). It cannot fully reflect the importance of different factors, or even the causal relationship of these factors with LE.

Academic studies of correlation of various factors to LE typically go further. For example, it[21] also considers a dataset involving multiple societies, but tries to divide the data into different groups, i.e., developing countries and regions and developed countries and regions, and discuss the results based on the importance of different factors for these groups. Another paper[22] considered LE data for different provinces of China and draw conclusions based on the geographical correlation of the factors. The approach of these studies is still basically correlation analysis, and has the same limitations mentioned above.

2.4 Data analysis leading to hypothesis

In this approach, one does not start with any factor(s) and study its correlation or contribution to LE. Rather, relevant data are first analyzed to better understand LE, in terms of components contributing to LE, such as mortality rates (for different age groups), causes of mortality and so on, and try to hypothesize reasons based on your data analysis.

A good example of this approach is Ref. [8]. In this case, the authors were able to obtain vast amount of data for the cause of death for specific cohorts of a population, and for many developed countries and regions (OECD, Republic of Korea, and Singapore), as well as Hong Kong, China. Based on these data, and the known attribution of smoking to mortality of various diseases, it is possible to derive the reduced rate of mortality in Hong Kong of China attributable to smoking. On another front, the authors also obtained data on smoking rates in Hong Kong of China compared to the comparison group over the years. By putting these evidences together, Ref. [8] made a strong case that smoking is a major factor for Hong Kong's advantage in LE. In comparison to the earlier studies on economic (pGDP) and social policy (education) factors, this study is more convincing. This approach would involve more domain experts, especially in the hypothesizing step.

Other examples of this approach include a recent article "The Mortality Trend in Hong Kong, 1986 to 2018", published by the Census and Statistics Department of Hong Kong Special Administrative Region Government as part of their monthly report[10]. The report analyzed the trend of mortality rate for different age groups, which are the components LE based on. Besides the trend, they also analyzed the mortality causes together with Hong Kong's LE standing, in a similar fashion as Ref. [8]. Although the report did not suggest reasons for Hong Kong's ranking, their analysis certainly provides some useful information for hypothesizing reasons.

Another paper[9] also analyzed Hong Kong's LE by decomposing it into the age-specific mortality rates. This decomposition allows the authors to discuss another metric called life disparity (variation in longevity), and compare Hong Kong of China with Japan in these two dimensions. It turns out that this decomposition method and its variations have been used since 1980s to understand trends in LE over a period of time[23]. Another interesting paper from Max-Planck Institute for Demographic Research used this approach to do a comparison between US and UK[24].

2.5 Causal inference and model building

Beyond correlation, statisticians and social scientists certainly also pay attention to causality. Given two variables that are correlated, does the first variable cause the behavior of the second variable, or vice versa? A humorous example illustrating two variables that are correlated but do not have causal relationship is the simultaneous growth in height of a boy and the tree in his yard (planted when he was born). By a similar logic, when two factors are shown to be correlated to a dependent variable, is one of the two factors caused by the other factor, hence not the real cause for the dependent variable? Such causal inference is very important for ensuring correct interpretation of correlation relationships. In the context of LE studies, besides pGDP (income), another factor highly correlated to LE is education. In questioning which of these two factors is causal, the authors of Ref. [25] asked, "if, for instance, the empirical association between income and health is not directly causal but rather due to a third factor such as education, then an increase in income, e.g., through policies directly aiming at economic growth would not result in the expected health improvements unless educational attainment also improved simultaneously". The method for doing causal inference usually involves setting up a null hypothesis (that one factor is the true cause), and then using experimentation and rigorous statistical methods to rule out likelihood for alternative hypothesis to be true. This approach is always the most challenging. Most of the time, it cannot be conducted in the real world as it is unethical to remove the needy services from the control group. In this case, we need more advice from the domain experts.

Ideally, the ultimate understanding we chase after is a model that explains how a bunch of factors cause certain result to occur. Such goal may be plausible for simple processes. It quickly becomes overwhelming for a complicated process like life expectancy for a particular population, when each person's life span can depend on myriad of things in a complicated way, though the aggregate behavior (of a population) can arguably be easier to characterize statistically than that of a simple person. Causal inference can be viewed as a small step towards the long journey of the model building efforts. In comparing the evaluation of medical factors and social factors affecting the life span of population, it is often possible to provide more domain specific evidence in the former case than the latter, due to the differences in physical sciences versus social sciences. Nonetheless, social factors are sometimes more complicated and dynamic than physical factors.

