Published by Oxford University Press on behalf of the International Epidemiological Association
© The Author 2007; all rights reserved. Advance Access publication 21 March 2007
International Journal of Epidemiology 2007;**36**:542–548
doi:10.1093/ije/dym008

The shape of health to come: prospective study of the determinants of 30-year health trajectories in the Alameda County Study

George A Kaplan,[1]* Peter T Baltrus[2] and Trivellore E Raghunathan[3]

Accepted 15 January 2007

Background Considerable evidence now exists indicating that incidence and progression of disease and disability are associated with socio-economic, behavioural, demographic and psychosocial factors. The emerging life course perspective suggests that these factors might be associated with not just the incidence of disease and death, but also trajectories of health over decades.

Methods Prospective study of a representative, population-based cohort studied on four occasions over 30 years. Trajectories of health over this period, combining self-rated health and date of death, were related to behavioural, psychosocial and socio-economic risk factors.

Results Trajectories of health were associated with behavioural, socioeconomic, and psychosocial risk factors, the strongest predictors being household income and physical activity. Those with an income 1 SD above the mean were ~25% more likely to die having previously consistently reported excellent health (Men: relative risk (RR) = 1.27, 95% CI = 1.10–1.46; Women: RR = 1.25, 95% CI = 1.06–1.47), were more likely to remain 'Alive in Excellent Health' (Men: RR = 1.35, 95% CI = 1.21–1.51; Women: RR = 1.30, 95% CI = 1.19–1.43) and were less likely to have shown an 'Unremitting Decline' (Men: RR = 0.72, 95% CI = 0.49–1.05; Women: RR = 0.71, 95% CI = 0.48–1.04). Those with low physical activity were ~50% less likely to die having consistently reported excellent health (Men: RR = 0.54, 95% CI = 0.39–0.76; Women: RR = 0.48, 95% CI = 0.33–0.71), and were five-times more likely to show an 'Unremitting Decline' (Men: RR = 5.05, 95% CI = 1.75–14.56; Women: RR = 5.00, 95% CI = 1.48–16.92). They were also less likely to be 'Alive in Excellent Health' (Men: RR = 0.41, 95% CI = 0.29–0.57; Women: RR = 0.44, 95% CI = 0.33–0.57).

Conclusions The burden of illness associated with behavioural, socio-economic and psychosocial risk factors extends beyond shortening of life to poorer trajectories of health over decades.

Keywords Socioeconomic, life course, aging, obesity, smoking, depression

Over the last several decades there has been increasing recognition that behavioural, social, psychological and socio-economic factors are associated with longevity in the elderly.[1–4] Increasing the length of life is, of course, only one goal and a growing interest in increased quality of life while aging has also been the focus of a number of analyses.[5–8]

Consider several possible trajectories of a generic measure of health and mortality generated for the three individuals,

[1] Center for Social Epidemiology and Population Health, University of Michigan School of Public Health, Ann Arbor, MI 48104, USA.

[2] Morehouse School of Medicine, National Center for Primary Care, Atlanta, GA 30310, USA.

[3] Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan 48109, USA.

* Corresponding author. Center for Social Epidemiology and Population Health, University of Michigan School of Public Health, 1214 South University Street, Ann Arbor, MI 48104, USA. E-mail: gkaplan@umich.edu

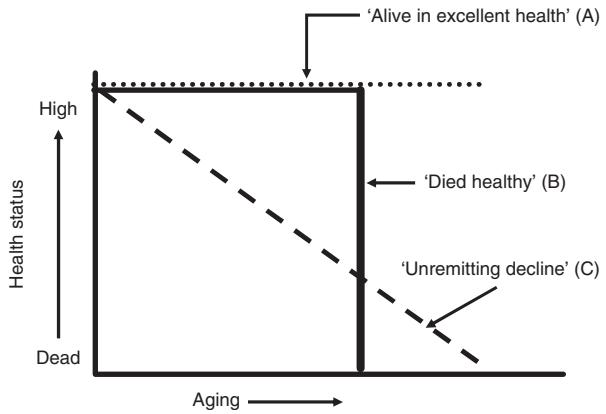

Figure 1 Three possible health trajectories over decades

portrayed in Figure 1. During the period of observation, which is meant to be decades, the person labelled 'Alive in Excellent Health (A)', survives as long as the person labelled 'Unremitting Decline (C)'. However, it is likely that few would choose scenario C over A. Other trajectories are also possible. For example, 'Died Healthy' (B) portrays a person who lives less long than C but does so in good health, the 'one hoss shay' example from Fries[9] in which a long period of health is followed by a very short period of illness leading to death. What these examples highlight is the importance of considering not only length of life, and health status during the later years, but also the 'shape' of the health trajectory over a substantial period of the life span. Attention to this issue, however, has been hampered by the lack of data on the health of individuals followed over a long period.

