



## Natural History of Leisure-time Physical Activity and Its Correlates: Associations with Mortality from All Causes and Cardiovascular Disease Over 28 Years

George A. Kaplan,<sup>1</sup> William J. Strawbridge,<sup>2</sup> Richard D. Cohen,<sup>2</sup> and Laura R. Hungerford<sup>2</sup>

The association between leisure-time physical activity and 28-year (1965–1993) risk of death from all causes and cardiovascular disease was studied in 6,131 adults who participated in the Alameda County Study in Northern California. Because study participants were interviewed on a number of occasions, it was possible to include in the analyses information on changes over time in levels of leisure-time physical activity as well as changes in a wide variety of other risk factors. There were 47,616 person-years of observation for males (639 deaths from all causes and 321 from cardiovascular disease) and 57,666 person-years of observation for females (587 deaths from all causes and 388 from cardiovascular disease). In analyses in which only the baseline values of all covariates were included, a four-point increase on the leisure-time physical activity scale, the interquartile range, was associated with reduced risk of death from all causes (relative risk (RR) = 0.90, 95% confidence interval (CI) 0.83–0.99) and cardiovascular disease (RR = 0.85, 95% CI 0.75–0.97). When time-varying information on leisure-time physical activity and all other covariates was included, there was still a protective effect for all-cause and cardiovascular disease mortality (RR = 0.84, 95% CI 0.77–0.92 and RR = 0.81, 95% CI 0.71–0.93, respectively). The association between leisure-time physical activity and risk of death was not altered when information on variations over time in leisure-time physical activity and many determinants and consequences of physical activity were explicitly included in survival models. *Am J Epidemiol* 1996; 144:793–7.

cardiovascular diseases; exercise; longitudinal studies; mortality; physical fitness; risk factors

A substantial number of epidemiologic studies have indicated that participation in leisure-time physical activity protects against the risk of a variety of physical and mental health outcomes, including death (1–6). These studies conducted in numerous locations, using a wide variety of measures of physical activity and different populations, have led to public health recommendations regarding the important preventive value of physical activity for the population (7, 8). While the evidence for these recommendations is compelling, there remain two issues which somewhat cloud interpretation of the epidemiologic evidence that supports these recommendations. Both relate to the design limitations of most epidemiologic studies that have studied the association between physical activity

and health outcomes. While there are some exceptions (9–11), most studies have examined the association between a single measure of physical activity, measured at study baseline, and subsequent health outcomes. Because physical activity, as with many behavioral and biologic risk factors, can change substantially over time, such analyses, particularly those with long periods of follow-up, can be subject to considerable misclassification of physical activity over time. Even more importantly, physical activity may be influenced by many other risk factors (12–14), and may also influence those risk factors (15). In such a state of affairs, quasi-causal interpretations of the results of observational studies of the associations between single measures of physical activity and other risk factors and health outcomes potentially can be misleading.

In the analyses reported here, we make use of physical activity and other risk factor data collected on multiple occasions between 1965 and 1983 in the Alameda County Study (16, 17) to examine the risk of death in this cohort during 1965–1993. The availability of multiple measures, over time, of physical activ-

Received for publication January 16, 1996, and accepted for publication August 16, 1996.

Abbreviations: CI, confidence interval; RR, relative risk.

<sup>1</sup> Human Population Laboratory, California Department of Health Services, Berkeley, CA.

<sup>2</sup> Human Population Laboratory, California Public Health Foundation, Berkeley, CA.

Reprint requests to Dr. George A. Kaplan, Human Population Laboratory, California Department of Health Services, 2151 Berkeley Way, Annex 2, Suite 300, Berkeley, CA 94704–1011.

ity and other risk factors enables us to directly address the design limitation described above.

## MATERIALS AND METHODS

### Study sample

The Alameda County Study is a longitudinal study of factors associated with health and mortality. Since 1965, it has followed 6,928 Northern California adults who were aged 16–94 years at baseline (16, 17). Subjects were originally selected through a stratified random sample of Alameda County households to be representative of Alameda County, California, and have been followed regardless of where they have moved. Survivors were resurveyed in 1974, 1983 (a 50 percent sample), and 1994. The percent response rates for the four surveys were 86, 85, 87, and 93, respectively. The current analyses utilize physical activity and other risk factor data collected in 1965, 1974, and 1983. Subjects who answered the 1965 physical activity questions and who had no missing responses on any of the adjustment variables used were included in the analyses ( $n = 6,131$ ) reported here.