In conclusion, we reviewed and categorized different data science methods applicable to studying the factors contributing to LE as a concept, from simple and easy to complicated and ambitious. Despite more abundant data that may become available, and innovations in methods to analyze data, the research question remains very important, interesting, and inspiring.

3 A Targeted Comparison Study

In this section, we report some data analysis for Hong Kong's LE, namely we do not start with specific factors and try to evaluate their contribution to LE, but rather present some targeted comparison for coming up with potential factors.

We do our targeted comparison in two steps: (1) come up with a few societies to compare Hong Kong of China with; (2) decompose the life expectancy metric into mortality rates (for different age groups) and compare on that basis. These are explained in the following two subsections.

Overview of our approach: first narrow down the societies for comparison, and then do the decomposition of life expectancy at age zero into a series of mortality rates and compare the latter. Based on the comparison, speculate of the factors with the experience of the research team. The authors try to argue that this is an approach better than previous approach of speculating without data analysis.

3.1 Selecting targets for comparison

Hong Kong of China is a city. It seems only fair to compare Hong Kong of China with other cities doing well in LE to see Hong Kong of China still stands out. Another reason for comparing with other metropolitan cities is that they are likely to be similar to Hong Kong of China in many dimensions, so that the comparison can better help us identify the areas of difference, whether Hong Kong of China is doing better or worse than its counterparts in these areas.

The reasoning is similar to that adopted in Refs. [21, 22], though they fixed the factor for comparison. In Ref. [21], they separated the countries and regions for comparison into two groups, developing vs developed, and tried to observe interesting differences in the factors that are important to each group. In Ref. [22], they compared provinces in China, noticing the correlations between neighboring provinces and used that to score the importance of factors.

While searching for data and studies of LE at the city level, we noticed an interesting phenomenon, that is high LE of a city often has a geographical impact, causing near-by towns and areas also have high LE. We call this regional radiant effect of LE. For example, according to a study of LE of different income groups in the US[26], the authors found that the cities with higher LE in the US are concentrated in California. Among the top 40 cities of highest LE in US in 2019, California accounted for 14, followed by Florida, Hawaii, New York, and their surrounding areas. In a comprehensive study of relationship of personal income levels to LE, Ref. [27] shows that lower income individuals whose city/town is near an affluent and high LE city have comparably higher LE, demonstrating such regional radiant effect.

Another study examined LE of 40+ cities in China[28]. They found the cities in the Yangtze River Delta (close to Shanghai), Bohai Bay (close to Beijing), and Pearl River Delta (close to Guangzhou, and Shenzhen) all have similarly high LE as Shanghai, Beijing, Guangzhou, and Shenzhen, even other characteristics (such as GDP and population density) may be different. This again demonstrates the regional radiant effect.

The regional radiant effect means it is reasonable to compare Hong Kong of China with small countries and regions with large metropolitan cities, as most areas of such small countries and regions would enjoy similar LE as the metropolitan cities there.

By looking around the world, we found a set of candidate cities in Table 1[29–35].

Among the cities in Table 1, we first exclude some that have really small population size (e.g., 1/60 of Hong Kong of China), which are resort-like places for the rich, very different than a metropolitan city like Hong Kong of China. Secondly, for those more metropolitan cities in this list, we note Hong Kong of China still comes out first among them. Because of the regional radiant effect, we decided to choose three other populations for comparison in the next subsection: Singapore, Japan, and Switzerland. The historical trends for all these places can be found in Fig. 2.

3.2 Decomposition

Although life expectancy is a useful metric to gauge the health situation for different countries and regions, it is hardly possible to represent the full story with a single number. One way to dissect the implication of this number is to use the decomposition method as shown in Ref. [9]. Let us first review how life expectancy is calculated.

In a population, each group of people born in the same time period is referred to as a cohort. For simplicity, let us assume the group of people born in the same year

Table 1 Some possible cities for comparison.