We used the Alameda County Study, a population-based, prospective epidemiologic study, with five additional waves of data collection and mortality ascertainment, of the association between a broad array of behavioural, psychosocial and socio-economic factors and risk of death and other outcomes[10–12] to study the risk of and factors associated with these trajectories of self-rated health and mortality over 30 years.

Methods

Study population

The Alameda County Study began in 1965 and used a stratified random household sampling procedure to recruit a representative sample of the adult, non-institutionalized population of Alameda County, California. A total of 6928 adults (86% of those eligible), ages 16–94 years, provided baseline information by written self-administered questionnaires on health history and status, and behavioural, psychosocial, socio-economic and demographic variables. Table 1 presents characteristics of this cohort at baseline in 1965. Subjects were tracked regardless of location and survivors were asked to complete questionnaires again in 1974, 1983 (50% sample), 1994, 1995 and 1999 with response rates (calculated as the percentage of questionnaires mailed to those not known to be dead at each wave that were returned completed) between 85% and 97%. For the present study, data from the 1965, 1974, 1983 and 1994 surveys were used in combination with mortality information up to 1994. After initial recruitment and determination of eligibility, questionnaires were sent to and returned from participants by mail. Detailed design and sampling procedures for this study have been reported previously.[11–13] This research protocol was approved by the California Health and Welfare Agency Committee for the Protection of Human Subjects and conducted in accordance with their guidelines regarding informed consent for participation in research in which the sole form of data collection is via mailed surveys.

Table 1 Alameda County Study – 1965

		N	%
Age	<20	49	0.7
	20–29	1590	23.0
	30–44	2229	32.2
	45–64	2170	31.3
	65+	890	12.9
Sex	Female	3770	54.4
	Male	3158	45.6
Race/Ethnicity	White NH	5464	78.9
	Black NH	862	12.5
	Hispanic	273	3.9
	Other	327	4.7
Education	<0–8	1243	18.0
	9–11	1229	17.8
	12	2094	30.4
	13+	2331	33.8
Self-rated health	Excellent	1723	24.9
	Good	3906	56.5
	Fair	1141	16.5
	Poor	138	2.0

Ascertainment of deaths, through 1994 in the present analyses, was based on annual search of mortality tapes from the California Department of Health Services Office of Vital Records, the National Death Index and the Social Security Death Index. Further deaths were ascertained during follow-up contacts for additional waves of data collection, and all death certificates were matched to participant information obtained in previous waves of data collection before a participant was identified as deceased.

Measurement of health status

Self-rated health status, rated as 'Excellent, Good, Fair, or Poor', was used as a generic measure of overall health. Self-rated health has been shown to be strongly associated with risk of death in the Alameda County Study[14] and in many other studies.[15] It is also strongly associated with a number of clinically measured indicators of poor health,[16] risk of hospitalization,[17] acute health deterioration,[18] comorbidity and disease burden[19–20] and physical function status.[21]

Measurement of health trajectories

All results reported are for those who rated their health as 'excellent' in 1965, estimated from models that included all respondents. Using information from subsequent waves of data collection in 1974, 1983 and 1994, and mortality information through 1994, a subset of three health trajectories were selected to be examined. The general strategy was to first estimate the probability of death between one wave and the next, and then, for those who survived, estimate the probability of change in health status, creating a probability 'tree' structure, which represents all deaths and health status transitions from 1965 to 1994. Based on this tree structure we identified three health trajectories plus death at any time. The 'Alive in Excellent Health' trajectory corresponds to reporting health as excellent in 1965, 1974 and 1983, and remaining alive through 1994 (Table 2). The 'Unremitting Decline' trajectory corresponds to a progressive decline in health status at each new wave of data collection with death, when it occurred, included as a possible lowest level of health status. The 'Died Healthy' reflects reporting 'Excellent Health' at all waves of data collection prior to death sometime between 1965 and 1994.

