# Risk factors for prostate cancer incidence and progression in the health professionals follow-up study

**Edward Giovannucci**[1,2,3,*], **Yan Liu**[2], **Elizabeth A. Platz**[4], **Meir J. Stampfer**[1,2,3], and **Walter C. Willett**[1,2,3]

1 *Channing Laboratory, Department of Medicine, Harvard Medical School and Brigham and Women's Hospital, Boston, MA 02115*

2 *Department of Nutrition, Harvard School of Public Health, Boston, MA 02115*

3 *Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115*

4 *Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205*

## Abstract

Risk factors for prostate cancer could differ for various subgroups, such as for "aggressive" and "non-aggressive" cancers or by grade or stage. Determinants of mortality could differ from those for incidence. Using data from the Health Professionals Follow-Up Study, we re-examined 10 factors (cigarette smoking history, physical activity, BMI, family history of prostate cancer, race, height, total energy consumption, and intakes of calcium, tomato sauce and α-linolenic acid) using multivariable Cox regression in relation to multiple subcategories for prostate cancer risk. These were factors that we previously found to be predictors of prostate cancer incidence or advanced prostate cancer in this cohort, and that have some support in the literature. In this analysis, only 4 factors had a clear statistically significant association with overall incident prostate cancer: African–American race, positive family history, higher tomato sauce intake (inversely) and α-linolenic acid intake. In contrast, for fatal prostate cancer, recent smoking history, taller height, higher BMI, family history, and high intakes of total energy, calcium and α-linolenic acid were associated with a statistically significant increased risk. Higher vigorous physical activity level was associated with lower risk. In relation to these risk factors, advanced stage at diagnosis was a good surrogate for fatal prostate cancer, but high-grade (Gleason ≥ 7 or Gleason ≥ 8) was not. Only for high calcium intake was there a close correspondence for associations among high-grade cancer, advanced and fatal prostate cancer. Tomato sauce (inversely) and α-linolenic acid (positively) intakes were strong predictors of advanced cancer among those with low-grade cancers at diagnosis. Although the proportion of advanced stage cancers was much lower after PSA screening began, risk factors for advanced stage prostate cancers were similar in the pre-PSA and PSA era. The complexity of the clinical and pathologic manifestations of prostate cancer must be considered in the design and interpretation of studies.

### Keywords

prostate cancer; risk factors

The wide international variation in incidence rates[1,2] and increases in prostate cancer incidence and mortality rates in migrants from countries with low rates to those countries with high rates[3–5] demonstrate the importance of modifiable etiologic factors for this cancer. However,

*Correspondence to: Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115. Fax: 617-432-2435., E-mail: edward.giovannucci@channing.harvard.edu.

the evidence for any specific factor has generally not been very consistent. The heterogeneous nature of prostate cancers, which range from relatively innocuous to highly aggressive in behavior, may contribute to inconsistent results. Because they may act on different biologic pathways, risk factors may be different for various sub-groups of prostate cancer, such as for "aggressive" and "non-aggressive" cancers, defined by grade, stage, or survival. The premise, usually implicit, that risk factors for initiation of relatively innocuous, well-differentiated prostate cancers should be the same as those that cause death from prostate cancer has little theoretical or empirical basis. Further, although many epidemiologic studies now combine cancers of advanced stage at the time of diagnosis and those with high Gleason grade to characterize "aggressive" prostate cancer, this practice implicitly supposes that grade—which reflects degree of differentiation—carries the same meaning as advanced stage, but a risk factor could influence the progression of a cancer independently of an effect on tumor grade. Thus, results across studies could vary depending on the specific prostate cancer sub-type examined.

In the Health Professionals Follow-Up Study, we have reported on various lifestyle and nutritional factors in relation to risk of prostate cancer. Most of the reports have addressed single factors, and focused on incident cancer, or advanced disease at diagnosis as an indicator of aggressive disease. In this report, we consider systematically 10 risk factors with cases from 1986 to 2002 in relation to various prostate cancer sub-types, specifically, total incident prostate cancer, fatal prostate cancer, advanced or non-advanced stage at presentation, high-grade or low-grade prostate cancer, and some combinations of stage and grade. We examined in multivariable analysis 10 factors that we previously found to be predictors of prostate cancer incidence or advanced prostate cancer, and that have some support in the literature. These include cigarette smoking history,[6] race,[7] family history of prostate cancer, physical activity, [8] body mass index (BMI),[9] height,[9] total energy consumption,[10] and intakes of calcium,[11] tomato sauce[12] and α-linolenic acid.[13,14] Further, we considered how alternative definitions of clinically advanced prostate cancer influenced the results. In addition, we considered the potential influence that PSA screening may have on identifying risk factors for total prostate cancer and advanced prostate cancer, because PSA screening advances the time of diagnosis and alters the spectrum of cancers that are diagnosed. Finally, we considered if risk factors differed for earlier-age onset and late-onset prostate cancer.

## Methods

### The study population

The Health Professionals Follow-up Study cohort was initiated in 1986, when 51,529 U.S. male health professionals, ages 40–75, completed a mailed questionnaire on age, marital status, height and weight, ancestry, medications, smoking history, disease history, leisure-time physical activity, diet (in the past year) and use of vitamin and mineral supplements. This cohort is predominantly Caucasian (>91%). Through biennial follow-up mailed questionnaires, we updated information on disease history, smoking history, body weight and physical activities, and every 4 years, we updated dietary information. We identified deaths in this cohort through information from family members, the postal system and the National Death Index. Through these methods, we ascertained at least 98% of the total deaths.[15] The conduct of this cohort study and these analyses was approved by the Human Subjects Committee of the Harvard School of Public Health.

### Identification of cases of prostate cancer

On the follow-up questionnaires, participants were asked to report new diagnoses of cancer. For each new report of prostate cancer, we re-contacted the study participant to acquire permission to obtain relevant hospital records and pathology reports. For men who had died, we contacted next of kin. The medical records were reviewed by study investigators, who were

blinded to any exposure information. For staging, we used information from the questionnaires derived from any procedures or tests conducted during the initial diagnosis and treatment, including prostatectomies and bone scans. The tumor-node-metastasis (TNM) system was used.[16] We excluded T1a cancers because these are relatively innocuous and may be especially prone to detection bias related to treatment for benign prostatic hyperplasia. In terms of total potential person-time accumulated up to the end of the study period (January 31, 2002), the response rate to follow-up questionnaires was 96%. We categorized a death as due to prostate cancer when there was evidence of extensive metastatic prostate cancer, and no other plausible cause of death, based on a review of the medical records by a study physician. Cause of death was rarely based on death certificate alone, when the primary medical records were not obtainable.