City	Life expectancy	Per capita GDP (US dollar)	Population size	Area (km^2)	Population density (per km^2)	Remark
Hong Kong, China (data during 2015−2020)	84.64	48 354	7 577 629	1106.34	6849.00	
Naples-Immokalee-Marco Island, Florida	86.60	76 025	384 902	5970.00	64.47	USA (data from 2019)
San Jose-Sunnyvale Santa Clara, California	84.55	130 865	1 990 660	466.10	4270.89	
San Francisco-Oakland-Hayward, California	83.37	114 696	4 731 803	6397.30	739.66	
Community of Madrid, Spain	85.20	39 843	6 661 949	8030.10	829.62	Europe (data during 2017−2019)
Ticino, Switzerland	85.00	95 664	350 986	2812.21	120.00	
Nagano	84.72	36 289	2 052 493	13 561.56	150.00	Japan (data during 2015)
Shiga	84.68	40 062	1 412 916	4017.38	350.00	
Fukui	84.41	34 921	265 904	536.41	495.71	
Kyoto	84.38	38 567	1 475 000	827.80	1781.83	
Kumamoto	84.36	30 290	740 822	390.30	1898.08	
Okayama	84.35	36 529	719 474	789.90	910.84	
Nara	84.31	24 970	360 310	276.80	1301.70	
Kanagawa	84.28	32 348	9 200 000	2416.00	3807.95	
Shimane	84.22	33 167	665 205	6708.00	99.17	
Hiroshima	84.21	38 493	1 194 000	906.70	1316.86	
Ōita	84.20	34 329	478 146	501.20	954.00	
Tokyo	84.17	68 776	14 043 239	2194.07	6363.00	
Shanghai, China	83.67	24 700	24 870 895	6341.00	3900.00	Chinese Mainland (data from 2020)
Beijing, China	82.43	25 823	21 893 095	16 410.50	1334.09	
Seoul, Republic of Korea	84.10	38 489	9 776 000	605.20	16 153.34	Asia and other (data during 2017−2019)
Gyeonggi, Republic of Korea	83.10	30 998	13 410 000	10 717.00	1318.45	
Singapore	83.39	65 640	5 704 000	728.60	7828.71	
Sydney, Australia	84.20	86 500	5 312 000	12 368.00	429.50	
Macao, China	84.05	86 117	640 446	115.30	5554.61	

forms a cohort, and people live up to 100 years, so we have a total of $n = 100$ cohorts. Based on the current (recent) size and death rate of the cohorts[♯], we derive the mortality rate of each cohort. For cohort x, the cohort size is denoted by $l(x)$ and the probability of dying in year $(x, x+1)$ is denoted by $q(x)$ respectively. For each cohort x, it is then possible to calculate the total person-year lived by this cohort from age i to $i+1$, denoted $L(i)$, as

$$L(i) = l(i+1) + 0.5q(i)l(i) \tag{1}$$

for $i = x, x+1, ..., n$. Note, the second term is the total person-years accumulated by those who died during year $(i, i+1)$, so on average each died person is assumed to accumulate only half year of life. The total person-years cohort x is expected to accumulate until everyone from cohort x has died, denoted by $T(x)$, is

$$T(x) = \sum_{i=x}^{n} L(i) \tag{2}$$

The life expectancy of cohort x, denoted $e(x)$, is thus given by

$$e(x) = \frac{T(x)}{l(x)} \tag{3}$$

Note, this way of calculating life expectancy assumes the mortality rate stays the same in future years. But since there is no sure way to predict future mortality rates, this is the best we can do. A somewhat theoretical explanation of this method of calculation can be found in Ref. [36], and a practical explanation of how the calculation is done by a life table can be found in Ref. [37].

For a whole population, when we speak of life

[♯]The time period used to define a cohort can differ, and of course some people can live beyond 100 years old. How to handle these and other details have standard treatments, and do not affect our explanations.

expectancy, we are referring to $e(0)$, the life expectancy of the cohort currently at age 0. But from the above derivation of how $e(0)$ is calculated, we can see $e(0)$ can be decomposed as

$$e(0) = C(0) + C(1) + C(2) + \cdots + C(100) \tag{4}$$

where

$$C(i) = \frac{L(i)}{T(0)} e(0) \tag{5}$$

$C(i)$ is the contribution to $e(0)$ from cohort i.

