Individual and Population Benefits of Daily Aspirin Therapy
A Proposal for Personalizing National Guidelines

Jeremy B. Sussman, MD, MS; Sandeep Vijan, MD, MS; HwaJung Choi, PhD; Rodney A. Hayward, MD

Background—Clinical practice guidelines that help clinicians and patients to understand the magnitude of expected individual risks and benefits would help with patient-centered decision-making and prioritization of care. We assessed the net benefit from taking daily aspirin to estimate the individual and public health implications of a more individualized decision-making approach.

Methods and Results—We used data from the National Health and Nutrition Examination Survey representing all US persons aged 30 to 85 years with no history of myocardial infarction and applied a Markov model based on randomized evidence and published literature to estimate lifetime effects of aspirin treatment in quality-adjusted life years (QALYs). We found that treatment benefit varies greatly by an individual’s cardiovascular disease (CVD) risk. Almost all adults have fewer major clinical events on aspirin, but for most, events prevented would be so rare that even a very small distaste for aspirin use would make treatment inappropriate. With minimal dislike of aspirin use (disutility, 0.005 QALY per year), only those with a 10-year cardiac event risk >6.1% would have a net benefit. A disutility of 0.01 QALY moves this benefit cut point to 10.6%. Multiple factors altered the absolute benefit of aspirin, but the strong relationship between CVD risk and magnitude of benefit was robust.

Conclusions—The benefits of aspirin therapy depend substantially on an individual’s risk of CVD and adverse treatment effects. Understanding who benefits from aspirin use and how much can help clinicians and patients to develop a more patient-centered approach to preventive therapy. (Circ Cardiovasc Qual Outcomes. 2011;4:00-00.)

Key Words: aspirin • prevention • risk factors • decision making • primary prevention • chronic ischemic heart disease • cardiovascular pharmacology

 Critics have faulted guidelines and related quality and performance measures for defining treatments using broad categories of patients (eg, aged >65 years or having diabetes) rather than of individual patient risks and benefits. Such broad, sweeping recommendations are further criticized for being inconsistent with the ethic of personalized medicine and patient-centered care. It is clear that even for worthwhile therapies, there can be dramatic variation in how much individual patients benefit (or are harmed) from a therapy. Research and guidelines that recognize these limitations could help clinicians and health systems to prioritize care and personalize their recommendations to their patients’ individual preferences and circumstances. However, means of quantifying individual risks and benefits often are not easily available. One important example of this difficulty is the confusion surrounding when aspirin should be used for the primary prevention of cardiovascular disease (CVD).

Aspirin for primary CVD prevention is one of the best-studied therapies in medicine, with 9 large, double-blind, randomized controlled trials (RCTs) completed; more under way; and several published meta-analyses. Despite this wealth of evidence, there is still a lack of clarity about who should receive therapy and limited understanding of the population implications of treatment cut points. Although some of this confusion is due to disagreement over treatment effectiveness in specific clinical subgroups, some of it is attributable to guidelines seeking a single cut point at which to recommend treatment rather than focusing on clinical trade-offs and the range of absolute benefit for the individual patient.

In this article, we describe a novel approach to personalizing treatment guidelines, using aspirin therapy as an example. We developed a Markov model that estimates the CVD risk reduction conferred by aspirin therapy in primary prevention, allowing individual estimates of the strength of expected net benefit or harm to patients. We also use these estimates to examine the benefit of aspirin therapy at different levels of CVD risk and the population-level (public health) benefit achieved by using different treatment cut points in targeting individuals for aspirin therapy.

Received September 16, 2010; accepted February 8, 2011.
From the Department of Veterans Affairs, VA Health Service Research and Development Center of Excellence, VA Ann Arbor Healthcare System (J.B.S., S.V., R.A.H.), and Department of Internal Medicine and The Robert Wood Johnson Clinical Scholars Program, University of Michigan (J.B.S., S.V., H.C., R.A.H.), Ann Arbor, MI.

The online-only Data Supplement is available at http://circoutcomes.ahajournals.org/cgi/content/full/CIRCOUTCOMES.110.959239/DC1.
Correspondence to Jeremy B. Sussman, MD, Robert Wood Johnson Clinical Scholars Program, University of Michigan Medical School, 6312 Medical Science Building I, Ann Arbor, MI 48109. E-mail jeremyssu@med.umich.edu
© 2011 American Heart Association, Inc.