### The food-frequency questionnaire

Diet was assessed through a semiquantitative food-frequency questionnaire; the methodology and validation of the questionnaire was previously described in detail.[17] The questionnaire, initially administered in 1986, contained a list of 131 food and beverage items, and included an open-ended section. We specified a commonly used unit or portion size for each item on the list, and asked the participant how often, on average, over the past year, he consumed the specified amount of each item. For each item, the participant chose from among 9 possible responses for frequencies. In addition, we assessed current use, brand, dose and duration of use of supplements, including calcium supplements. Information based on specific brand and type of multivitamin was taken into account when calculating nutrient intakes. In 1990, 1994 and 1998, we re-administered the dietary questionnaire, which had slight modifications to take into account changes in the food supply that occurred over time. We computed nutrient intakes by multiplying the consumption frequency of each unit of food by the nutrient content of the specified portions, using composition values from U.S. Department of Agriculture sources supplemented with other data. We evaluated the validity of nutrient and food consumption measured by the questionnaire among a sample of 127 cohort members from the Boston area. [17,18] After adjusting for week-to-week variation in dietary records, the median correlation for the major nutrients between the dietary record and the questionnaire assessment was 0.65.

### Assessment of PSA utilization and other non-dietary exposure information

On the 1986 baseline questionnaire, the men reported their current height and weight, and weight at age 21. Current weight was also assessed on each biennial questionnaire. We used BMI ($kg/m^2$) to estimate total adiposity.[19] To assess physical activity, we used MET-hours/week calculated from a list of activities reported in 1986 and updated biennially thereafter. [20] One MET-hour is the metabolic equivalent of sitting at rest for 1 hr. Detailed lifetime tobacco use was assessed at baseline and current smoking status biennially. We initially asked about screening history for prostate cancer by PSA test in 1994, and thereafter, inquired about PSA testing biennially for the prior 2-year period. In 1990 and 1996, we asked each man whether his biologic father or any brothers had ever had a diagnosis of prostate cancer.

### Choice of risk factors for multivariable analysis

We examined in multivariable analysis 10 factors that we previously found to be risk factors for prostate cancer incidence or advanced prostate cancer in the Health Professionals Follow-Up Study. These include recent smoking history,[6] race,[7] family history of prostate cancer, vigorous physical activity,[8] BMI,[9] height,[9] total energy consumption,[10] and intakes of calcium,[11] tomato sauce,[12] and α-linolenic acid.[13,14] For all these factors, there is additional support for their potential importance besides the Health Professionals Follow-Up Study, although data may be inconsistent.[21] We also included in the models BMI at age 21 years and updated history of diabetes mellitus, as well as processed meat, which is a risk factor in this

cohort,[22] but has less support in general. Age was adjusted in 1-month time intervals and time period in 2-year intervals. Vitamin E supplements were also included in the model, but as in our previous analysis,[23] no overall association was observed except for an inverse association for advanced cancer limited to men who had smoked within the previous 10 years.

## Statistical analysis

At baseline, we excluded men with diagnosed cancer (except for non-melanoma skin cancer), and men who did not adequately complete a food frequency questionnaire (3% of the total). We excluded these men because the presence of another cancer could result in behavioral, dietary or pathophysiologic changes (for example, weight loss), and it could complicate the assignment of metastatic death to prostate cancer. These exclusions left 47,750 men, each of whom accrued follow-up time beginning on the month of return of the baseline questionnaire and ending on the month of diagnosis of prostate cancer, month of death from other causes, or January 31, 2002, whichever came first. For prostate cancer sub-types, we considered total, advanced stage (at diagnosis), non-advanced stage, high-grade (Gleason $\geq 7$), low-grade (Gleason $<7$), and fatal prostate cancer. We calculated rates for men in a specific category of an exposure by dividing the number of incident cases by the number of person-years.

Advanced stage cases were considered those that had invaded other organs, including the seminal vesicle or lymph nodes, as well as distant organs at the time of diagnosis, or that were fatal by January 31, 2002. In our analyses, we typically include N0 M0 cancers with seminal vesicle invasion (stage T3b or T4 or C2) as advanced, but those with invasion only through the capsule (stage C1, or T3a) as non-advanced because seminal vesicle involvement is a strong predictor of clinical failure but capsular invasion alone may not be.[24] Because the stage C1 (or T3a) can arguably be considered as either advanced or non-advanced, we analyzed these separately to examine whether, in relation to risk factors, these were more similar to advanced or to non-advanced cancers. Non-advanced cancers were considered to be those characterized as T1 or T2 and N0 M0, but excluding stage T1a (incidental histologic cancer found in 5% or less of tissue resected for benign prostatic hyperplasia).

For nutrients and foods in the models (see above), to best assess long-term cumulative exposure intake, we used cumulative updating, which uses 1986 diet assessment to assess risk from 1986 to 1990, the average of 1986 and 1990 assessments to assess risk from 1990 to 1994, and the average of 1986, 1990 and 1994 intakes to assess risk from 1994 to 1998, and the average of 1986, 1990, 1994 and 1998 intakes to assess risk from 1998 to 2002.[25] We used a similar cumulative average approach for BMI and physical activity level. For both nonfatal and fatal prostate cancer, we updated exposure data up until the time of the diagnosis.

We computed relative risks (RR) defined as the incidence rate of disease in one category (e.g. high category of calcium intake) divided by the incidence rate in the specified reference category (e.g. low category of calcium intake). We used Cox proportional hazards modeling to control for multiple time-varying variables simultaneously and to compute 95% confidence intervals (CI). Age was controlled using 1-month categories and calendar time in 2-year intervals as stratified variables in the Cox models. We tested for trend across categories controlling for multiple covariates by modeling the median values of the variable as a continuous variable in the multivariable model. In addition, we examined the risk factors in relation to total and advanced prostate cancer in 2 time periods, the pre-PSA screening era, defined here as beginning in February 1986 (start of the study) until January 31, 1992, and the PSA screening era, beginning in February 1, 1992 until the end of follow-up, January 31, 2002. All reported $p$-values are two-sided. $p$-values $\leq 0.05$ were considered to be statistically significant.