Predictors of trajectories of health

We examined the association between cigarette smoking (current, former or never), household income, body mass index (BMI; categorized into: low weight: $<20\,kg/m^2$;

Table 2 Relative probabilities of 1965–94 health trajectories by risk factor status: Alameda County Study

	'Died'		'Died Healthy'		'Unremitting Decline'		'Alive in Excellent Health'	
	Men	Women	Men	Women	Men	Women	Men	Women
Smoking								
Current	2.23	2.46	0.77	0.72	2.68	2.92	0.53	0.58
	1.63–3.05	1.72–3.53	0.53–1.10	0.47–1.12	0.92–7.80	0.91–9.38	0.40–0.72	0.45–0.76
Former	1.46	1.57	0.98	0.95	1.84	2.18	0.88	0.87
	1.04–2.03	1.16–2.12	0.68–1.43	0.70–1.28	0.55–2.50	0.49–4.34	0.73–1.05	0.71–1.06
Never (reference)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Income								
Mean + 1 SD	0.75	0.75	1.26	1.25	0.72	0.71	1.36	1.30
	0.66–0.86	0.65–0.88	1.10–1.46	1.06–1.47	0.49–1.05	0.48–1.04	1.21–1.51	1.19–1.43
Mean (reference)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Mean–1 SD	1.33	1.36	0.76	0.77	1.37	1.40	0.70	0.73
	1.17–1.54	1.15–1.61	0.64–0.91	0.63–0.93	0.93–2.02	0.87–2.21	0.59–0.81	0.64–0.83
BMI								
<20.0	1.31	1.36	1.07	1.05	1.62	1.64	0.91	0.96
	0.94–1.82	0.87–2.13	0.70–1.61	0.60–1.83	0.48–5.44	0.42–6.39	0.66–1.24	0.74–1.24
20.0–24.9 (reference)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25.0–29.9	0.73	0.67	1.30	1.41	0.62	0.53	0.84	0.83
	0.55–0.96	0.48–0.93	0.99–1.71	0.99–2.00	0.23–1.71	0.17–1.63	0.66–1.05	0.68–1.04
⩾30.0	0.98	0.99	0.75	0.73	2.14	2.21	0.49	0.48
	0.63–1.51	0.59–1.64	0.46–1.22	0.41–1.29	0.73–6.27	0.66–7.44	0.31–0.79	0.31–0.76
Alcohol								
Abstain	1.33	1.26	0.85	0.90	1.62	1.67	0.69	0.72
	1.00–1.76	1.06–1.94	0.61–1.19	0.63–1.29	0.75–3.50	0.72–3.88	0.50–0.92	0.56–0.94
1–2 drinks/day (reference)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
>2 drinks/day	1.58	1.72	0.74	0.71	1.75	2.25	0.85	0.87
	1.18–2.15	1.12–2.65	0.52–1.05	0.45–1.12	0.67–4.53	0.70–7.31	0.63–1.15	0.66–1.13
Physical activity								
Low	1.76	2.05	0.54	0.48	5.05	5.00	0.41	0.44
	1.31–2.38	1.46–2.87	0.39–0.76	0.33–0.71	1.75–14.56	1.48–16.92	0.29–0.57	0.33–0.57
Medium	1.49	1.67	0.60	0.54	4.20	4.42	0.63	0.65
	1.12–1.98	1.17–2.38	0.44–0.82	0.36–0.80	1.40–12.61	1.19–16.42	0.49–0.81	0.53–0.81
High (reference)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Depression								
⩾5 symptoms	1.38	1.35	0.65	0.67	1.21	1.18	0.55	0.57
	0.97–1.97	0.86–2.12	0.41–1.04	0.32–1.42	1.08–1.36	1.06–1.32	0.37–0.81	0.41–0.97
<5 symptoms (reference)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

normal weight: 20 to <25 kg/m^2; overweight: >25 to <30 kg/m^2 and obese: 30+ kg/m^2), alcohol consumption (0 drinks per day; >0–2 drinks per day, and 2+ drinks per day), physical activity, and depression (all assessed at each wave) and the trajectories by computing RRs, defined as the ratio of probabilities of belonging to a trajectory group for two different risk factor profiles. Physical activity level was constructed from three separate items: swimming or taking long walks, active sports and doing physical exercise, each scores as 'often' = 2 points, 'sometimes' = 1, 'never' = 0. The scores from each item were summed to create a physical activity score which was categorized as follows: Low Activity = 0 or 1, Medium Activity = 2 or 3, High Activity = 4+ (2). Depression was defined as reporting five or more depressive symptoms on a fifteen-item depressive symptom inventory.[22]

