Appendix 1 | Detailed description of methods

Design overview

The WHS was a randomized trial evaluating the effect of 100mg of aspirin on alternate days compared with placebo for primary prevention of CVD and cancer in 39,876 women of 45 years of age or older, without a history of cardiovascular disease or cancer. Detailed methods and outcomes have been described previously[1-4]. Written informed consent was obtained from all participants and the trial was approved by the Institutional Review Board of Brigham and Women's Hospital and was monitored by an external data and safety monitoring board. Endpoints were ascertained using yearly questionnaires and were confirmed using medical records. All relevant information was reviewed by an endpoints committee comprising physicians blinded to treatment allocation[1, 2]. After the end of randomized treatment on 31 March 2004, with an average 10 years of follow-up, participants were invited for further observational follow-up[4]. Of the survivors 33,682 (88.6%) women agreed to continue participation. During the posttrial follow-up, use of aspirin was allowed for women from both study arms. The posttrial use of aspirin for at least three days per month was higher in the randomized aspirin group (46%) compared to the placebo group (43%). Women who used nonstudy aspirin during the posttrial follow-up used aspirin for a median of three years (IQR: 2-5 years)[4]. Information on outcomes was collected and confirmed in a similar manner as during the trial period. End point review is complete for 95% of reported cancer cases, 95% of myocardial infarctions, and 94% of strokes. The confirmation rate among participants with records is 82% for cancer, 61% for myocardial infarction, and 68% for stroke. For the present study, only events confirmed by medical records and deaths with confirmed cause were used. Reports of gastrointestinal bleeding were collected intermittently during posttrial follow-up and were not confirmed[4]. The present analyses include end points accrued and confirmed through 14 March 2012, using data of participants who provided an adequate baseline plasma sample (n=27,939).

Model development

Data of women who provided a baseline plasma sample (n=27,939) were used for model development. For the 10-year predictions, endpoints that occurred during the trial period were used. In order to capture any delayed effects of aspirin on cancer risk[4, 5], the cancer outcomes were also modeled using cases ascertained during the entire follow-up, for prediction of 15-year treatment effect. Since the effects of aspirin on CVD and bleeding seem to be more immediate[4, 6] and the randomized aspirin intervention stopped after 31 March 2004, modeling these outcomes using posttrial data would likely lead to underestimation of the treatment effect. Hence, 15-year predictions for CVD and bleeding were obtained by extrapolating the 10-year risk estimates under the assumption of exponential risk over time, to mimic the effects of taking aspirin for a duration of 15-years. As the CVD endpoint included all strokes, hemorrhagic strokes were not evaluated separately.

To minimize over-fitting, predictors for each outcome were selected based on existing risk scores and/or literature[7-11]. Only predictors that were deemed to be easily available in clinical practice were selected. As a result, the following predictors, besides aspirin treatment, were used for major cardiovascular events (CVD): age, current smoking, body mass index (BMI), systolic blood pressure (SBP), use of blood pressure lowering medication, total cholesterol, high density lipoprotein cholesterol (HDLc), high sensitivity C-reactive protein (hs-CRP), family history of premature coronary heart disease (CHD) and hemoglobin A1c (HbA1c) if diabetic; for colorectal cancer: age, ever smoking, BMI, height, diabetes mellitus, alcohol use (no. of drinks per day), menopausal status, hormone replacement therapy use, family history of colorectal cancer; for non-colorectal cancer: age, ever smoking, BMI, height, diabetes mellitus, alcohol use, menopausal status, hormone replacement therapy use, family history of breast, colorectal, or ovarian cancer; for major bleeding events: age, current smoking, BMI, alcohol use, diabetes mellitus, history of dyspepsia.

The relative treatment effect of aspirin was assumed constant in the main analysis. Findings of effect modification by any risk factors are inconsistent in previous studies[1, 2, 12-14], although significant effect modification was found by age and smoking for CVD in the WHS[2]. To evaluate these potential relative subgroup effects, sensitivity analyses were performed in which treatment interactions with age, smoking status and BMI were considered. These interactions terms were chosen based on previous findings of interaction[1, 2, 15] and/or strong pathophysiological evidence[16, 17]. To avoid including non-relevant treatment interactions, estimation of model coefficients with implicit variable selection was done using component-wise likelihood-based boosting[18]. Aspirin use was included as an mandatory (unpenalized) covariable, whereas the other candidate predictors and treatment interactions were subjected to penalization in penalized partial likelihood estimation. The optimal number of boosting steps was determined by 10-fold cross-validation[19].