# Results

From 1986 to the end of the study period, January 31, 2002, in 673,706 person years, we documented 3,544 incident prostate adenocarcinoma cases after excluding 71 cases (about 2% of total) of stage T1a cancers. We documented 90% of the 3,544 cases through medical records and pathology reports, and the remaining 10% by supporting data for the diagnosis that participants provided (typically, evidence of treatment). Based on the pathology report, we had information on Gleason sum for 2,701 cases (76.2% of the total). We documented 1,110 high-grade cancers defined as Gleason $\geq 7$ (and 322 as Gleason $\geq 8$), and 1,601 low-grade cases (Gleason $\leq 6$). Also, 523 cases were considered to be advanced stage or fatal, of which 312 cases were defined as fatal prostate cancer, having been the underlying cause of death, by January 31, 2002.

## Incident and fatal prostate cancer

We first examined each factor in relation to incident prostate cancer and fatal prostate cancer. The results are displayed in Figure 1. Only 4 factors had a statistically significant association with incident prostate cancer: higher tomato sauce intake was associated with a decreased risk, and African–American race, a positive family history of prostate cancer, and higher α-linolenic acid intake were associated with an increased risk. In contrast, most items were significantly associated with risk of fatal prostate cancer. Specifically, recent smoking history, positive family history of prostate cancer, taller height, higher BMI, and high intake of total energy, calcium and α-linolenic acid were associated with a statistically significant increased risk, whereas higher vigorous physical activity level was associated with lower risk of fatal prostate cancer. For tomato sauce, the magnitude of the inverse association was similar for fatal as for total prostate cancer, and the smaller numbers may have prevented us from detecting a statistically significant association with fatal prostate cancer.

## Advanced and non-advanced stage prostate cancer

The results for advanced and non-advanced prostate cancer were largely similar as those for fatal and incident prostate cancer respectively (see Table I). However, some differences were notable. While smoking within the past 10 years was associated with a higher risk of fatal prostate cancer, it was not significantly associated with risk of advanced prostate cancer. In addition, high tomato sauce intake was associated with a decreased risk of advanced prostate cancer, although the test for linear trend was not statistically significant (RR = 0.66; 95% CI = 0.44–1.00, for intake $\geq 2$ *vs.* <0.25 servings/week; *p*(trend) = 0.12). Although more frequent vigorous physical activity was strongly associated with a decreased risk of advanced and fatal prostate cancer, it was associated with a modest but statistically significant increased risk of non-advanced prostate cancer (RR = 1.19; 95% CI = 1.03–1.37, for high versus low quintile; *p*(trend) < 0.01). For N0 M0 prostate cancers with capsular invasion alone (stage C1 or T3a), none of the risk factors was associated significantly with these cancers except for family history of prostate cancer, and suggestively for calcium (positively) and tomato sauce (inversely). Thus, this endpoint in general (except possibly for calcium intake) appears to respond to risk factors similarly to the non-advanced as opposed to advanced cancers.

## High-grade and low-grade prostate cancer

We next examined high-grade and low-grade prostate cancer separately. Risk factors for high-grade prostate cancer (Fig. 2) were not consistently reflected in relation to fatal prostate cancer (Fig. 1). Analyses defining high-grade as Gleason $\geq 8$ gave similar results as those with high-grade defined as Gleason $\geq 7$ but with wider confidence intervals because of the much smaller number of cases (data not shown). Only high calcium intake exhibited a close correspondence for high-grade cancer, advanced and fatal prostate cancer. Taller height was also associated with an increased risk of high-grade prostate cancer, but the magnitude of the association was

much weaker than that for fatal prostate cancer and it was non-monotonic. African–American race was associated with a significantly higher risk of high-grade prostate cancer, but not with low-grade prostate cancer. Similar to its association with non-advanced prostate cancer, vigorous activity was associated with a modest but statistically significant greater risk of low-grade prostate cancer. α-linolenic acid intake was positively and tomato sauce intake was inversely associated with risk of low-grade prostate cancer. In a separate analysis, we examined cancers that were high-grade but non-advanced at the time of diagnosis ($n = 860$ cases). None of the variables indicated a significant trend; high intake of calcium was suggestively associated with an increased risk (multivariable RR = 1.51, 95% CI = 0.95–2.41). In a case-only analysis, trends for a higher likelihood of a diagnosed cancer being high-grade versus low-grade were evident only for higher calcium intake ($p = 0.002$), taller height ($p = 0.01$), and marginally for less vigorous physical activity ($p = 0.06$).

## Low-grade, advanced prostate cancer

Poor differentiation (*i.e.*, high Gleason grade) is a well-established strong determinant of progression and fatality from prostate cancer. However, a small proportion of lower grade prostate cancers do progress to advanced stage and lead to mortality. In our data, only 83 cases were defined as low-grade and advanced stage. Higher intakes of α-linolenic acid (multivariable RR = 2.23, 95% CI = 1.11–4.48; $p$(trend) = 0.04) and tomato sauce (multivariable RR = 0.27, 95% CI = 0.10–0.96; $p$(trend) = 0.02), which were associated with overall low-grade prostate cancer, had especially strong relationships with low-grade advanced prostate cancers, suggesting that these factors may influence progression of well-and moderately-well differentiated lesions into advanced stages. No other factor was associated with this endpoint.

## Effect modification by time period as a surrogate of PSA era

The widespread use of PSA screening in the United States beginning around the early 1990's had a profound influence on the apparent incidence of prostate cancer, as well as the spectrum of case types being diagnosed. We examined the risk factors in relation to total and advanced prostate cancer in the pre-PSA screening era and the PSA screening era. As shown in Table II, the proportion of advanced cases decreased dramatically in the PSA era (from 27.5% to 10.4%). The association patterns for total and for advanced cancers were largely similar for the 2 time periods, with several exceptions: calcium intake was associated with increased risk of total prostate cancer in the earlier era but not in the PSA era, and the associations with family history of prostate cancer and African–American race tended to be stronger for both total and advanced prostate cancer in the pre-PSA era.

## Effect modification by age

We also examined if risk differed by age group, using the median age of cases as the cutpoint (<68 years; ≥68 years). For total prostate cancer, only for BMI was there strong statistical evidence of an age interaction ($p = 0.001$); for younger men, an inverse association was observed (RR = 0.81, 95% CI = 0.58–1.13 for BMI ≥ 30 versus <21 kg/m$^2$, $p$(trend) = 0.01), whereas for older men, a positive association was observed (RR = 1.47; 95% CI = 1.10–1.96, $p$(trend) = 0.02). For fatal prostate cancer, these respective RR's for younger and older men were RR = 0.88; 95% C = 0.41–1.91, $p$(trend) = 0.12 and RR = 2.73; 95% C = 1.44–5.14, $p$(trend) = 0.03. For physical activity, the result for interaction for fatal cancer was not statistically significant based on a test for trend, but the inverse association was stronger for older men for the high versus low quintile (RR = 0.78, 95% CI = 0.36–1.66 for men <68 years and RR = 0.46, 95% CI = 0.21–1.00 for men ≥68 years).