Statistical methods

Probabilities for the three trajectories were based on a series of logistic and polytomous regression models. Logistic regression models were used for mortality, and polytomous response logistic regression models were used for self-rated health status (Figure 2). For example, the first mortality model estimated the probability of death between 1965 and 1974. The first polytomous regression model estimated the probability of change in perceived health status from 1965 to 1974 (improved, declined, stayed the same), conditional on being alive and participating in 1974. Age, sex and baseline (1965) health status (as dummy variables) were included in all models. The process was repeated for each subsequent follow-up interval. Separate models were estimated for each risk factor, with risk factors updated at each wave to enable the estimation of probabilities using the most recent data. Probabilities of each trajectory by risk factor status were calculated by traversing through all possible paths in the probability tree leading to that trajectory. Thus, the probabilities represent the probability of a particular trajectory for a man or woman of average age who had excellent baseline health by risk factor level. For example, the association of physical activity with an 'unremitting decline' is expressed as the RR of that outcome in those who were active at each wave compared with those who were inactive at each wave. The standard errors and the confidence intervals (CI) for probabilities and RR were estimated using 100 bootstrap samples.

Results

In 1965, 24.9% of the participants (870 men, 853 women) reported being in 'Excellent Health'. Between 1965 and 1994, 227 (26.1%) of the men and 187 (21.9%) of the women died. Based on inspection of patterns of outcomes across waves of data collection, 'Died Healthy' was found for 16.8% (146) of the men who reported 'excellent health' and 13.1% (112)