Once we decompose life expectancy into contributions from different cohorts, when we compare the life expectancy of two populations, we can compare them on the basis of contribution from different cohorts of the populations as well.

In our case, when we compare Hong Kong of China and Japan (HKC-JP), Hong Kong of China and Switzerland (HKC-SW), and Hong Kong of China and Singapore (HKC-SP), we express the differences in terms of the contributions from the corresponding cohorts of each pair of countries and regions, namely, $C_{\text{HKC-JP}}(i)$, $C_{\text{HKC-SW}}(i)$, and $C_{\text{HKC-SP}}(i)$, and plot them in Fig. 3. Note, we have chosen to let each cohort come from a larger age range, namely, [0], [1, 24], [25, 59], [60, 74], [75, 84], and [85+]. Also, since there are significant differences between male and female, they are compared separately.

In all three comparisons, Hong Kong of China enjoys a positive gap at this moment, but the recent trends have been somewhat different. In the HKC-JP case, Hong Kong of China has been behind until very recently; in the HKC-SW case, Hong Kong of China had a fluctuating case for male, but increasing positive gap for female. In the HKC-SP case, Hong Kong of China enjoyed a relatively larger gap, but it is gradually shrinking recently.

In all cases, the mortality rate advantage of the 85+ group contributed to the gap. In three of the cases, HKC-SW female, HKC-SP male and female, Hong Kong's gap is contributed by mortality rates for 75−84, and 60−74 as well. But for two cases, HKC-JP female and HKC-JP male, Hong Kong's 75−84 mortality rate contributed negatively, which indicates more disparity among the elderly cohort arriving at the advanced age. Noticeably for all cases, Hong Kong's mortality rate for the 25−59 range is a secondary contributing factor; except for HKC-SW male, the contribution of 25−59 mortality is more significant.

Although the above analysis did not identify factors for explaining why Hong Kong of China is leading the world in life expectancy, it does provide evidence for making some hypotheses, which may be useful for setting the agenda for further studies.

4 Discussion and Conclusion

While the research question in the title of this paper is so innocent and inviting, the answer can be infinitely complicated. There are so many factors that can affect people's life span, be it environmental, social or medical, and so many ways their effects can be combined to create a result or trend we observe. So what answers would be expected? We believe it is unrealistic to expect anything comprehensive and thorough, but rather anything that helps peel off the complexity and lead us to some insights and ability to see the trends, and focus on practical things we can do to improve the situation, all based on a solid and convincing methodology. And this is indeed what we learned from published studies of a similar question in the literature.

The decomposition method leads us to think of factors contributing to long life expectancy more from a cohort point of view. The cohort of people aged above 85 in recent years were all born before 1930. What factors in their life experience, be that medical, environmental, or sociopolitical, would help them enjoy comparatively lower mortality rate today? There are surely many factors, but by carefully selecting the target for comparison, we have narrowed down the range of factors to those contributing to the gap in the comparison, which should match our curiosity the most.

As suggested by Ref. [8], less smoking in previous life is certainly an important factor keeping the 85+ years old in Hong Kong of China relatively healthy. But by this factor alone, would it be sufficient to explain the life longevity of this cohort? We would like to bring up a couple of other factors, that we believe are also necessary. First, most of this age 85+ cohort immigrated from Mainland of China in their younger days (reportedly more than 70% of this group). These immigrants may be more physically and psychologically fit and resilient than ordinary, as suggested by (d) in the CNN report[2], as well as the commentary of Ref. [7]. Second, Hong Kong's elderly care may also have been a necessary factor. Hong