### Measures

**Physical activity.** A scale was constructed from three questions on physical activity. Subjects were asked how often they did physical exercise, took part in active sports, and took long walks or swam. Responses to these questions were never, sometimes, or often and were scored as 0, 2, or 4, respectively. The resulting physical activity scale had a range of 0 to 12 for each follow-up. The items have been used in a number of other analyses from the Alameda County Study, and scores which include them have been importantly associated with risk of death (2, 9, 17, 18).

**Other risk factors.** Sociodemographic variables included age, sex, race/ethnicity (black vs. other), and education (less than high school graduate, high school graduate or more). Health conditions included self-reported mobility impairment (trouble climbing stairs or going outdoors), perceived health (good or excellent vs. fair or poor), and the self-reported presence or absence in the past 12 months of diabetes mellitus, stroke, heart trouble, asthma, high blood pressure, and shortness of breath. Other risk factors included cigarette smoking (current, former, or never) and body mass index (weight (kg)/height (m)<sup>2</sup>), which was used to divide subjects into weight tertiles. Social isolation was based on an isolated response to two of three questions: whether subjects had less than three close friends, less than three close relatives, and saw less than three of these close friends or relatives monthly. These risk factors have been shown to predict both

mortality and morbidity in Alameda County Study respondents and elsewhere (2, 9, 17–19).

**Mortality from 1965 through 1993.** California State mortality tapes were searched annually for information on study participants. Out-of-state deaths were ascertained during follow-up contact, in which case death certificates were requested from the state of residence. Death certificates were used to verify cause of death. Cardiovascular disease deaths were based on *International Classification of Diseases*, Eighth Revision, codes 390–458.

### Data analyses

Cox proportional hazards models, with and without time-dependent covariates, were used to analyze the relation between physical activity and subsequent mortality (20, 21). Deaths were included through 1993. Subjects known to have died and not lost to follow-up were kept in the analysis until the year of their death. Subjects not known to have died and not lost to follow-up were given a survival time of 28 years. Subjects lost to follow-up in either the 1974 or 1983 surveys were censored on the survey date when loss to follow-up occurred. For analyses of cardiovascular disease deaths, subjects who died of causes other than cardiovascular disease were censored in the year of their death. Calculations were performed using PROC PHREG in SAS, Version 6.09 (SAS Institute, Cary, North Carolina). Tied survival times were handled using the “discrete” option.

A sequence of three models was used. In the first model, age, sex, ethnicity, and education were included. In the second model, health conditions were added. The third model added smoking, body mass index, and social isolation.

Two sets of analyses were conducted. In the fixed covariate analyses, all values for physical activity and other risk factors were held constant at their 1965 baseline value. In the time-dependent covariate analyses, all variables, except age, sex, and ethnicity, were allowed to change based on data from subsequent survey periods. Thus, a subject who reported smoking cigarettes in 1965 but not in 1974 would be scored as a current smoker from 1965 to 1973, but a former smoker from 1974 until date of death or until results changed as a result of the next survey.

## RESULTS

### All-cause mortality

Basic descriptive data on the sample are presented in table 1. Table 2 presents all-cause and cardiovascular disease deaths by sex, person-years of follow-up, and crude mortality rates for each measure of physical

**TABLE 1. Baseline characteristics, Alameda County Study, Northern California**

Characteristic	
Age (years) (mean)	43.0
Sex (% male)	46.2
Race (% black)	11.9
Physical activity	
Exercise (% often)	17.6
Active sports (% often)	15.5
Long walks/swimming (% often)	16.5
Education (% <high school)	34.5
Mobility impairment (%)	5.3
Perceived health (% fair/poor)	17.9
Chronic conditions (mean no. out of six possible)	0.34
Smoking (% current)	45.0
Body mass index (kg/m <sup>2</sup> ) (mean)	23.8
Social isolation (%)	25.0