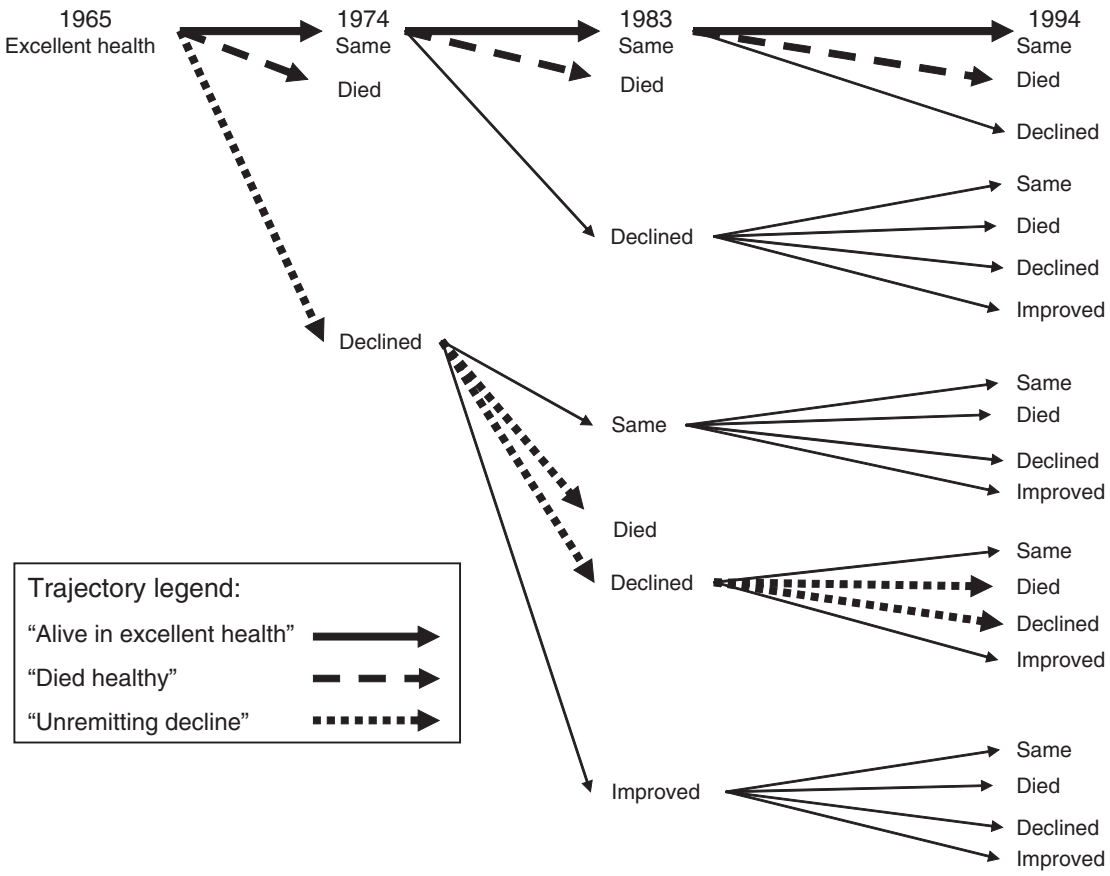

Figure 2 Tree structure illustrating calculation of health trajectories

of the women in 'excellent health'. 'Unremitting Decline' were observed for 8.0% (70) of the men, and 6.2% (53) of the women, and 40.34% (351) of the men and 39.0% (333) of the women remained 'Alive in Excellent Health' throughout the follow-up period. Other trajectories were found for 34.8% (303) of the men, and 41.6% (355) of the women.

There were considerable differences in the probability of each trajectory by risk factor levels. As expected, both male and female current smokers compared with never smokers had a greater than 2-fold risk of death (Men: RR = 2.23, 95% CI = 1.63–3.05; Women: RR = 2.46, 95% CI = 1.72–3.53). Current smokers compared with never smokers were also less likely to have 'Died Healthy' (Men: RR = 0.77, 95% CI = 0.53–1.10; Women: RR = 0.72, 95% CI = 0.47–1.12), were more likely to suffer an 'Unremitting Decline' (Men: RR = 2.68, 95% CI = 0.92–7.80; Women: RR = 2.92, 95% CI = 0.91–9.38) and were less likely to have remained 'Alive in Excellent Health' (Men RR = 0.53, 95% CI = 0.40–0.72; women RR = 0.58, 95% CI = 0.45–0.76). Generally, the results for former smokers showed similar but smaller elevated (or reduced) RRs.

Health trajectories were also associated with BMI. Low BMI (<20) was not associated with any of the outcomes. The overweight (BMI 25–29.9) group was less likely to die (Men: RR = 0.73, 95% CI = 0.55–0.96; Women: RR = 0.67, 95% CI = 0.48–0.93) than the normal weight reference group (BMI 20–24.9). Overweight compared with normal weight participants also tended to have a higher probability of 'Dying Healthy' (Men: RR = 1.30, 95% CI = 0.99–1.71; Women: RR = 1.41, 95% CI = 0.99–2.00). On the other hand, those who were obese (BMI $\geqslant$ 30) were 50% less likely to be 'Alive in Excellent Health' (Men: RR = 0.49, 95% CI = 0.31–0.79; Women: RR = 0.49, 95% CI = 0.31–0.76).

Men who abstained from alcohol were at a 33% (RR = 1.33, 95% CI = 1.00–1.76) greater probability of dying and 31% (RR = 0.69, 95% CI = 0.50–0.92) less likely to be 'Alive in Excellent Health' than moderate drinkers (>0–2 drinks per day). Female abstainers were 28% (RR = 0.72, 95% CI = 0.56–0.94) less likely to be 'Alive in Excellent Health' than moderate drinkers. Men who drank more than two drinks a day were at a 59% (RR = 1.59, 95% CI = 1.18–2.15) greater risk of death than moderate drinkers, while female heavy drinkers were at a 72% (RR = 1.72, 95% CI = 1.12–2.65) greater risk of death

Physical activity was strongly associated with health trajectories for both men and women. Men with a low physical activity level were 76% (RR = 1.76, 95% CI = 1.31–2.38) more likely to die than those with a high activity level, low activity women had twice the risk of death (RR = 2.04, 95% CI = 1.46–2.87). Low activity men and women were about 50% less likely to 'Die Healthy' (Men: RR = 0.54, 95% CI = 0.39–0.76; Women: RR = 0.48, 95% CI = 0.33–0.71). A 5-fold risk of an 'Unremitting Decline' was found for low activity men and women (Men: RR = 5.05, 95% CI = 1.75–14.56; Women: RR = 5.00, 95% CI = 1.48–16.92). Low activity men were 59% (RR = 0.41, 95% CI = 0.29–0.57) less likely to be 'Alive in Excellent Health'. Low activity women were 56% (RR = 0.44, 95% CI = 0.33–0.57) less likely to be 'Alive in Excellent Health'. Associations for medium activity level were weaker than for low activity level, but all were substantial.