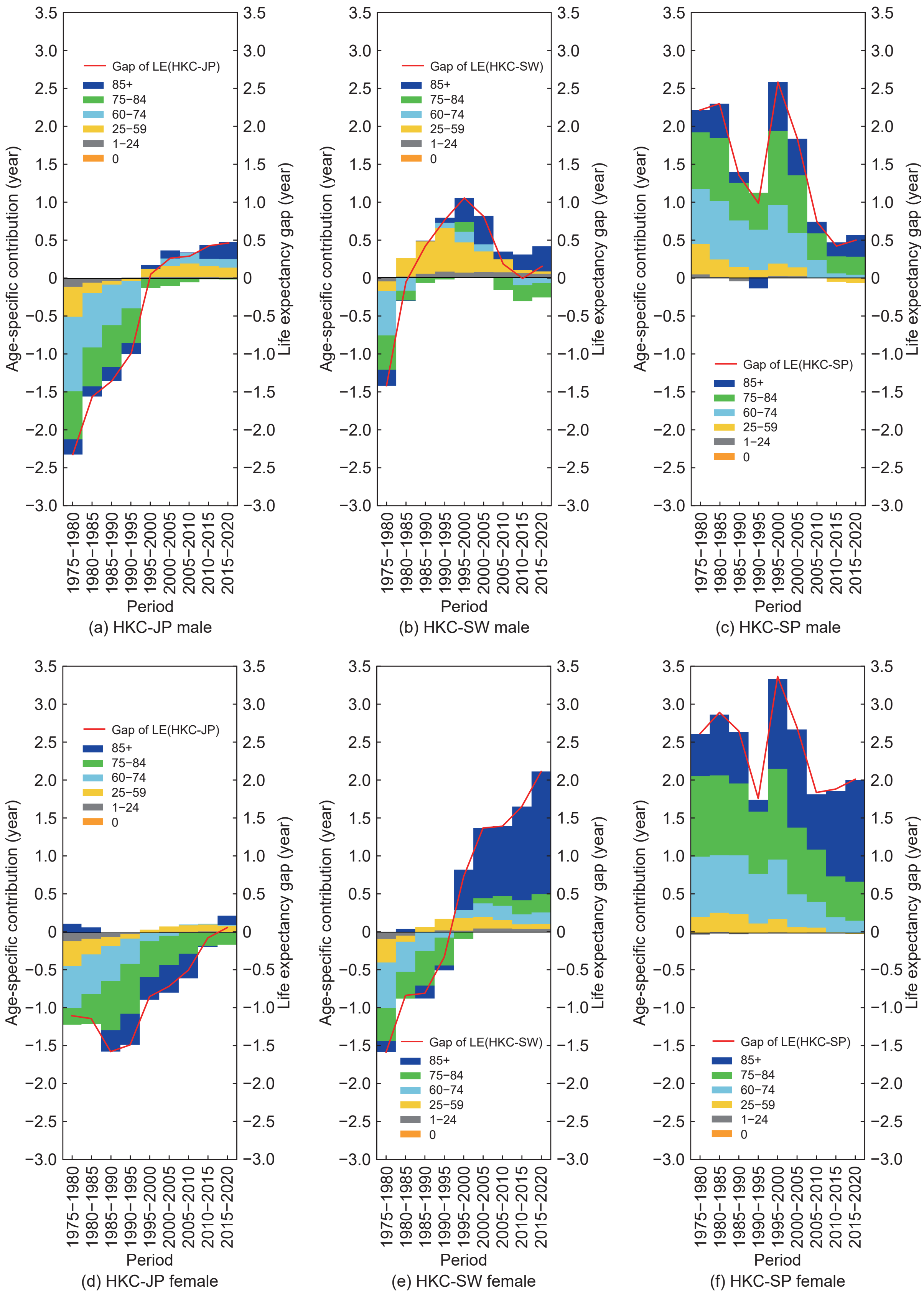


Fig. 3 Effect of mortality rates of different ages on life expectancy comparison of the HKC-JP, HKC-SW, and HKC-SP cases, for male and female.

Kong of China has a public hospital system that served many elderly, especially those who cannot afford private hospital care. A specific percentage of the hospital funding goes to Geriatrics dedicated to take care the elderly. For those elderly retiring at home, traditional Chinese culture may help build tighter bond between elderly and their descendants, who become informal caregivers. Also, Hong Kong of China has a system for importing domestic helpers from neighboring Asian countries and regions, their number accounting for around 4.6%※ of the population of Hong Kong of China. A significant fraction of these domestic helpers is hired specially to take care the elderly people on one-to-one basis. Besides, Hong Kong of China has a significantly higher percentage of elderly staying in government supported senior homes and nursery homes, the percentage is two to four times higher compared to other Asian countries and regions according to Ref. [38]. We argue that the low smoking rate, the above two factors we point out here (and possibly some additional important factors in the collective experience by this cohort of elderly), together explain their life longevity, rather than a single factor (such as low smoking rate) alone can differentiate Hong Kong of China from the target we are comparing to.