Similar to previous analysis of the WHS[1, 4], no effect of aspirin on non-colorectal cancer was observed in the present competing risks analysis (HR 1.02, 95%CI 0.95-1.09). Since the incidence of non-colorectal cancer is high compared to the other competing outcomes, even a small non-significant coefficient could potentially have considerable effects on the overall treatment effect predictions. To evaluate these effects

and to test the robustness of the results, sensitivity analysis were performed in which the treatment effect of aspirin on non-colorectal cancer was assumed null. Accordingly, the competing risks endpoint was adjusted in these analyses.

One or more covariable data were missing in 865 (3.1%) participants and these were singly imputed using bootstrapping and predictive mean matching (aregImpute-algorithm in R, Hmisc-package)[20]: family history of premature CHD (n=464), SBP (n=292), HbA1c (n=140), hormone replacement therapy use (n=55), menopausal status (n=51), smoking status (n=36), BMI (n=23), blood pressure lowering medication use (n=18), diabetes mellitus (n=15), total cholesterol (n=1), HDLc (n=1), alcohol use (n=6), family history of cancer (n=865) and height (n=18). To limit the effect of outliers, continuous predictors were truncated at the 1st and 99th percentile. Continuous predictors that were not linearly associated to the outcome were transformed to optimize model fit[21]. Accordingly, HDLc, total cholesterol, systolic blood pressure and hsCRP were log-transformed.

Model validation

An estimate of the optimism in the calibration slope was obtained for all models by repeating the complete modeling process in 500 bootstrap samples. The optimism was 0.9% for the CVD model, 9.7% for the 10-year colorectal cancer model, 7.7% for the 15-year colorectal cancer model, 4.1% for the 10-year non-colorectal cancer model, 3.2% for the 15-year non-colorectal cancer model and 4.9% for the bleeding model. Subsequently, the obtained uniform shrinkage factors were applied to the models to adjust for overfitting[21].

The proportional subdistribution hazards assumptions were assessed graphically by plotting the scaled Schoenfeld residuals against failure time and formally by a Wald test of the interaction term of a specific covariable with the logarithm of time. Some non-proportionality was observed for age and family history of cancer in the 15-year model for non-colorectal cancer (p-values: <0.001 and 0.039, respectively). In addition, the proportionality assumption appeared to be violated for history of dyspepsia in the gastro-intestinal bleeding model (p-value: 0.044). Hence, the reported coefficients for these predictors should be interpreted as the weighted average effect over follow-up[22].

Discriminatory ability of each model was evaluated using an inverse probability of censoring weighted estimate of the c-index that is adapted for competing risks[23]. C-indices were truncated at 10 or 15-year and corrected for optimism by repeating the complete modeling process in 500 bootstrap samples. Calibration was assessed graphically using calibration plots.

Net benefit assessment

To evaluate the clinical value of prediction-based treatment with aspirin in a primary prevention setting, a decision analytic approach as proposed by Vickers et al. [24] was used. This method focuses on the effects of (changes in) treatment decisions that result from a treatment strategy (e.g. prediction-based treatment) and is based on calculation of 'net benefit'. Net benefit is defined as the treatment benefit (reduction in event rate) minus the treatment harm (adverse effects, costs, etc.), where the relative weighting of treatment harm is given by a treatment threshold (*i.e.* ARR at which one would opt for treatment). This treatment threshold is the reciprocal of the maximum acceptable number-needed-to-treat (NNT) to prevent one event or 'number-willing-to-treat' (NWT)[7, 25]. Consequently, the net benefit of a certain treatment strategy is calculated as the observed decrease in event rate minus the treatment rate multiplied by the treatment threshold. Using the aggregated ARRs of all outcomes for each individual, the clinical value of the combination of the benefit and harm models can be assessed.Net benefit was calculated for the following treatment strategies: (I) treat no one (reference, *i.e.* net benefit equals zero), (II) treat everyone, (III) treat according to guidelines[26] , *i.e.* women ≥65 years and (IV) prediction-based treatment. Since major gastro-intestinal bleeding is already incorporated in the total ARR, the treatment threshold for aspirin is mainly determined by less serious complications, inconvenience of taking pills and costs. As the appropriate treatment threshold (or NWT) is subjective and can vary among different patients and clinicians, the net benefit was calculated for threshold values ranging from 0 to 5% (10-/15-year NWT between infinite and 20). Net benefit for the different treatment strategies was also calculated applying a weight of 0.5, 0.25 and 0.1 for gastro-intestinal bleeding. The net benefit results were presented graphically as decision curves after local polynomial regression fitting.

All analyses were performed in R, version 3.0.2 (R Core Team, Vienna, Austria; packages: 'Hmisc', 'pec', 'riskRegression').