Circ Cardiovasc Qual Outcomes is available at http://circoutcomes.ahajournals.org

DOI: 10.1161/CIRCOUTCOMES.110.959239
WHAT IS KNOWN

- Aspirin use reduces first heart attack and ischemic stroke by about 18% and 14%, respectively, but it also increases rates of intracerebral hemorrhage and gastrointestinal bleeding.

WHAT THE STUDY ADDS

- Only ≈28% of American adults are in the high cardiovascular risk group for whom aspirin for primary prevention is likely to be highly beneficial.
- For most adult Americans, the potential benefit of aspirin is small, and treatment decisions should depend almost completely on personal preferences.
- However, most official guidelines do not identify these individuals as candidates for aspirin therapy.
- Few adults would increase their net risk substantially by using aspirin.

Methods

Overview

To assess the individual and population benefits of aspirin use, we used evidence from population-weighted cohorts, RCTs, and prospective studies to develop a Markov simulation model of aspirin-related outcomes. We used validated prediction indices to estimate the risk of each of 4 primary outcomes in the absence of aspirin therapy: myocardial infarction (MI), ischemic stroke (ISC), gastrointestinal bleeding (GIB), and intracerebral hemorrhage (ICH). A meta-analysis of clinical trials supplied estimates of the relative effects of aspirin on these outcomes. Using evidence from this meta-analysis, we assumed that the relative risk reduction of treatment (RRRRx) from aspirin is constant across all risk factors, which allowed us to examine the effect of heterogeneity of individual risk of having CVD events on absolute risk reduction. Finally, the estimated event rates, with and without aspirin, were used to estimate quality-adjusted life years (QALYs) gained and lost (Table 1 and Figure 1).

Development of the Population

To assess the population effect of primary prevention of CVD with aspirin, we developed a nationally representative sample of the target population. We used data from all study subjects aged 35 to 85 years with no history of CVD and with fasting laboratory values from 2001 to 2006 of the National Health and Nutrition Examination Survey (NHANES). NHANES is a nationally representative, multistage probability sample survey of the noninstitutionalized population of the United States.

To simulate a 0.1% sample of the eligible population, we first expanded the data by applying sample weight to 417,138 participants (a 0.3% sample) and then performed 2 simulation steps. First, we conducted a first-order Monte Carlo simulation by obtaining predicted values from multivariate regressions and adding residuals randomly drawn from a normal distribution. Of this sampling pool, we randomly sampled 176,000 simulated participants (0.1% sample of the eligible population) as our primary population. The purpose of this step was to establish a sample population that was based on observed NHANES data but more robust. We could have used sample weights, but that would have risked having certain subgroups overly influenced by small groups of NHANES participants. The size of the sample was chosen on the basis of estimates of the sample size needed for stable output and was later verified when repeated samples showed that our results were highly stable. To account for the 4% of the population that had missing values in systolic and diastolic blood pressure, we imputed using switching regression an iterative multivariable regression technique.

Individual Risk Assessment

To assess individual risk of events, risk scores were used or developed for estimating fatal and nonfatal events of 4 major

Table 1. Model Attributes and Assumptions*

<table>
<thead>
<tr>
<th></th>
<th>MI</th>
<th>ISC</th>
<th>ICH</th>
<th>GIB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event rate</td>
<td>FHS13†</td>
<td>FHS13†</td>
<td>From Sturgeon et al14‡</td>
<td>From Saini et al15‡</td>
</tr>
<tr>
<td>Relative risk of event with aspirin</td>
<td>0.828</td>
<td>0.868</td>
<td>1.328</td>
<td>1.548</td>
</tr>
<tr>
<td>Event mortality rates</td>
<td>Derived from NCVS16,17</td>
<td>Derived from NCVS16,17</td>
<td>0.518</td>
<td>Age-based19</td>
</tr>
<tr>
<td>SMR mortality after year 1§</td>
<td>2.018</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QALY loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year of event</td>
<td>0.8823,23</td>
<td>0.6723–25</td>
<td>0.5823–25</td>
<td>0.9424,25</td>
</tr>
<tr>
<td>Per year, afterwards</td>
<td>0.9023</td>
<td>0.9023</td>
<td>0.9023–25</td>
<td>0.9924,25</td>
</tr>
</tbody>
</table>

Details are provided in online-only Data Supplement Appendix 1. FHS indicates Framingham Heart Study score; GIB, gastrointestinal bleeding; ISC, intracerebral hemorrhage; ICH, ischemic stroke; MI, myocardial infarction; NCVS, National Center for Vital Statistics; QALY, quality-adjusted life years; SMR, standardized mortality ratio.