## Discussion

In the Health Professionals Follow-Up Study (summarized in Table III) we found only low tomato sauce and high α-linolenic acid intakes, a positive family history of prostate cancer, African–American race and possibly calcium intake were associated with increased incidence, but additional factors were associated with fatal prostate cancer. These results suggest that reducing mortality from prostate cancer through lifestyle and diet may generally be more feasible than preventing its occurrence. Because prostate cancers probably have a multiple decades-long induction or latent phase, initiating and early-acting factors may have occurred many years prior to our follow-up, whereas progression factors are operative late and thus are more easily identifiable in a cohort of middle-aged and elderly men. On the other hand, tallness, presumably a surrogate of a factor acting during the growth period (*e.g.,* adolescent IGF-1 levels) strongly predicted prostate cancer mortality but did not predict incidence; this finding suggests early-operative carcinogenic processes could affect aggressive behavior in cancers, or increase the incidence of a sub-set of prostate cancer with a greater propensity to progress.

We speculate that pathways through which specific factors may influence death from prostate cancer may be quite diverse, and may involve: (*i*) increasing incidence (*e.g.* high α-linolenic acid, low tomato sauce intake, family history of prostate cancer, African–American race), (*ii*) increasing the likelihood of poor differentiation (as assessed by Gleason score) in a prostate cancer (*e.g.* calcium, and possibly tallness and African–American race), (*iii*) increasing promotion or progression preferentially of better differentiated prostate cancers (*e.g.* high α-linolenic acid, low tomato sauce intake) into advanced stages, and (*iv*) by increasing mortality independently of incidence and grade (*e.g.* high BMI, physical inactivity, cigarette smoking). If all or even a sub-set of these patterns whereby risk factors increase risk are causal, one would predict substantial heterogeneity in the prostate cancer literature. Different susceptibility to risk factors based on age may additionally contribute to heterogeneity.

Potential mechanisms underlying these associations have been discussed in our previous publications, and we consider here in general terms the plausibility of mechanisms acting through these endpoints. That any factor that increases the incidence of a cancer could ultimately increase the mortality is intuitive, unless the factor only increased risk of a sub-set of lesions that are defined as cancers histologically, but that could not progress further. The existence of such a sub-set of tumors is entirely plausible, in view of the natural history of prostate cancer and the enormously high prevalence of apparently innocuous disease at older ages. Influencing cancer risk by affecting differentiation status is also a plausible mechanism. For example, such a role has been suggested for vitamin D, a differentiation agent.[26,27] In contrast, some risk factors may preferentially influence the promotion or progression of well-differentiated lesions. Interestingly, as for tomato sauce and α-linolenic acid, in the Physicians' Health Study we had previously shown that high circulating IGF-1 levels were preferentially associated with increased risk of low-grade but advanced prostate cancers,[28] a finding confirmed in preliminary analyses in the Health Professionals Follow-Up Study and by others. [29] A possible explanation is that growth in poorly differentiated cancers may be more autonomous because these cancers may have extensive mutations in the IGF-1 growth factor signaling pathway.[30,31] In contrast, better-differentiated cancers may have a relatively intact signaling pathway responsive to circulating IGF-1 levels. Finally, that some factors would influence survival (and thus mortality) independently of incidence or tumor grade would not be surprising; examples may be through effects on angiogenesis, or on progression of metastatic clones.

The findings for BMI illustrate the complexity of studying prostate cancer. Our current findings are consistent with our previous reports of no overall association with total prostate cancer, but with an indication of an inverse association in younger men and a suggestive positive

association in older men.[9,32] In previous analyses, we did not have adequate power to study fatal prostate cancer; in the current analysis, we found BMI positively associated with risk of fatal prostate cancer, especially in older men. Most studies have not shown an association between BMI and incident prostate cancer,[33,34] but an association with fatal prostate cancer has been observed more frequently.[35–38] The mechanisms underlying the heterogeneity by age are unknown, but may relate to specific hormones, such as testosterone, estrogen, insulin, and IGF-1, that change in concentration over the lifespan and that are influenced by BMI or influence BMI.[32]

Chance seems unlikely as the primary explanation because each factor considered has been consistently associated with risk in multiple time periods throughout our follow-up period. For bias, differential detection through PSA screening is of most concern, as groups with more frequent screening may be more likely to have a cancer detected, especially at an earlier stage. In general, PSA screening intensity is very high in this population and fairly equal across categories of various exposures; the percentage of men reporting having had a PSA test in the prior 2 years asked biennially from 1994 to 2000 for the low and high categories, respectively, for the following exposures are: height: 74 and 75%; BMI: 70 and 73%; vigorous activity: 74 and 77%; calories: 77 and 72%; calcium: 71 and 76%; α-linolenic acid: 74 and 75%; tomato sauce: 74 and 75%; 75% for never smokers and 70% for current smokers; family history, no: 74% and family history yes: 81%; Caucasian: 75% and African–American: 77%. These modest differences in PSA screening were unlikely to produce substantial bias. Perhaps the most compelling argument against detection bias is that the associations with advanced and fatal prostate cancer were observed prior to widespread PSA screening, or before PSA screening could plausibly affect mortality. Finally, confounding cannot be ruled out in observational studies. An advantage of our cohort is that detailed data on many factors allowed us to consider confounding carefully. However, residual or uncontrolled confounding could still occur, especially with error in measuring exposures and covariates. Yet, even if confounding did account for some of our findings (for example, if α-linolenic acid were acting as a marker of another type of fat), our general conclusions regarding the differential effects of risk factors on various endpoints would remain valid. The possibility that bias or confounding may have contributed to our findings has been further considered in detail for each factor in our previous publications.

A limitation of our study is that to assess Gleason score, we relied on pathology reports, which may result in measurement error. Further, Gleason grading may have undergone secular changes.[39] However, our results were similar when we excluded cases with Gleason score of 7, and when we considered cancers with scores of 8 or greater as high-grade. More importantly, one of our major goals for this analysis was to assess the utility of various endpoints in epidemiologic studies, so our findings are directly relevant to epidemiologic studies that use pathology reports to assign Gleason score. Both inherent limitations in the Gleason score to predict death, as well as measurement error in assigning the Gleason score, ultimately are relevant factors regarding its utility.