Household income was a predictor of all out outcomes except 'Unremitting Decline'. Those with an income 1 SD above the mean were 25% less likely to die (Men: RR = 0.75, 95% CI = 0.66–0.86; Women: RR = 0.75, 95% CI = 0.65–0.88), while those with an income 1 SD below the mean were ~35% more likely to die (Men: RR = 1.34, 95% CI = 1.17–1.54; Women: RR = 1.36, 95% CI = 1.15–1.61). Those with higher income were ~25% more likely to 'Die Healthy' (Men: RR = 1.27, 95% CI = 1.10–1.46; Women: RR = 1.25, 95% CI = 1.06–1.47), those with lower income were 23% less likely to 'Die Healthy' (Men: RR = 0.77, 95% CI = 0.64–0.91; Women: RR = 0.77, 95% CI = 0.63–0.93). Higher income was also a predictor of being 'Alive in Excellent Health' (Men: RR = 1.35, 95% CI = 1.21–1.51; Women: RR = 1.30, 95% CI = 1.19–1.43), while lower income was inversely related to the same outcome (Men: RR = 0.69, 95% CI = 0.59–0.81; Women: RR = 0.73, 95% CI = 0.64–0.83). Lower income was positively associated with an 'Unremitting Decline' (Men: RR = 1.37, 95% CI = 0.93–2.02; Women: RR = 1.39, 95% CI = 0.87–2.21), while higher income was protective (Men: RR = 0.72, 95% CI = 0.49–1.05; Women: RR = 0.71, 95% CI = 0.48–1.04).

The association between depression and death and dying healthy was weaker, and important for men only (RR = 1.38, 95% CI = 0.97–1.97 for death; RR = 0.65, 95% CI = 0.41–1.04 for 'Dying Healthy'). An ~20% elevated risk of an 'Unremitting Decline' was found for men and women with five or more depressive symptoms (Men: RR = 1.21, 95% CI = 1.08–1.36; Women: RR = 1.18, 95% CI = 1.06–1.32). Men and women with five or more depressive symptoms were >40% less likely to be alive and in excellent health (Men: RR = 0.55, 95% CI = 0.37–0.81; Women: RR = 0.57, 95% CI = 0.41–0.79).

Conclusion

In this representative population-based cohort followed for 30 years, risk of death and trajectories of health among the living were importantly associated with behavioural, socio-economic, and psychosocial risk factors. Not surprisingly, risk of death was greatest among those who smoked, had low incomes, abstained from or consumed more than a moderate amount of alcohol and were physically inactive. In addition, as reported previously,[23] those with BMI in the 'Overweight' category (25.0–29.9) survived longer than those in higher or lower BMI.

Most importantly there was considerable variation in temporal patterns of self-rated health, in relation to potentially modifiable risk factors. The strongest and most consistent results were found for income and physical activity. Those with low income or relatively low levels of physical activity were more likely to have died and less likely to be 'Alive and in Excellent Health', and if they had died they were less likely to have reported 'Excellent' levels of health in the waves of data collection prior to their death and more likely to have experienced deteriorating health prior to their death. As low income and residence in poor areas are associated with declines in physical activity[24] one might interpret these results as reflecting an impact of income, and what it represents, on physical activity, but because all risk factors were measured cross-sectionally, at baseline and at subsequent waves, we cannot be confident of such an interpretation.

As a summary measure, self-rated health has been shown to be cross-sectionally and prospectively related to clinically significant outcomes,[14–21] but it is not possible to determine in this study the contribution of various disease pathways to its trajectory. Based on previous analyses,[20] we do know that self-rated health is strongly related to disease symptoms. However, analyses of studies with both repeated examinations and long-term follow-up will be necessary to fully understand the pathways linking risk factors and long-term trajectories of health.

Most longitudinal studies model the probability or time to initial occurrence of health compromising conditions or, for recurrent conditions, the risk of re-occurrence, and relate such outcomes to baseline or updated risk factor levels. To our knowledge, this is the first study to examine the relationship between risk factor levels and the shape or trajectory of health over long periods, distinguishing between length of survival and level of health. The results suggest that the burden of illness associated with behavioural, socio-economic and psychosocial risk factors, particularly low income and physical inactivity, extends beyond the shortening of life leaving an imprint on patterns of health over decades.

Acknowledgements

This study was funded by awards from the National Institute of Child Health Development (P50 HD038986) and the National Institute on Aging (R37AG11375) to G.A.K.