References

1. Cook NR, Lee I-m, Gaziano JM, *et al.* Low-Dose Aspirin in the primary prevention of cancer. JAMA 2005;294:47-55.
2. Ridker PM, Cook NR, Lee IM, *et al.* A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. The New England Journal of Medicine 2005;352(13):1293-304.
3. Buring JE, Hennekens CH. The Women's Health Study: Summary of the study design. Journal of Myocardial Ischemia 1992;4:27-9.
4. Cook NR, Lee IM, Zhang SM, *et al.* Alternate-Day, Low-Dose Aspirin and Cancer Risk: Long-Term Observational Follow-up of a Randomized Trial. Ann Intern Med 2013;159(2):77-85.
5. Thun MJ, Jacobs EJ, Patrono C. The role of aspirin in cancer prevention. Nat Rev Clin Oncol 2012;9(5):259-67.
6. Patrono C, Coller B, Dalen JE, *et al.* Platelet-active drugs : the relationships among dose, effectiveness, and side effects. Chest 2001;119(1 Suppl):39S-63S.
7. Dorresteijn JAN, Visseren FLJ, Ridker PM, *et al.* Aspirin for primary prevention of vascular events in women: individualized prediction of treatment effects. European Heart Journal 2011;32:2962-9.
8. Freedman AN, Slattery ML, Ballard-Barbash R, *et al.* Colorectal cancer risk prediction tool for white men and women without known susceptibility. J Clin Oncol 2009;27(5):686-93.
9. Ridker PM, Buring JE, Rifai N, *et al.* Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA 2007;297(6):611-9.
10. Wei EK, Colditz GA, Giovannucci EL, *et al.* Cumulative risk of colon cancer up to age 70 years by risk factor status using data from the Nurses' Health Study. Am J Epidemiol 2009;170(7):863-72.
11. de Groot NL, Hagenaars MP, Smeets HM, *et al.* Primary non-variceal upper gastrointestinal bleeding in NSAID and low-dose aspirin users: development and validation of risk scores for either medication in two large Dutch cohorts. J Gastroenterol 2013.
12. Berger JS, Lala A, Krantz MJ, *et al.* Aspirin for the prevention of cardiovascular events in patients without clinical cardiovascular disease: a meta-analysis of randomized trials. Am Heart J 2011;162(1):115-24 e2.
13. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med 2009;150(11):795-802.
14. Rothwell PM, Price JF, Fowkes FGR, *et al.* Short-term effects of daily aspirin on cancer incidence, mortality, and non-vascular death: analysis of the time course of risks and benefits in 51 randomised controlled trials. The Lancet 2012;379:1602-1612.
15. Garcia Rodriguez LA, Hernandez-Diaz S, de Abajo FJ. Association between aspirin and upper gastrointestinal complications: systematic review of epidemiologic studies. Br J Clin Pharmacol 2001;52(5):563-71.
16. Patrono C, Garcia Rodriguez LA, Landolfi R, *et al.* Low-dose aspirin for the prevention of atherothrombosis. N Engl J Med 2005;353(22):2373-83.
17. Rocca B, Dragani A, Pagliaccia F. Identifying determinants of variability to tailor aspirin therapy. Expert Rev Cardiovasc Ther 2013;11(3):365-79.
18. Binder H, Allignol A, Schumacher M, *et al.* Boosting for high-dimensional time-to-event data with competing risks. Bioinformatics 2009;25(7):890-6.
19. Verweij PJ, Van Houwelingen HC. Cross-validation in survival analysis. Stat Med 1993;12(24):2305-14.
20. Donders AR, van der Heijden GJ, Stijnen T, *et al.* Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006;59(10):1087-91.
21. Steyerberg EW. Clincial prediction models: a practical approach to development, validation, and updating. New York, USA: Springer. 2009.
22. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol 2009;170(2):244-56.
23. Wolbers M, Blanche P, Koller MT, *et al.* Concordance for prognostic models with competing risks. Biostatistics 2014.
24. Vickers AJ, Kattan MW, Daniel S. Method for evaluating prediction models that apply the results of randomized trials to individual patients. Trials 2007;8:14.
25. Dorresteijn JAN, Visseren FLJ, Ridker PM, *et al.* Estimating treatment effects for individual patients based on the results of randomised clinical trials. British Medical Journal 2011;343:d5888-d5888.
26. Mosca L, Benjamin EJ, Berra K, *et al.* Effectiveness-based guidelines for the prevention of cardiovascular disease in women--2011 update: a guideline from the american heart association. Circulation 2011;123(11):1243-62.