The findings have important implications for the design and interpretations of studies of prostate cancer. First, in cancer epidemiology, incidence has typically been considered as a more relevant endpoint than mortality, but this may not be true for prostate cancer. Studies focusing primarily on incidence may miss important associations with fatal prostate cancer. This limitation of total prostate cancer incidence as an endpoint is exacerbated by PSA screening, leading mainly to diagnosis of non-advanced prostate cancer cases, with few deaths. These include a small proportion of cases that are not advanced at time of diagnosis, but have lethal potential if untreated. Such cases cannot be readily distinguished from innocuous cancers. Second, results from analyses of cancers with strong indicators of disease progression at the time of diagnosis (*e.g.* seminal vesicle or lymph node involvement) parallel results for

fatal prostate cancer, but these clinical manifestations are also becoming rare in the PSA era. Extension beyond the prostate but not into adjacent organs does not appear to be an adequate marker of aggressive prostate cancer; in many studies conducted in the PSA era, these may comprise a substantial proportion of lesions that are defined as advanced.

Third, some factors may act on the progression of well-differentiated prostate cancers, but may not influence poorly-differentiated lesions. This finding, if confirmed, could have some important practical implications. Well-differentiated prostate cancers (for example, Gleason score of 6 or lower) have a relatively good prognosis, but a proportion of these do progress leading to difficult decisions regarding treatment. Some interventions, such as tomato sauce or lycopene, could potentially act to reduce progression of low-grade cancers. Finasteride may be an example of an agent that inhibits (or even regresses) growth in better differentiated lesions, but not for poorly differentiated lesions.[40] If the propensity for progression could be further reduced in these cancers, perhaps treatment could be deferred or even averted. However, treatments that influence incidence or progression of primarily well-differentiated cancers may do little to prevent mortality from poorly differentiated cancers, which account for the majority of fatal cancers.

A fourth implication is that although high-grade, non-advanced prostate cancers are relatively common in the PSA era, and are often used as an endpoint to characterize aggressive prostate cancer, the pattern of most risk factors for high-grade, non-advanced cancers is not the same as that for fatal prostate cancer. Even though high-grade cancers are more likely to progress than low-grade cancers, the majority of high-grade cancers are not fatal, especially those diagnosed in the PSA era. Moreover, lesions that do progress may take many years to do so, and the endpoint of organ-confined, high-grade cancers does not necessarily help identify most risk factors for progression. Calcium intake, the factor most strongly associated with high-grade *vs.* low-grade cancer, was associated with advanced prostate cancer in both the pre-PSA and the PSA eras, but with total prostate cancer only in the pre-PSA era (Table II). A possible reason is that in the pre-PSA era, a much higher proportion of cancers were diagnosed due to clinical progression, for which high-grade is an important determinant. In the PSA era, these cancers may have been largely diluted by the high prevalence of PSA-detected cancers. Interestingly, we had previously shown that high calcium intake was associated with a non-significant inverse association with organ-confined, low-grade prostate cancer in the PSA era.[41] A fifth implication is that heterogeneity by age at risk may occur for some risk factors, most notably BMI and possibly physical activity. Other sources of heterogeneity that we did not assess are possible.

The etiology of prostate cancer remains enigmatic. Results of many studies have been conflicting, especially in the PSA era. Currently, studies relying solely on incidence may have limited applicability to identifying means to prevent dying from prostate cancer. Using Gleason sum to characterize aggressiveness may be informative in some contexts, but generally does not seem useful in pointing to risk factors that influence disease progression. Fatal and advanced stage prostate cancer may be informative endpoints, although advanced stage should be based on clear indicators, such as invasion into the seminal vesicle or other regional structures or metastasis to the lymph nodes, bone, or other organs. The complexity of the clinical and pathologic manifestations of prostate cancer must be considered in the design and interpretation of studies.

## References

1. Parkin, DM.; Whelan, SL.; Ferlay, J.; Teppo, L.; Thomas, D. Cancer incidence in five continents, International agency for research on cancer. 155. Lyon: IARC Scientific Publications; 2003.

2. Parkin, DM.; Muir, CS.; Whelan, SL.; Gao, YT.; Ferlay, J.; Powell, J. Cancer incidence in five continents, International agency for research on cancer. Lyon, France: IARC Scientific Publications; 1992.

3. Haenszel W, Kurihara M. Studies of Japanese migrants. I Mortality from cancer and other diseases among Japanese in the United States. J Natl Cancer Inst 1968;40:43–68. [PubMed: 5635018]

4. Yu H, Harris RE, Gao YT, Gao R, Wynder EL. Comparative epidemiology of cancers of the colon, rectum, prostate and breast in Shanghai, China versus the United States. Int J Epidemiol 1991;20:76–81. [PubMed: 2066247]

5. Shimizu H, Ross RK, Bernstein L, Yatani R, Henderson BE, Mack TM. Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles county. Br J Cancer 1991;63:963–6. [PubMed: 2069852]

6. Giovannucci E, Rimm EB, Ascherio A, Colditz GA, Spiegelman D, Stampfer MJ, Willett WC. Smoking and risk of total and fatal prostate cancer in United States health professionals. Cancer Epidemiol Biomarkers Prev 1999;8:277–82. [PubMed: 10207628]

7. Platz EA, Rimm EB, Willett WC, Kantoff PW, Giovannucci E. Racial variation in prostate cancer incidence and in hormonal system markers among male health professionals. J Natl Cancer Inst 2000;92:2009–17. [PubMed: 11121463]

8. Giovannucci E, Leitzmann M, Spiegelman D, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. A prospective study of physical activity and prostate cancer in male health professionals. Cancer Res 1998;58:5117–22. [PubMed: 9823321]

9. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Willett WC. Height, body weight, and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev 1997;6:557–63. [PubMed: 9264267]

10. Platz EA, Leitzmann MF, Michaud DS, Willett WC, Giovannucci E. Interrelation of energy intake, body size, and physical activity with prostate cancer in a large prospective cohort study. Cancer Res 2003;63:8542–8. [PubMed: 14679023]

11. Giovannucci E, Rimm EB, Wolk A, Ascherio A, Stampfer MJ, Colditz GA, Willett WC. Calcium and fructose intake in relation to risk of prostate cancer. Cancer Res 1998;58:442–7. [PubMed: 9458087]

12. Giovannucci E, Ascherio A, Rimm EB, Stampfer MJ, Colditz GA, Willett WC. Intake of carotenoids and retinol in relation to risk of prostate cancer. J Natl Cancer Inst 1995;87:1767–76. [PubMed: 7473833]

13. Leitzmann MF, Stampfer MJ, Michaud DS, Augustsson K, Colditz GA, Willett WC, Giovannucci E. Dietary intake of omega-3 and omega-6 fatty acids and the risk of prostate cancer. Am J Clin Nutr 2004;80:204–16. [PubMed: 15213050]