We want to clarify that our arguments above are not necessarily saying that Hong Kong's public hospital system and nursing homes are providing adequate services. The commentary of Ref. [7] has pointed out some limitations (e.g., inequality) in Hong Kong's public hospital system. Many of the nursing homes in Hong Kong of China are notoriously cramped and short-handed in staff; it is widely recognized that the quality of life of those elderly staying in nursing homes is rather poor. But they may be doing a sufficient job as far as life longevity of this cohort is concerned.

The recent Omicron outbreak between February to April 2022 caused nearly 9000 deaths in Hong Kong of China. Most of them are elderly. This will likely affect the life expectancy standing of Hong Kong of China in the 2022 tally. In the same way as the above mentioned factors, Covid-19 will thus become a significant negative factor that affect the elderly cohort of population that has thus far contributed most to the world-leading LE ranking for Hong Kong of China.

For Hong Kong's advantage in the 25−59 group mortality rate, we can also try to come up with some factors pertaining most to this cohort's life experiences as hypothesis. Just as with the case for the elderly cohort, less smoking and many other social and environmental factors can be simultaneously relevant. We leave this question to future studies; in particular, the significantly larger contribution of the 25−59 mortality rate for the deficit of SW is puzzling and worthy of future studies.

In summary, we believe the decomposition method and cohort-based analysis, together with carefully choosing comparison targets, led us to a novel approach to explain the relative performance of a population in LE. That is, to focus on the cohorts in a population that performed particularly well (or poorly), relative to a good comparison target, and try to identify a set of factors that together explain the specific cohorts' performance. The factor analysis method is not new, it would still be through hypothesizing, correlation analysis, and causal inference; the decomposition method and the comparison method are not invented by us, but this approach of finding factors and explanations are insightful, and have not been used in the literature we studied, and in particular have not been used by the recent papers by domain experts studying this interesting question.

In conclusion, we first explored different ways to apply data analysis to study the factors that contribute to the life expectancy achievement of a society, e.g., Hong Kong of China. We categorized them into four approaches. We then reported our analysis, in making careful selection of societies to compare, and in applying decomposition analysis to help making hypotheses. Our data analysis suggested a cohort-based approach to explain Hong Kong's high ranking in life expectancy standing. In explaining a particular cohort's life expectancy, we take an approach that identifies multiple critical factors together to explain the gap in LE for that cohort, rather than trying to argue one or two factors are the most important factors. We believe our approach, based on both data analysis and social science arguments, is useful for further studies of this topic.

In closing, we observe that there is no need to be number one, but living longer and happier is always a blessing.

※Based on 2022 statistic of around 339 thousand domestic helpers, and 2021 Census of around 7.4 million population for Hong Kong of China.

Acknowledgment

The authors wish to thank professor Roger Chung from CUHK for his valuable comments after reading an earlier version of the paper. The authors also wish to acknowledge the support of funding (No. UGC/IDS (R)11/21) from the Hong Kong SAR Government.

References

[1] Centre for Health Protection, Life expectancy at birth in Hong Kong, 1971–2020, https://www.chp.gov.hk/en/statistics/data/10/27/111.html, 2022.

[2] CNN Health Story, This urban population is leading the world in life expectancy, https://edition.cnn.com/2018/03/02/health/hong-kong-worldlongest-life-expectancy-longevity-intl/index.html, 2018.

[3] The Standard, Life expectancy of Hong Kong men and women reach 82.9 and 88 years, https://www.thestandard.com.hk/breaking-news/section/4/182966/Lifeexpectancy-of-Hong-Kong-men-and-women-reach-82.9-and-88-years, 2022.

[4] South China Morning Post Video, Hong Kong has the world's highest life expectancy, here's why, https://www.scmp.com/video/scmporiginals/3154653/hong-kong-has-worlds-highest-life-expectancy-heres-why, 2021.