*The analyses use a 3% discount rate and varied the disutility of taking aspirin from 0 to 0.005 QALY per year.
†Different scores are used for cardiac events and strokes.
§Both ICH and GIB are multivariate risk scores that are derived from the cited works.
§The SMR is the relative effect of an event on the patient’s age-, sex-, and race-standardized life expectancy.
aspirin-related outcomes and competing risks (MI, ISC, GIB, and ICH) and dying of another cause or remaining healthy. We estimated overall fatal plus nonfatal MI and ISC event rates with Framingham Risk Scores. Measurement of left ventricular hypertrophy by ECG, an insensitive test that is included in some risk prediction models but is no longer routine in health maintenance examinations, was not used in our calculations. For GIB, we used the model developed by Siani et al and based on Rockall et al, which used age, steroid use, and concomitant anticoagulant or antiplatelet use as risk factors. To develop a risk score for stroke subtypes, we converted an existing multivariable risk model for ICH into a predictive model. ISC rates were Framingham Stroke Scale rates minus ICH rates. The risk factors used in the ICH and GIB models were consistent with those reported in other research. Competing risk (mortality due to anything other than MI, ISC, ICH, and GIB) was modeled from Centers for Disease Control and Prevention Life Expectancy calculators.

To assess fatal versus nonfatal events, we applied the event risk estimators to the entire NHANES population and calibrated event and fatality rates to sex, race, and age-specific National Center for Health Statistics Cause of Death Statistics and other published literature to ensure reliability and population-level accuracy. Calibration and discrimination of these models were assessed in multiple ways, including comparing the calculated fatal event rates with overall Cause of Death Statistics and comparing the rates of each event on the basis of other published literature (online-only Data Supplement Appendix 1). These techniques are important for population-level models to ensure accurate overall event rates.

### Assessment of the Benefits and Risks of Aspirin

All risk reductions used were obtained from a major meta-analysis of RCTs. This meta-analysis found that the RRR of aspirin is consistent across a range of cardiac risks. We used the RRR of aspirin in the meta-analysis for each of the 4 outcomes assessed in the present study (MI, ISC, ICH, and GIB). We closely examined the literature for possible treatment interactions, that is, situations where the RRR of aspirin could be different for some patients than for others. The only interaction regularly addressed in the literature is the possibility that aspirin functions differently in preventing stroke among women and MI among men. We address this issue in detail in online-only Data Supplement Appendix 1.

### The Markov Model and Assessing QALY Gains and Losses

We separately assessed QALYs lost for fatal and nonfatal events. Nonfatal MI, ISC, ICH, and GIB could affect a patient’s future QALYs in 3 ways: a reduction in the patient’s utility for the year following the event, a reduced utility for all subsequent years, and a reduction in life expectancy due to a higher mortality rate in the years after an event occurs (eg, those who experienced an MI have a higher mortality rate, other factors being equal, than those who had not had an MI). All components and disutilities came from the published literature (see online-only Data Supplement Appendix 1 for details).

The relative effects of aspirin on the primary outcomes were obtained from a pooled analysis of 6 RCTs (Table 1). This meta-analysis used patient-level data from the original RCTs and is, to our knowledge, the most reliable, robust information available at this time.