activity and the combined index. In all cases, the least active category has the highest mortality rate, and the rates generally show a consistent increasing gradient with decreasing amounts of activity. Table 3 presents the results for all-cause mortality separately for the fixed-covariate and time-dependent covariate models for the 1,226 deaths between 1965 and 1993. With all variables held constant at their baseline value, physical activity was protective with a four-unit increase, the interquartile range, associated with a 20 percent lower risk of death when only age, sex, ethnicity, and education were included (RR = 0.80, 95 percent CI 0.74–0.87). When health conditions were included, the relation between physical activity weakened somewhat but remained statistically significant (RR = 0.87, 95 percent CI 0.80–0.95). The addition of other risk factors had little further effect on the observed relation (RR = 0.90, 95 percent CI 0.83–0.99).

In the time-dependent covariate analyses, the relations between physical activity and mortality were stronger in all models compared to the fixed covariate models. With adjustment for age, sex, ethnicity, and education, a four-point increase in the physical activity scale was associated with 28 percent lower risk of death (RR = 0.72, 95 percent CI 0.65–0.78). In the model with all covariates, a four-point increase was associated with more than 50 percent greater protective effect (16 percent vs. 10 percent) in the time-dependent covariate analysis compared to the fixed covariate analysis.

### Cardiovascular disease mortality

Table 2 presents the crude results for cardiovascular disease mortality, and table 4 presents the results for the fixed-covariate and time-dependent models for cardiovascular disease mortality. There were 609 car-

diovascular disease deaths over the 28-year period. In all models, the relation between physical activity and mortality was stronger than it was for all-cause mortality, and the protective association diminished somewhat with adjustment for health status and other risk factors. While the protective effect of physical activity is stronger in the time-dependent covariate analyses, the differences between the fixed and time-dependent analyses are somewhat smaller than they were for deaths from all causes.

### Validity of covariates

The persistence of a protective effect of leisure-time physical activity with adjustment for baseline health status and risk factors as well as changes over time could reflect poor measurement of health status and other risk factors. As discussed earlier, risk factors included in the regression models were chosen because they were strong predictors of morbidity and mortality in other analyses using this cohort. In preliminary bivariate analyses, we verified that they all were strong and significant predictors of all-cause and cardiovascular disease mortality. In the baseline covariate regression model with all covariates simultaneously included, the following covariates were associated with all-cause mortality (relative risk with 95 percent CI): age (1.10, 1.09–1.10), sex (0.64, 0.61–0.72), perceived health (1.24, 1.07–1.44), mobility impairment (1.34, 1.10–1.63), heart trouble (1.34, 1.09–1.57), high blood pressure (1.27, 1.09–1.48), diabetes (1.98, 1.52–2.57), shortness of breath (1.23, 1.03–1.46), current smoking (1.85, 1.60–2.13), low body mass index (1.17, 1.01–1.36), and social isolation (1.17, 1.02–1.34). In the time-dependent model, stroke and former smoking were also significantly associated with all-cause mortality.

### Effect of early deaths

Baseline covariate analyses were repeated with elimination of deaths which occurred during the first 3 years of follow-up ( $n = 154$ ). The magnitude of the effect of physical activity changed only in the second and third decimal places.

### DISCUSSION

These analyses provide information to address the two methodological issues that we raised in the introduction. Taking into account changes in physical activity over the follow-up period resulted in a stronger protective effect for physical activity than when such changes were not considered. This is consistent with the smaller association in the fixed covariate model, reflecting misclassification over time of the level of

**TABLE 2. Baseline physical activity and all-cause and cardiovascular disease mortality, Alameda County Study, Northern California, 1965-1993**