[5] Blog from Borgen Project, This urban population is leading the world in life expectancy, https://borgenproject.org/life-expectancy-in-hong-kong/, 2020.

[6] D Tam, Hong Kong ranked #1 in life expectancy but health quality is questioned, Blog from food for life, https://foodforlifehk.com/blogs/healthtips/hong-kong-ranked-1-in-life-expectancy-but-health-quality-is-questioned, 2021.

[7] R. Chung and M. Marmot, People in Hong Kong have the longest life expectancy in the world: Some possible explanations, https://nam.edu/people-in-hong-kong-have-the-longest-life-expectancy-in-the-world-some-possible-explanations/, 2020.

[8] M. Y. Ni, V. Canudas-Romo, J. Shi, F. P. Flores, M. S. Chow, X. I. Yao, S. Y. Ho, T. H. Lam, C. M. Schooling, A. D. Lopez, et al., *The Lancet Public Health*, vol. 6, no. 12, pp. e919–e931, https://www.sciencedirect.com/science/article/pii/S2468266721002085, 2021.

[9] Y. Zheng, C. Mengni, and P. S. Yip, *International Journal of Health Policy and Management*, vol. 10, no. 1, pp. 5–13, https://pubmed.ncbi.nlm.nih.gov/32610743/, 2021.

[10] Census and Statistics Department, Hong Kong Special Administrative Region, The mortality trend in Hong Kong 1986 to 2018, https://www.statistics.gov.hk/pub/B71911FB2019XXXXB0100.pdf, 2020.

[11] Blog from Alea, Highest life expectancy: How did Hong Kong get there? https://alea.care/resources/highest-life-expectancy-world-hong-kong, 2022.

[12] S. H. Preston, The changing relation between mortality and level of economic development, *International Journal of Epidemiology*, vol. 36, no. 3, pp. 484–490, 2007.

[13] UN life expectancy data, http://data.un.org/Search.aspx?q=life+expectancy, 2022.

[14] WHO life experience data, https://www.who.int/data/gho/indicator-metadata-registry/imr-details/65, 2022.

[15] Our world in data, https://ourworldindata.org/lifeexpectancy, 2022.

[16] Macrotrends, https://www.macrotrends.net/countries/HKG/hong-kong/life-expectancy, 2022.

[17] Global AgeWatch, https://www.helpage.org/global-agewatch/, 2022.

[18] Hong Kong elder quality of life index, Institute of Ageing-AgeWatch Index, https://cuhk.edu.hk, 2022.

[19] Kaggle dataset on life expectancy from WHO, https://www.kaggle.com/datasets, 2022.

[20] H. Hussain, Data science: Which public health factors have the greatest impact on life expectancy, Medium, https://towardsdatascience.com/data-science-whichpublic-health-factors-have-the-greatest-impact-on-lifeexpectancy-52d024ee3d27, 2020.

[21] Z. Chen, Y. Ma, J. Hua, Y. Wang, and H. Guo, Impacts from economic development and environmental factors on life expectancy: A comparative study based on data from both developed and developing countries from 2004 to 2016, *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, p. 8559, 2021.

[22] Y. Wu, K. Hu, Y. Han, Q. Sheng, and Y. Fang, Spatial characteristics of life expectancy and geographical detection of its influencing factors in China, *International Journal of Environmental Research and Public Health*, vol. 17, no. 3, p. 906, 2020.

[23] E. Arroaga, Measuring and explaining the change in life expectancies, *Demography*, vol. 21, no. 1, pp. 83–96, 1984.

[24] D. A. Jdanov, V. M. Shkolnikov, A. A. van Raalte, and E. M. Andreev, Decomposing current mortality differences into initial differences and differences in trends: The contour decomposition method, *Demography*, vol. 54, pp. 1579–1602, 2017.

[25] W. Lutz and E. Kebede, Education and health: Redrawing the Preston curve, *Population and Development Review*, vol. 44, no. 2, pp. 343–361, 2018.

[26] R. Chetty, M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler, The association between income and life expectancy in the United States, 2001–2014, *Jama*, vol. 315, no. 16, pp. 1750–1766, 2016.