We next used these results in a Markov state transition model with 1-year cycles (Figure 1). Every simulated patient was assessed with and without aspirin use and in multiple sensitivity analyses. In addition to disutility associated with major bleeding complications from aspirin, we examined the impact of assuming a very small treatment disutility (varying it from 0.001 to 0.010 QALY per year) to account for the minor negative effects of aspirin use, such as stomach upset, costs, and the bother of taking a medication. These values are much smaller than the range in utilities seen in surveys of patient opinion, and even the highest value is roughly one third as large as the estimated disutility of having a single dry eye. We used a 3% discount rate for life years. Although men have higher event rates than women for major CVD events, following the findings of the individual patient meta-analysis, we used the same estimates of RRR for aspirin in women and men, unlike other recent meta-analyses and decision models. We describe our reasons for this decision and the implications in detail in online-only Data Supplement Appendix 2.

#### Analyses

Using nationally representative demographic information and the risk tools described previously, we calculated the lifetime effects of 5 years of aspirin treatment on CVD, bleeding events, and QALYs for the entire population in the present study (MI, ISC, ICH, and GIB). We closely examined the literature for possible treatment interactions, that is, situations where the RRR of aspirin could be different for some patients than for others. The only interaction regularly addressed in the literature is the possibility that aspirin functions differently in preventing stroke among women and MI among men. We address this issue in detail in online-only Data Supplement Appendix 1.

### Study Population

Our nationally representative simulated population closely matches previous reports of the US population aged 35 to 85 years with no history of CVD (Table 2). In a baseline simulation in which the entire population is untreated, we estimate that MI would account for 46% of all the events measured (670,000 first MIs nationwide in our age range per year) with >25 million QALYs lost due to first MI per 5 years of use, which represents 56% of all QALYs lost from the 4 major events studied. Our event rates are well calibrated with national data (see online-only Data Supplement Appendix 1 for details). GIB would account for 17% of the events.
Almost one half of the eligible population is in the lower 10% of risk (Framingham Heart Study risk score; QALY, quality-adjusted life year). If the risk is placed at 0.005 QALY per year, every individual percentile of risk would benefit from use. If the risk is placed at 0.005, then people with a 10-year estimated cardiac risk >6.1% would benefit. Almost one half of the eligible population is in this group. If the risk is 0.01 QALY per year, then benefit is at risk >10.6%. For everyone, those with a greater chance of an event have a greater chance of benefiting from aspirin. FHS indicates Framingham Heart Study risk score; QALY, quality-adjusted life year.

but <3% of the QALYs lost, which is unsurprising given its lesser clinical severity.

Overall Benefit of Aspirin
As illustrated in Figure 2, the absolute benefit of aspirin therapy increases continuously and substantially as a person’s predicted event rate rises. However, so few individuals will benefit from aspirin that even a small harm from or dislike of aspirin use (i.e., “treatment disutility,” an individual patient’s level of dislike for the medication) can have a large effect on which patients receive net benefit. If a patient truly has almost no aversion to taking a daily aspirin (disutility =0.001 QALY per year), then for virtually all Americans, aspirin use could be considered appropriate. However, even a small increase in distaste for aspirin (disutility=0.005 QALY per year) dramatically changes the number of patients who benefit so that only those with a 10-year cardiac event risk (Framingham score) >6.1% would have a net benefit, which is just more than half of the eligible US population. If the disutility is increased to 0.010 QALY per year, the cut point for net benefit moves to a cardiac risk of >10.6%, or 28% of the eligible population. This number is near the 10% cut point of the current American Heart Association and European Society of Cardiology guideline recommendations.9,41

Population-Level Benefit of Aspirin
Most practice guidelines impose a single treatment threshold along this continuous risk-benefit trade-off, above which treatment would be recommended. We found that successful implementation of the American Heart Association aspirin recommendations (treat if 10-year coronary artery disease risk is >10%, ie, our substantial benefit group) for a 5-year period would result in treating ~54 million Americans with a daily aspirin and result in a nationwide lifetime net gain of 670 000 QALYs per year. Treating the substantial benefit group of Framingham score >10.6% (those who would benefit even with a disutility of 0.01 QALY) would treat 50 million people for a 650 000-QALY benefit. Treating an additional 49 million people in the small benefit group would increase this benefit only slightly, resulting in a 35% smaller benefit per person of treating those at higher risk (online-only Data Supplement Table 4).

Because far more men than women have a Framingham risk score >10%, following this guideline treats many more men (39 million) than women (14 million), resulting in 500 000 QALYs saved versus 170 000 QALYs. The benefit per person treated, however, has a nearly identical number needed to treat of 20.4 years treatment per QALY saved for men and 20.9 years for women.