Conflict of interest: None declared.

KEY MESSAGES

- We found that trajectories of self-reported health over 30 years were importantly associated with a broad range of behavioural, socio-economic and psychosocial risk factors.
- People with low income and those who were physically inactive were most likely to die early, to report 'Unremitting Declines' in health before death, and if they remained alive over three decades were less likely to have consistently reported excellent levels of health.
- Further analyses of other long-term, multi-wave longitudinal studies may provide additional evidence that the imprint of socioeconomic, behavioural, demographic and psychosocial is not only on the risk of disease but also on trajectories of health over decades.

References

[1] Kaplan GA, Seeman TE, Cohen RD, Knudsen LP, Guralnik J. Mortality among the elderly in the Alameda County Study: behavioral and demographic risk factors. *Am J Public Health* 1987;**77**:307–12.

[2] Kaplan GA, Haan MN. Is there a role for prevention among the elderly? Epidemiologic evidence from the Alameda County Study. In: Ory MG and Bond K (eds). *Aging and Health Care: Social Science and Policy Perspectives*. London: Tavistock Publications; 1989, pp. 27–51.

[3] Kaplan GA, Haan MN, Wallace RB. Understanding changing risk factor associations with increasing age in adults. *Ann Rev Public Health* 1999;**20**:89–108.

[4] Seeman TE, Crimmins E. Social environment effects on health and aging: integrating epidemiological and demographic approaches and perspectives. In: Weinstein M, Hermalin AI and Stoto MA (eds). *Population Health and Aging: Strengthening the Dialogue Between Epidemiology and Demography, Ann N Y Acad Sci*. 2001, pp. 88–117.

[5] Guralnik JM, Kaplan GA. Predictors of healthy aging: prospective evidence from the Alameda County Study. *Am J Public Health* 1989;**79**:703–08.

[6] Kaplan GA, Strawbridge WJ. Behavioral and social factors in healthy aging. In: Abeles RP, Gift HC and Ory MG (eds). *Aging and Quality of Life*. New York: Springer Publishing Company; 1994, pp. 57–78.

[7] Berkman LF, Seeman TE, Albert M *et al*. High, usual and impaired functioning in community-dwelling older men and women: findings from the MacArthur Foundation Research Network on successful aging. *J Clin Epidemiol* 1993;**46**:1129–40.

[8] Camacho TC, Strawbridge WJ, Cohen RD, Kaplan GA. Functional ability in the oldest old: cumulative impact of risk factors from the preceding two decades. *J Aging Health* 1993;**5**:439–54.

[9] Fries JF, Crapo LM. *Vitality and Aging*. San Francisco, CA: WH Freeman; 1981.

[10] Hochstim JR. Health and ways of living—the Alameda County, California, population laboratory. In: Kessler IJ and Levin ML (eds). *The Community as an Epidemiologic Laboratory*. Baltimore, MD: Johns Hopkins University Press; 1970, pp. 149–76.

[11] Berkman LF, Breslow L. *Health and Ways of Living: The Alameda County Study*. New York, NY: Oxford University Press; 1983.

[12] Kaplan GA. Health and aging in the Alameda County Study. In: Schaie KW, Blazer D and House JS (eds). *Aging, health Behaviors, and Health Outcomes*. New Jersey: Lawrence Erlbaum Associates; 1992, pp. 69–88.

[13] Lynch JW, Kaplan GA, Shema SJ. Cumulative impact of sustained economic hardship on physical, cognitive, psychological, and social functioning. *N Engl J Med* 1997;**337**:1889–95.

[14] Kaplan GA, Camacho T. Perceived health and mortality: a nine–year follow–up of the Human Population Laboratory cohort. *Am J Epidemiol* 1983;**117**:292–304.

[15] Benyamini EL. Self-rated health and mortality: a review of twenty-seven community studies. *J Health & Soc Behav* 1997;**38**:21–37.

[16] Kaplan GA, Goldberg DE, Everson SA *et al*. Perceived health status and morbidity and mortality: evidence from the Kuopio ischaemic heart disease risk factor study. *Int J Epidemiol* 1996;**25**:259–65.

[17] Kennedy BS, Kasl SV, Vaccarino V. Repeated hospitalizations and self-rated health among the elderly: a multivariate failure time analysis. *Am J Epidemiol* 2001;**153**:232–41.