[27] G. J. Xu, K. Xu, and Q. Q. Lu, Does the extension of life expectancy promote economic growth—An Empirical Analysis Based on 121 economies, (in Chinese), *Economic Theory and Business Management*, vol. 41, no. 3, pp. 97–112, 2021.

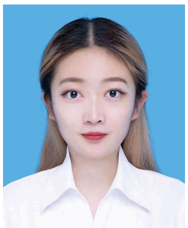
[28] Wikipedia, List of cities in China by life expectancy, https://en.wikipedia.org/wiki/List_o_cities_in_China_by_life_expectancy, 2022.

[29] Xiaoxiang Morning News, Latest life expectancy for Shanghai, (in Chinese), https://baijiahao.baidu.com/s?id=1724924641854556492&wfr=spider&for=pc, 2022.

[30] Xinhua News, Latest life expectancy for Beijing, (in Chinese), http://www.bj.xinhuanet.com/2021-07/27/c_1127698817.html, 2021.

[31] Wikipedia, List of European countries by life expectancy, https://en.wikipedia.org/wiki/List_of_European_countries_by_life_expectancy, 2022.

[32] Wikipedia, List of Japanese prefectures by life expectancy, https://en.wikipedia.org/wiki/List_of_Japanese_prefectures_by_life_expectancy, 2022.
[33] USA Today, Long life: Cities across the US where people live the longest include San Francisco and Boulder, https://www.usatoday.com/story/money/2020/12/24/cities-where-peoplelive-the-longest/115182590/, 2020.
[34] National Statistics, Life tables for Korea, 2017, http://kostat.go.kr/portal/eng/pressReleases/8/6/index.board?bmode=download&bSeq=&aSeq=373128&ord=1, 2018.
[35] Australian Bureau of Statistics, https://www.abs.gov.au/, 2022.
[36] N. Fergany, On the human survivorship function and life table construction, *Demography* , vol. 8, no. 3, pp. 331–334, 1971.
[37] S. David, S. Robert, and B. Jordan, Life expectancy project: The life table, https://www.lifeexpectancy.org/lifetable.shtml, 2022.
[38] Hong Kong Policy Research Institute, Aging in place, (in Chinese), https://www.hkpri.org.hk/research/detail/aging-in-place, 2017.



Ting Xu received the master degree from Chinese University of Hong Kong in 2022, and worked on this research while taking a MSc research course. She received the bachelor degree from Beijing University of Posts and Telecommunications in 2021. She has joined the State Key Laboratory of Information Photonics and Optical Communications in Beijing University of Posts and Telecommunications and published one paper in *IEEE ACCESS* 2019, one in *OECC* 2020, and one in *ACP* 2020. Her current research interests are data analysis and social science computing following professor Dah Ming Chiu.

Ming-sum Tsui is a professor and dean of Felizberta Lo Padilla Tong School of Social Sciences, Caritas Institute of Higher Education (CIHE). Before joining CIHE, he was an associate dean of the Faculty of Health and Social Sciences, Hong Kong Polytechnic University. He received the MSW degree from McGill University in 1985 and the PhD degree from University of Toronto in 2001. His research interests include theory & practice in social work, social work supervision, human service management, and social work education. He has conducted 20 funded research projects, and published 12 books and 78 refereed journal articles. His journal article *From resilience to resistance: A reconstruction of the strengths perspective in social work practice* received the Frank Turner Best Paper Award 2010 from *International Social Work Journal* (*ISWJ*). He is now serving as the co-chief editor of *ISWJ* and editorial board member for other 9 academic journals.

Dah Ming Chiu received the bachelor degree from Imperial College London in 1975 and the PhD degree in applied mathematics from Harvard University in 1980. He has worked at AT&T Bell Labs, DEC, and Sun Microsystems in the US. He joined the Department of Information Engineering at Chinese University of Hong Kong in 2002, served as the department chair from 2009 to 2015, and is now emeritus professor in the department. Currently, he is also a research professor and the director of the Data Science Research Center at Caritas Institute of Higher Education. He is an IEEE fellow. His current research interest is in applying data science to various practical problems, especially in social services and policies.