Individual-Level Benefit of Aspirin
Individual benefits from aspirin therapy varied substantially depending on baseline risk. For individuals with a 10-year CVD risk >10.6%, a net gain in 1 QALY required treatment of 20 people for 5 years, whereas for individuals with a 10-year risk between 6.1% and 10.6%, 43 people would need to be treated to achieve this same net QALY gain (online-only Data Supplement Table 6). The people who are most likely to benefit from aspirin, those with a 10-year event risk >20% (average number needed to treat, 13), comprise only about 9% of the US population aged 35 to 85 years and usually have multiple risk factors.

The vast bulk of the benefit of aspirin is attributable to a reduction in MI (online-only Data Supplement Table 6). A total of 150 individuals with a 10-year risk >10.6% would need 5 years of therapy to prevent an MI as opposed to 420 to prevent an ISC. In those with a 10-year risk >10.6%, aspirin would cause 1 ICH per 2200 people treated and 1 GIB per 280 treated (online-only Data Supplement Table 6). In online-only Data Supplement Appendix 3, we describe a series of representative patients and how this information would aid their treatment decisions.

Graded Recommendations
This evaluation, which recognizes treatment not by population-averaged effectiveness or cost-effectiveness but by quantified individual effectiveness, could potentially allow clearer shared decision-making. As seen in Figure 2, a clinical practice guideline based on these analyses need not involve presenting patients with excessive analytic detail but could provide clinicians and patients with qualitative assessments of the overall importance of treatments and how sensitive the recommendation should be to an individual’s treatment disutilities. Listing the results as “minimal benefit/do not recommend,” “small benefit/patient preference,” and “larger benefit/recommend” could emphasize the relationship between benefit and importance of recommendation.

Testing the Model With RCTs
Although some advocates would suggest that a treatment trial is necessary to validate our results, our estimates suggest that an RCT to verify the benefit of aspirin in the minimal and small net benefit range would be extremely time consuming.
and expensive. The largest existing trial of aspirin for primary prevention, the Women’s Health Study,42 has just over 400,000 years of participation. A new RCT to assess the benefit of aspirin for people with 10-year risk between 6.1% and 10.6% would require ~400,000 person-years (online-only Data Supplement Appendix 1). Testing whether aspirin is beneficial in the lowest benefit range of the population (ie, those with a 10-year risk <5.5%) would require >1 million patient-years of therapy.

**Sensitivity Analyses**

The strong relationship between underlying CVD risk and increasing magnitude of the benefits of aspirin was unaltered across a broad range of assumptions in sensitivity analyses (Table 3). Multiple factors, however, did alter the absolute benefit of aspirin. This absolute benefit was influenced most strongly by varying the RRR due to aspirin and overall event rates for each condition. Even in these instances, the main impact was to move those near the top or bottom range of 1 category into the adjacent category (ie, some individuals initially categorized as having a benefit at a disutility of 0.005 QALY only benefited if the disutility was decreased to 0.001 QALY). Although these analyses further demonstrate the large number of patients for whom the overall value of aspirin use is debatable, they also demonstrate that the decisions made by these individuals will not significantly affect the overall population-level benefit.

**Discussion**

Although most aspirin guidelines propose a single threshold for treatment decisions, the present study demonstrates how the clinical importance of aspirin varies tremendously and continuously as a function of a patient’s coronary artery disease risk and how individual-level benefit estimates can be produced to aid personalized patient decision-making. Our results further demonstrate that for most adult Americans, the potential benefit of aspirin is small and the decision to take it should depend almost completely on personal preference, yet most individuals are not currently identified by official guidelines as candidates for daily aspirin use.

The relationship among the overall cardiac risk, individual treatment disutility, and the risk-benefit profile of aspirin has not been fully assessed previously. Across our population, there was almost no net harm from aspirin, even in low-risk patients, as the cardiovascular benefits outweigh bleeding risks for nearly all patient groups. For the remainder of the population, the appropriateness of treatment depends on the individual’s preferences and utilities for treatments and specific health outcomes. Guidelines, pay-for-performance models, and clinical care could adopt a more-efficient and patient-centered approach by (1) targeting people who are likely to have a substantial benefit, (2) providing a more-tempered recommendation for treatment of those with small expected benefit, (3) neither encouraging nor discouraging therapy in those with uncertain benefit, and (4) recommending against treatment in those in whom harm likely outweighs benefit (online-only Data Supplement Table 7).