Type of physical activity	Males					Females				
	Person-years	All-cause mortality		Cardiovascular disease mortality		Person-years	All-cause mortality		Cardiovascular disease mortality	
		No. of deaths	Crude mortality rate/1,000 person-years	No. of deaths	Crude mortality rate/1,000 person-years		No. of deaths	Crude mortality rate/1,000 person-years	No. of deaths	Crude mortality rate/1,000 person-years
Exercise										
Often	9,086	83	9.13	38	4.18	10,912	63	5.77	21	1.92
Sometimes	22,188	233	10.50	110	4.96	26,732	181	6.77	85	3.18
Never	16,342	323	19.77	173	10.59	20,024	343	17.13	182	9.09
Active sports										
Active	10,762	81	7.53	30	2.79	7,349	24	3.27	10	1.36
Sometimes	20,826	168	8.07	76	3.65	20,044	98	4.89	35	1.75
Never	16,028	390	24.33	215	13.41	30,274	465	15.36	243	8.03
Long walks/swimming										
Often	8,702	95	10.92	45	5.17	9,498	77	8.11	31	3.26
Sometimes	27,506	289	10.51	140	5.09	31,004	213	6.87	88	2.84
Never	11,408	255	22.35	136	11.92	17,166	297	17.30	169	9.85
Physical activity score*										
Tertile 3 (8-12)	14,104	107	7.59	42	2.98	13,140	51	3.88	15	1.14
Tertile 2 (4-6)	22,163	252	11.37	130	5.87	25,728	197	7.66	89	3.46
Tertile 1 (0-2)	11,349	280	24.68	149	13.13	18,798	339	18.03	284	15.11

\* The physical activity score was determined from a scale constructed from three questions on physical activity. Subjects were asked how often they did physical exercise, took part in active sports, and took long walks or swam. Responses to these questions were never, sometimes, or often and were scored as 0, 2, or 4, respectively. The items have been used in previous analyses from the Alameda County Study (2, 9, 17, 18).

**TABLE 3. Association between leisure-time physical activity and all-cause mortality with fixed and time-dependent covariates, Alameda County Study, Northern California, 1965-1993**

Model and covariates included	Fixed covariates			Time-dependent covariates		
	RR*	95% CI*	<i>P</i> value	RR	95% CI	<i>P</i> value
I. Age, sex, ethnicity, and education	0.80	0.74-0.87	<0.0001	0.72	0.65-0.78	<0.0001
II. Model I plus health conditions	0.87	0.80-0.95	<0.01	0.81	0.75-0.89	<0.0001
III. Model II plus health behaviors and social isolation	0.90	0.83-0.99	<0.05	0.84	0.77-0.92	<0.0001

\* RR, relative risk; CI, confidence interval. All relative risks and confidence intervals are based on interquartile range of leisure-time physical activity score.

**TABLE 4. Association between leisure-time physical activity and cardiovascular disease mortality with fixed and time-dependent covariates, Alameda County Study, Northern California, 1965-1993**

Model and covariates included	Fixed covariates			Time-dependent covariates		
	RR*	95% CI*	<i>P</i> value	RR	95% CI	<i>P</i> value
I. Age, sex, ethnicity, and education	0.73	0.65-0.83	<0.0001	0.67	0.59-0.77	<0.0001
II. Model I plus health conditions	0.82	0.72-0.93	<0.01	0.78	0.69-0.90	<0.001
III. Model II plus health behaviors and social isolation	0.85	0.75-0.97	<0.05	0.81	0.71-0.93	<0.001

\* RR, relative risk; CI, confidence interval. All relative risks and confidence intervals are based on interquartile range of leisure-time physical activity score.

physical activity. Furthermore, while health status and other factors that influence physical activity may change over time, taking these changes into account

does not eliminate the protective effect of physical activity for either all-cause or cardiovascular disease mortality.

To some extent, the results of the time-dependent covariate analyses may reflect the consequences of adjusting for both antecedent and intervening risk factors. For example, in the Alameda County Study, low levels of physical activity are associated with increased incidence of social isolation (15) and social isolation is associated with decreases in physical activity over time (12). Such interdependency of risk factors may be the rule rather than the exception, resulting in considerable complexity in understanding causal pathways in epidemiologic studies.

While we have adjusted for both baseline health status and updated measures of health status at each wave of data collection, it may still be possible, to some extent, that the results reflect the impact of health status of physical activity. However, in the baseline model in which the effect of health status on the physical activity association with mortality should be the greatest (because there is no subsequent updating of health status information and physical activity information), the effect of elimination of early deaths was trivial.

