Appendix 2: Details regarding statistical methods

Restricted cubic spline and hazard ratio curve

The restricted cubic spline function uses piecewise cubic polynomials that are connected across different intervals of a continuous variable. It can fit sharply curving shapes, with the additional advantage that only $k - 1$ parameters must be estimated (where k is the number of knots).[1] We chose 4 knots at quantiles 0.05, 0.35, 0.65 and 0.95, since it has been suggested that 4 knots offer an adequate fit of the model, representing a good compromise between flexibility and the loss of precision caused by overfitting a small sample.[1] Use of the restricted cubic spline in Cox models has been described in detail by Harrell.[1] Elsewhere, we have further described how to use the spline Cox model to derive plots of hazard ratios for examining full-range associations in a more understandable way.[2,3] Briefly, the hazard ratio between 2 points of variable X can be estimated by $\exp(Y_2 - Y_1)$, where Y_1 and Y_2 are the corresponding spline function values of the 2 points of variable X. If we select a proper point Y_1 as the referent, $\exp(Y_2 - Y_1)$ stands for the hazard ratio of point 2 versus point 1. In this study, the value of LDL cholesterol with the lowest cancer risk was used as the referent, and the hazard ratios of all other LDL cholesterol levels relative to the referent value were calculated and plotted against their respective LDL cholesterol levels.

Rationale for yes/no coding scheme

We previously developed a drug adjustment scheme for abnormal lipid outcomes: the adjusted time proportion of drug use during the follow-up period.[4] The drug adjustment scheme considered cumulative effects, total follow-up time and diminishing effects after termination of the drugs. In this analysis for the outcome of cancer, we further compared the fit of the model using yes/no coding for use of drugs from enrolment to the outcome of cancer, death or censoring, whichever came first, with the drug adjustment scheme. Because yes/no coding was associated with the best-fit model for the outcome of cancer among all possible combinations of yes/no coding, the unadjusted time proportions of drug use and the adjusted time proportion of drug use, the yes/no coding scheme was used for all models in the analysis.

References

1. Harrell F. *Regression modelling strategies with applications to linear models, logistic regression, and survival analysis*. New York: Spinger-Varlag; 2001.
2. So WY, Yang X, Ma RC, et al. Risk factors in V-shaped risk associations with all-cause mortality in type 2 diabetes — the Hong Kong Diabetes Registry. *Diabetes Metab Res Rev* 2008;24:238-46.
3. Yang X, Ma RC, So WY, et al. Impacts of chronic kidney disease and albuminuria on associations between coronary heart disease and its traditional risk factors in type 2 diabetic patients — the Hong Kong diabetes registry. *Cardiovasc Diabetol* 2007;6:37.
4. Yang X, So WY, Ma R, et al. Effects of albuminuria and renal dysfunction on development of dyslipidaemia in type 2 diabetes–the Hong Kong Diabetes Registry. *Nephrol Dial Transplant* 2008 Mar 27. [Epub ahead of print]. doi:10.1093/ndt/gfn149:gfn149.