14. Giovannucci E, Rimm EB, Colditz GA, Stampfer MJ, Ascherio A, Chute CG, Willett WC. A prospective study of dietary fat and risk of prostate cancer. J Natl Cancer Inst 1993;85:1571–9. [PubMed: 8105097]

15. Stampfer MJ, Willett WC, Speizer FE, Dysert DC, Lipnick R, Rosner B, Hennekens CH. Test of the national death index. Am J Epidemiol 1984;119:837–9. [PubMed: 6720679]

16. Schroder FH, Hermanek P, Denis L, Fair WR, Gospodarowicz MK, Pavone-Macaluso M. The TNM classification of prostate cancer. Prostate Suppl 1992;4:129–38. [PubMed: 1574453]

17. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of a expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135:1114–26. [PubMed: 1632423]

18. Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB, Willett WC. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. J Am Diet Assoc 1993;93:790–6. [PubMed: 8320406]

19. Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett WC. Validity of self-reported waist and hip circumferences in men and women. Epidemiology 1990;1:466–73. [PubMed: 2090285]

20. Chasan-Taber S, Rimm EB, Stampfer MJ, Spiegelman D, Colditz GA, Giovannucci E, Ascherio A, Willett WC. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. Epidemiology 1996;7:81–6. [PubMed: 8664406]

21. Chan JM, Gann PH, Giovannucci EL. Role of diet in prostate cancer development and progression. J Clin Oncol 2005;23:8152–60. [PubMed: 16278466]

22. Wu K, Hu FB, Willett WC, Giovannucci E. Dietary patterns and risk of prostate cancer in U.S. men. Cancer Epidemiol Biomarkers Prev 2006;15:167–71. [PubMed: 16434606]

23. Chan JM, Stampfer MJ, Ma J, Rimm EB, Willett WC, Giovannucci EL. Supplemental vitamin E intake and prostate cancer risk in a large cohort of men in the United States. Cancer Epidemiol Biomarkers Prev 1999;8:893–9. [PubMed: 10548318]

24. Shipley, W.; Zeitman, AL. Salvage radiation after radical prostatectomy. In: Kantoff, PW.; Carroll, PR.; D'Amico, AV., editors. Prostate cancer principles and practice. Philadelphia, PA: Lippincott Williams & Wilkins; 2002.

25. Hu FB, Stampfer MJ, Rimm E, Ascherio A, Rosner BA, Spiegelman D, Willett WC. Dietary fat and coronary heart disease: a comparison of approaches for adjusting total energy intake and modeling repeated dietary measurements. Am J Epidemiol 1999;149:531–40. [PubMed: 10084242]

26. Leibowitz SB, Kantoff PW. Differentiating agents and the treatment of prostate cancer: vitamin D3 and peroxisome proliferator-activated receptor gamma ligands. Semin Oncol 2003;30:698–708. [PubMed: 14571417]

27. Krishnan AV, Peehl DM, Feldman D. The role of vitamin D in prostate cancer. Recent Results Cancer Res 2003;164:205–21. [PubMed: 12899524]

28. Chan JM, Stampfer MJ, Ma J, Gann P, Gaziano JM, Pollak M, Giovannucci E. Insulin-like growth factor-I (IGF-1) and IGF binding protein-3 as predictors of advanced-stage prostate cancer. J Natl Cancer Inst 2002;94:1099–109. [PubMed: 12122101]

29. Li L, Yu H, Schumacher F, Casey G, Witte JS. Relation of serum insulin-like growth factor-I (IGF-I) and IGF binding protein-3 to risk of prostate cancer (United States). Cancer Causes Control 2003;14:721–6. [PubMed: 14674736]

30. Cardillo MR, Monti S, Di Silverio F, Gentile V, Sciarra F, Toscano V. Insulin-like growth factor (IGF)-I, IGF-II and IGF type I receptor (IGFR-I) expression in prostatic cancer. Anticancer Res 2003;23:3825–35. [PubMed: 14666684]

31. Liao Y, Abel U, Grobholz R, Hermani A, Trojan L, Angel P, Mayer D. Up-regulation of insulin-like growth factor axis components in human primary prostate cancer correlates with tumor grade. Hum Pathol 2005;36:1186–96. [PubMed: 16260272]

32. Giovannucci E, Rimm EB, Liu Y, Leitzmann M, Wu K, Stampfer MJ, Willett WC. Body mass index and risk of prostate cancer in U.S. health professionals. J Natl Cancer Inst 2003;95:1240–4. [PubMed: 12928350]

33. MacInnis RJ, English DR. Body size and composition and prostate cancer risk: systematic review and meta-regression analysis. Cancer Causes Control 2006;17:989–1003. [PubMed: 16933050]

34. Engeland A, Tretli S, Bjorge T. Height, body mass index, and prostate cancer: a follow-up of 950000 Norwegian men. Br J Cancer 2003;89:1237–42. [PubMed: 14520453]

35. Andersson SO, Wolk A, Bergstrom R, Adami HO, Engholm G, Englund A, Nyren O. Body size and prostate cancer: a 20-year follow-up study among 135006 Swedish construction workers. J Natl Cancer Inst 1997;89:385–9. [PubMed: 9060961]

36. Rodriguez C, Patel AV, Calle EE, Jacobs EJ, Chao A, Thun MJ. Body mass index, height, and prostate cancer mortality in two large cohorts of adult men in the United States. Cancer Epidemiol Biomarkers Prev 2001;10:345–53. [PubMed: 11319175]

37. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 2003;348:1625–38. [PubMed: 12711737]

38. Rodriguez C, Freedland SJ, Deka A, Jacobs EJ, McCullough ML, Patel AV, Thun MJ, Calle EE. Body mass index, weight change, and risk of prostate cancer in the Cancer Prevention Study II Nutrition Cohort. Cancer Epidemiol Biomarkers Prev 2007;16:63–9. [PubMed: 17179486]

39. Albertsen PC, Hanley JA, Barrows GH, Penson DF, Kowalczyk PD, Sanders MM, Fine J. Prostate cancer and the Will Rogers phenomenon. J Natl Cancer Inst 2005;97:1248–53. [PubMed: 16145045]

40. Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG, Lieber MM, Cespedes RD, Atkins JN, Lippman SM, Carlin SM, Ryan A, et al. The influence of finasteride on the development of prostate cancer. N Engl J Med 2003;349:215–24. [PubMed: 12824459]