[18] Diehr P, Williamson J, Patrick DL, Bild DE, Burke GL. Patterns of self-rated health in older adults before and after sentinel health events. *J Am Geriatr Soc* 2001;**49**:36–44.

[19] Orfila F, Ferrer M, Lamarca R, Alonso J. Evolution of self-rated health status in the elderly: cross-sectional vs. longitudinal estimates. *J Clin Epidemiol* 2000;**53**:563–70.

[20] Cott CA, Gignac MA, Badley EM. Determinants of self rated health for Canadians with chronic disease and disability. *J Epidemiol Community Health* 1999;**53**:731–36.

[21] Jylha M, Guralnik JM, Balfour J, Fried LP. Walking difficulty, walking speed, and age as predictors of self-rated health: the women's health and aging study. *J Gerontol A-Bio Sci Med Sci* 2001;**56**:M609–17.

[22] Kaplan GA, Roberts RE, Camacho TC, Coyne JC. Psychosocial predictors of depression. Prospective evidence from the Human Population Laboratory studies. *Am J Epidemiol* 1987;**125**:206–20.

[23] Strawbridge WJ, Walhagen MI, Shema SJ. New NHLBI clinical guidelines for obesity and overweight: Will they promote health? *Am J Pub Health* 2000;**90**:340–43.

[24] Yen IH, Kaplan GA. Poverty area residence and changes in physical activity level: Evidence from the Alameda County Study. *Am J Public Health* 1998;**88**:1709–12.

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2007;**36**:548–549

doi:10.1093/ije/dym083

Commentary: Trajectories, Selection and Cumulative Causation

Andrew Pickles

Accepted 22 March 2007

Recent advances in life-course epidemiology have been considerable. Researchers now have a taste for examining a range of questions that involve contrasting combinations of patterns of exposure and patterns of impact on health outcomes that may be distant or occur over an extended time period, and data sets are becoming available where such questions can potentially be feasibly addressed.[1,2] However, the statistical frameworks in which such ideas can be formalized and tested are currently limited in scope and the analytical problem has many different facets that require solution. Kaplan *et al*.[3] present an interesting and surprisingly simple method of analysis that may offer useful insight. It suffers, however, from a variety of shortcomings.

One task in dealing with complex repeated measures is to simplify the outcome profile, and the reduction to a small number of trajectory classes offers an attractive option. Kaplan *et al*. estimate probabilities for each possible raw data trajectory under a simple model involving time updated exposures, and then sum these over a priori-defined classes of 'similar' trajectories. A contrasting model based approach, in common use in criminology and psychology,[4] is to empirically identify latent trajectory classes. Croudace *et al*.[5] provide an epidemiological example. Application of this approach usually takes one of two forms. The major classes of trajectory are first identified without reference to risk factors, a step that is essentially a form of model-based cluster analysis. Then the association of risk factors to each subject's posterior trajectory class membership probabilities or maximum a posteriori class assignment is then examined as a second step. This is potentially inefficient, particularly in a context of high levels of sporadic missing data or attrition. Alternatively, covariates can be linked to class membership probabilities through a multinomial model that forms part of the cluster analysis. While this works fine for baseline measures, the inclusion of later risk exposures raises serious concerns of allowing future values of exposure to influence trajectory probabilities where part of the trajectory has occurred prior to exposure. While latent trajectory models can be elaborated to consider latent transitions, with time updated risk factors influencing transitions in an appropriate time-ordered fashion, such models are complex and need considerable care in their formulation and implementation, and to the uninitiated seem to involve many arbitrary assumptions. By contrast, the Kaplan approach makes proper use of time-updated exposures in the determination of trajectories in a surprisingly straight-forward manner. However, many of these assumptions made in latent transition modeling merely make explicit assumptions implicit in such simpler analysis. For example, in the Kaplan *et al*. study, an explicit assumption is required as to the age-independence of some of the measurement properties of self-rated health.

A second major issue is the simultaneous consideration of both the state of health when alive and, when it occurs, the event of death. Such a joint consideration has been commonly cast in a Quality of Life Years framework, but in the Kaplan *et al*. study, the much more rudimentary categorization of death as a worsening of health has the appeal of simplicity. The occurrence of death also brings censoring of future observations. Having a distinct binomial model for inter-wave death

Biostatistics, Informatics and Health Economics Group, School of Community Medicine, University of Manchester, England.

E-mail: andrew.r.pickles@manchester.ac.uk