The present study results build on previous work in this area38,43 in several ways. Most importantly, our study examines the relationship between clinical decisions for individual patients and their public health benefit. We examined how individual decisions scale up to population-level effects and found that a minority of people in the United States fit within the typically described subgroups of treat or do not treat. Instead, we found that the majority of the population receives a small benefit from aspirin but one that might warrant treatment in those with certain preference profiles. This finding that the majority of the population benefits is nested within a relatively small proportion of individuals is a common and often-neglected phenomenon44 and merits more careful attention not only in clinical guidelines,2 but also in the analysis and reporting of clinical trials.45

Although previous decision analyses have recognized the value of risk in assessing the benefits of aspirin use,38,43 none have examined the implications of using clinically relevant measures of risk as a way to guide clinical therapy. Similarly,
although some other studies apply a disutility for aspirin use to examine its impact on population-averaged effects, none have assessed the importance that individual variation (patient heterogeneity as opposed to model parameter uncertainty) in disutility would have on effective individual patient decision-making. Furthermore, previous probability models have assessed risk and benefit using risk scores for MI and ISC but, to our knowledge, not for ICH and GIB. We believe that together, this has made our analyses substantially more relevant and useful for clinical decision-making while retaining a utility for policy-level instruction.

Although there are clinical practice guidelines that follow our 3-group model for guidelines (ie, recommend against treatment in low-risk patients, recommend for treatment in high-risk patients, and recommend shared decision-making in patients at intermediate risk), our model improves on these by quantifying the benefit at these risk levels while being based on widely available clinical risk tools. Basing care around the risk-benefit of treatment, as we have done, may have the weaknesses of proving more complicated to communicate than a dichotomous decision rule and the encouraging individualized decisions that require more nuanced quality measures.

The main limitations of our study relate to the uncertainty inherent in the model assumptions. For example, there is debate about whether the clinical effects of aspirin significantly differ in men and women, a disagreement that explains much of the substantial differences between the American Heart Association and US Preventive Services Task Force guidelines. We believe that the clinical evidence that the biological effects of aspirin differ between men and women is too weak to systematically under-treat women, as discussed in online-only Data Supplement Appendix 2. Therefore, our analysis uses the same assumptions as Baigent et al and assumes the same RRR for men and women. As shown in online-only Data Supplement Appendix 2, if aspirin is found to have different clinical effects in women versus men, this would change who is likely to benefit, but it would only make more important the point that the effectiveness of aspirin varies dramatically as a function of a patient’s baseline risk and is of indeterminate benefit in a large proportion of the population.

Models such as the one presented here often are believed to provide insufficient evidence to influence clinical practice, being relegated as “hypothesis generation” studies. However, only the overall population effects of our analyses are based on a simulation model; the findings that we recommend for use in clinical practice come from trial-based estimates of clinical benefit for individual patients. Although it has become traditional to use the average risks and benefits found in RCTs to make recommendations, we propose that our method provides more clinically useful and accurate estimates of an individual’s net benefit from treatment. For example, for diagnostic testing, we would not apply a study’s average positive predicted value to individual patients. Better estimates for individual patients are achieved by using the studies’ likelihood ratios together with an individual patient’s pretest probability using Bayes formula to estimate the positive or negative predictive value. Similarly, estimates of the RRR from RCTs should be obtained from RCTs (including examinations for interaction effects) and clinical practice guidelines, and clinicians should consider a patient’s risks if not treated in order to individualize care along the lines recommended in this article. We contend that this approach is more evidence based than one that ignores the large variations in absolute risk reduction or increase observed in RCTs.

However, there are also reasons that our proposed approach could improve adherence by better targeting patients for whom treatment is a high priority or is inappropriate. Whether the slight increase in complexity alters guideline adherence merits investigation in effectiveness studies. Although we do not believe that a new RCT to further test the efficacy of our proposed strategy on hard outcomes is needed or feasible, an implementation study could examine competing guidelines effectiveness in promoting the 2 most important measures of high quality: (1) high-risk-benefit patients who receive treatment (low undertreatment) and (2) patients who do not receive inappropriate treatment (low overtreatment and mistreatment).