41. Giovannucci E, Liu Y, Stampfer MJ, Willett WC. A prospective study of calcium intake and incident and fatal prostate cancer. Cancer Epidemiol Biomarkers Prev 2006;15:203–10. [PubMed: 16492906]
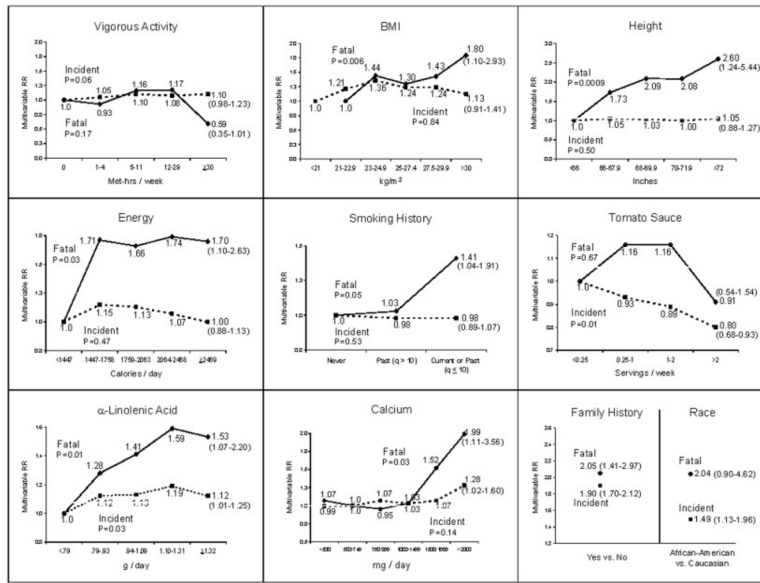
**Figure 1.**
Multivariable relative risks and 95% confidence intervals (in ()'s) for the high category versus low or reference category for selected variables from Cox models separately for total prostate cancer incidence and fatal prostate cancer in the Health Professionals Follow-Up Study. The models included age, time period, BMI at age 21 years, height, cigarette pack-years in the previous 10 years, vigorous physical activity level, family history of prostate cancer, history of diabetes mellitus, race, and intakes of total calories, processed meat, fish, α-linolenic acid, tomato sauce, and vitamin E supplements. For BMI, reference group is <23 kg/m$^2$ for fatal prostate cancer. P denotes *p*-value for trend.
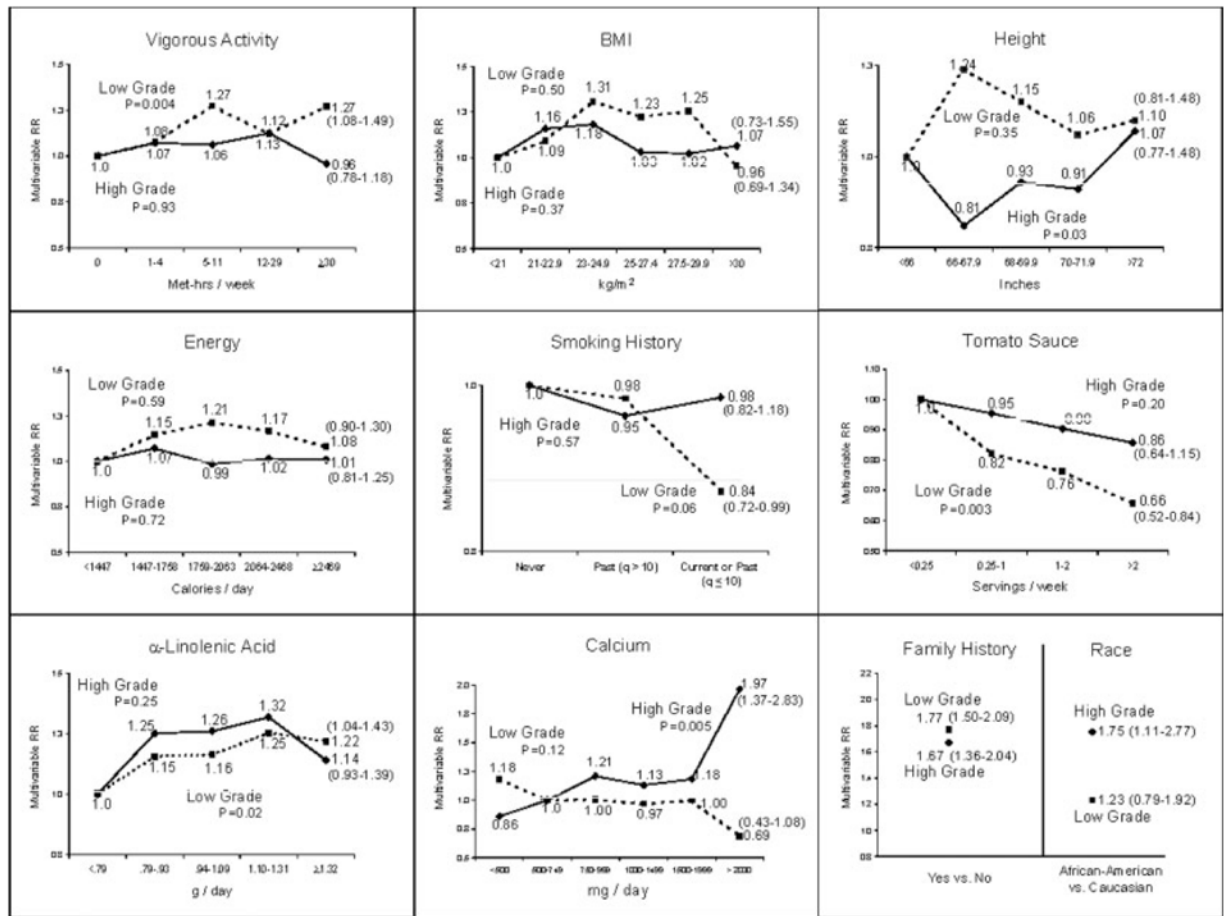
**Figure 2.**
Multivariable relative risk, and 95% confidence intervals (in ()'s) for the high category versus low or reference category for selected variables from Cox models separately for high-grade prostate cancer (Gleason ≥ 7) and low-grade prostate cancer (Gleason ≤ 6) in the Health Professionals Follow-Up Study. The models included age, time period, BMI at age 21 years, height, cigarette pack-years in the previous 10 years, vigorous physical activity level, family history of prostate cancer, history of diabetes mellitus, race, and intakes of total calories, processed meat, fish, α-linolenic acid, tomato sauce, and vitamin E supplements. P denotes *p*-value for trend.