There are several features of the current analyses that require comment. Measurement of leisure-time physical activity is difficult, and the measure used in the current analyses, although significantly associated with mortality risk, is rather crude. It is likely that the comparison of the fixed and time-dependent analyses would be more dramatic with a better measure of physical activity, because better measures seem to show greater protective effects (7). Similarly, because of the long intervals between data collection, the current analyses involve, at the most, only three measures of physical activity. Thus, while these analyses go beyond previous ones, trajectories of physical activity and other risk factors over the follow-up period would be better assessed by more frequent data collection. Despite these concerns, the present analyses provide further support for the importance of physical activity, and indicate that the protective effect of physical activity is a robust one.

#### ACKNOWLEDGMENTS

Partially supported by grant no. 1R37AG11375 from the National Institute on Aging.

#### REFERENCES

- Paffenbarger RS, Hyde RT, Wing AL, et al. Physical activity, all-cause mortality, and longevity of college alumni. *N Engl J Med* 1986;314:605-13.
- Kaplan GA, Seeman TE, Cohen RD, et al. Mortality among the elderly in the Alameda County Study: behavioral and demographic risk factors. *Am J Public Health* 1987;77:307-12.
- Powell KE, Thompson PD, Casperson CJ, et al. Physical activity and the incidence of coronary heart disease. *Ann Rev Public Health* 1987;8:253-87.
- Blair SN, Goodyear NN, Gibbons LW, et al. Physical activity and incidence of hypertension in healthy normotensive men and women. *JAMA* 1984;252:487-90.
- Helmrich SP, Ragland DR, Leung RW, et al. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med* 1991;325:147-52.
- Camacho TC, Roberts RE, Lazarus NB, et al. Physical activity and depression: evidence from the Alameda County Study. *Am J Epidemiol* 1991;134:220-31.
- Pate RR, Pratt M, Blair SN, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA* 1995;273:402-7.
- Fletcher GF, Blair SN, Blumenthal J, et al. AHA medical/scientific statement on exercise. *Circulation* 1992;86:340-4.
- Kaplan GA, Haan MN. Is there a role for prevention among the elderly? Epidemiologic evidence from the Alameda County Study. In: Ory MG, Bond K, eds. *Aging and health care: social science and policy perspectives*. London: Routledge, 1989:27-51.
- Paffenbarger RS, Kampert JB, Lee I-M, et al. Changes in physical activity and other lifeway patterns influencing longevity. *N Engl J Med* 1993;328:538-45.
- Blair SN, Kohl HW, Barlow CE, et al. Change in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *JAMA* 1995;273:1093-8.
- Kaplan GA, Lazarus NB, Cohen RD, et al. Psychosocial factors in the natural history of physical activity. *Am J Prev Med* 1991;7:12-17.
- Lazarus NB, Kaplan GA, Cohen RD, et al. Smoking and body mass in the natural history of physical activity: prospective evidence from the Alameda County Study. *Am J Prev Med* 1989;5:127-35.
- Salonen JT, Slater JS, Tuomilehto J, et al. Leisure time and occupational physical activity: risk of death from ischemic heart disease. *Am J Epidemiol* 1988;127:87-94.
- Kaplan GA. Reflections on present and future research on bio-behavioral risk factors. In: Blumenthal SJ, Matthews K, Weiss SM, eds. *New research frontiers in behavioral medicine: proceedings of a national conference*. Washington, DC: US GPO, 1994:119-34. (NIH publication no. 94-3772).
- Hochstim JR. Health and ways of living: the Alameda County population laboratory. In: Kessler II, Levin ML, eds. *The community as an epidemiologic laboratory*. Baltimore, MD: John Hopkins University Press, 1970:149-76.
- Berkman LF, Breslow L. *Health and ways of living: the Alameda County Study*. New York: Oxford University Press, 1983.
- Kaplan GA, Cohn BA, Cohen RD, et al. The decline in ischemic heart disease mortality: prospective evidence from the Alameda County Study. *Am J Epidemiol* 1988;127:1131-42.
- Kaplan GA, Kotler PL. Self-reports predictive of mortality from ischemic heart disease: a nine-year follow-up of the Human Population Laboratory cohort. *J Chronic Dis* 1985;38:195-201.
- Cox DR, Oakes D. *Analysis of survival data*. New York: Chapman & Hall, 1984.
- Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. New York: John Wiley and Sons, 1980.