**TABLE I**

MULTIVARIABLE RELATIVE RISK (RR) AND 95% CONFIDENCE INTERVAL (CI) FOR HIGH VERSUS LOW CATEGORY OF THE SPECIFIED VARIABLE BY STAGE OF PROSTATE CANCER

| Factor[1] | Organ confined | Minimally extraprostatic | Advanced |
|---|---|---|---|
| | T1 or T2 and N0M0 (*n* = 2161) | T3a and N0M0 (*n* = 345) | T3b or T4 or N1 or M1 (*n* = 523) |
| Cigarette smoking (prior 10 y) | 0.88 (0.77–1.01)[2] | 1.11 (0.82–1.52) | 1.14 (0.90–1.45) |
| Height | 1.00 (0.79–1.28) | 0.80 (0.46–1.40) | 2.12 (1.23–3.65)[3] |
| Vigorous physical activity | 1.19 (1.03–1.37)[3] | 1.25 (0.89–1.75) | 0.64 (0.44–0.94)[2] |
| Body mass index | 1.02 (0.76–1.36) | 0.91 (0.48–1.72) | 1.34 (0.79–2.26)[3] |
| Total energy intake | 1.04 (0.89–1.21) | 0.77 (0.52–1.14) | 1.68 (1.20–2.36)[2] |
| Calcium intake | 0.96 (0.68–1.34) | 1.98 (1.04–3.78) | 1.91 (1.20–3.03)[3] |
| α–linolenic acid | 1.12 (0.98–1.29)[2] | 1.18 (0.83–1.69) | 1.57 (1.19–2.07)[3] |
| Tomato sauce intake | 0.75 (0.61–0.92)[3] | 0.64 (0.37–1.10) | 0.66 (0.44–1.00) |
| Family history of CaP | 1.77 (1.54–2.04)[3] | 2.43 (1.78–3.32)[3] | 2.27 (1.73–2.97)[3] |
| African–American | 1.51 (1.06–2.15)[3] | 0.26 (0.04–1.87) | 1.75 (0.90–3.41) |

[1] Multivariable RR based on the high versus low category for each specified item (see Fig. 1 for variables in the model and cutpoints for the categories).

[2] *p* value $0.06 < p \leq 0.10$, for test for trend.

[3] *p* value $\leq 0.05$, for test for trend.

**TABLE II**

MULTIVARIABLE RELATIVE RISK (RR) AND 95% CONFIDENCE INTERVAL (CI) FOR HIGH VERSUS LOW
CATEGORY OF THE SPECIFIED VARIABLE IN THE PRE-PSA (1986–1/1992) AND PSA (2/1992–1/2002) ERAS

| Factor[2] | Total Prostate Cancer | | Advanced Prostate Cancer[1] | |
|---|---|---|---|---|
| | Pre-PSA Era (*n* = 894) | PSA Era (*n* = 2650) | Pre-PSA Era (*n* = 246) | PSA Era (*n* = 277) |
| Cigarette smoking (prior 10 y) | 1.07 (0.89–1.28) | 0.93 (0.82–1.06) | 1.17 (0.84–1.64) | 1.10 (0.77–1.55) |
| Height | 1.11 (0.78–1.57) | 1.03 (0.83–1.29) | 1.66 (0.81–3.41)[4] | 2.79 (1.20–6.50)[4] |
| Vigorous physical activity | 0.95 (0.74–1.22) | 1.15 (1.01–1.30)[4] | 0.57 (0.30–1.06) | 0.69 (0.43–1.11) |
| Body mass index | 0.94 (0.61–1.47) | 1.21 (0.94–1.56) | 1.21 (0.55–2.67)[4] | 1.43 (0.70–2.91) |
| Total energy intake | 1.14 (0.88–1.47) | 0.95 (0.82–1.09) | 1.90 (1.17–3.10)[4] | 1.50 (0.93–2.42) |
| Calcium intake | 1.77 (1.26–2.48)[3] | 1.01 (0.74–1.39) | 1.82 (0.97–3.45) | 2.08 (1.05–4.10) |
| α-linolenic acid | 1.12 (0.91–1.38) | 1.13 (0.99–1.28)[3] | 1.62 (1.09–2.42)[4] | 1.51 (1.03–2.23)[4] |
| Tomato sauce intake | 0.70 (0.52–0.93)[3] | 0.86 (0.70–1.04)[3] | 0.69 (0.40–1.20) | 0.63 (0.34–1.17) |
| Family history of CaP | 2.70 (2.22–3.27)[4] | 1.66 (1.45–1.89)[4] | 3.17 (2.22–4.54)[4] | 1.56 (1.02–2.37)[4] |
| African–American | 2.00 (1.25–3.22)[4] | 1.31 (0.93–1.85) | 2.49 (1.09–5.67)[4] | 1.13 (0.36–3.54) |

[1] Stringent definition of advanced prostate cancer (*e.g.*, T3b, T4 or N1 or M1).

[2] Multivariable RR based on the high versus low category for each specified item (see Fig. 1 for cutpoints for the categories).

[3] *p* value 0.06 < *p* ≤ 0.10, for test for trend.

[4] *p* value ≤ 0.05, for test for trend.

**TABLE III**

SUMMARY OF RESULTS FOR RISK FACTORS FOR PROSTATE CANCER ENDPOINTS[1] IN MULTIVARIABLE ANALYSIS IN THE HEALTH PROFESSIONALS FOLLOW-UP STUDY (1986–2002)[2]

| | Incident | Non-advanced | Low-grade | Fatal | Advanced | High-grade |
|---|---|---|---|---|---|---|
| Vigorous activity | | | | ↓ | | |
| Body mass index | ↓ | | | ← | (↓) | |
| Calorie intake | ← | ← | | ← | ← | |
| Height | ↑ | ↑ | ↑ | ↑ | ↑ | (↑) |
| Tobacco (last 10 yr) | | | | ← | | |
| Tomato sauce | ↓ | ↓ | (↓) | (↓) | | |
| α-linolenic acid | ← | (↑) | ↓ | (↑) | ← | |
| Calcium | (↑) | | | ← | (↑) | ← |
| African–American | ← | ← | | | | ← |
| Family history of CaP | ← | ← | ← | ← | ← | ← |

---

[1] Advanced cancers are those with extension into seminal vesicle, lymph nodes, distant metastatic, or which are fatal. High grade cancers are those of Gleason grade ≤7.

[2] ↑ denotes positive association; ↓ denotes inverse association. If no () is given, both the test for trend and the comparison for extreme categories are statistically significant. () denotes that only one of these is statistically significant.