In summary, given the variation in the clinical benefit of aspirin and the importance of individual preferences in assessing a person’s risk and benefit from aspirin, treatment guidelines could be made more patient centered by discarding dichotomous thresholds in favor of a more nuanced approach. Our findings suggest that guidelines should focus more on helping clinicians and patients to understand individual risk/benefit profiles and focus less on establishing treatment recommendations based on a single threshold. For most adults, aspirin would have at most a small benefit, and shared decision-making is a more sensible approach than definitive treatment recommendations. Finally, the present study develops a model for clinical guidelines that relies on clinical trials for determining a treatment’s relative effects, but applies those results to individuals by considering their personal risks and benefits, including their attitudes about the treatment, which can vary substantially even for people of the same age or similar medical conditions. By tailoring the relative treatment effects found in clinical trials to an individual person’s risks and preferences, we can foster clinical care that is both evidence based and patient centered.

Acknowledgments
We thank Emily Wu for her contributions toward this article.

Sources of Funding
This work was supported by the Robert Wood Johnson Foundation Clinical Scholars Program and an associated Veterans Affairs (VA) Advanced Fellowship as well as the VA Health Services Research and Development Service’s Quality Enhancement Research Initiative (QUERI DIB 98-001) and the Measurement Core of the Michigan Diabetes Research and Training Center (National Institute of Diabetes and Digestive and Kidney Diseases) of the National Institutes of Health (P60 DK-20572). None of the funding sources had any role in the design and conduct of the study; collection, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.
Disclosures
Dr Sussman has received research grant support from the Robert Wood Johnson Foundation Clinical Scholars Program and a VA Advanced Fellows Training Grant. Dr Vijan has received research grant support from the VA Health System. Dr Hayward has received research grant support from the VA Health System and the National Institute of Diabetes and Digestive and Kidney Diseases.

References


SUPPLEMENTAL MATERIAL

Appendix 1: Details of the simulation model

The model was calibrated to estimate rates of fatal and nonfatal events and the relative event rate of each event for men and women at each age group. We used multiple sources of data, focusing on reliable sources. When inconsistencies appeared, we first assumed the validity of National Center for Health Statistics data,1, 2 then American Heart Association,3 then Framingham Heart Study,4, 5 then other data.

Below we describe how we estimated event rates, fatality per event, effect of non-fatal events on future mortality, and effect of fatal and non-fatal events on discounted quality-adjusted life-years.

Event rates

The rates of MI and total stroke were derived from the National Center for Health Statistics 1, 2 and American Heart Association statistics.3 The risk factors for MI and total stroke were derived from Framingham Risk Scores.4, 5 The rate of ICH was developed from previously-published multivariate risk assessment from a large, prospective database.6 Risk factors for ICH in the model were older age, African-American race, hypertension, lower triglycerides, and lower LDL. We utilized a similar method for development of the GIB scale. Based upon the scale development of Saini et al,7 the GIB scale included age and use of oral steroids, anticoagulants or NSAIDs, and history of GI bleed. Overall rates of GIB were based upon population data.8
Event rates on aspirin were all multiplied by the relative risk reduction or increase from aspirin determined in Baigent, et al⁹ and listed in Appendix Table 1.

Appendix Table 1: Fatal MI and CVA rates compared to NCHS Cause of Death for all event-related mortality to results of the model. Rates are events per 100,000 person-years. Note that the CDC estimates all event-related mortality and our model estimates first-event related mortality. Since second MI are substantially more common and have higher fatality rates than second CVA, the difference between the CDC findings and the model and it’s widening with age were both intended.

<table>
<thead>
<tr>
<th></th>
<th>MI CDC</th>
<th>MI Model</th>
<th>CVA CDC</th>
<th>CVA Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>280</td>
<td>158</td>
<td>61</td>
<td>62</td>
</tr>
<tr>
<td>35-54</td>
<td>60</td>
<td>49</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>55-85</td>
<td>593</td>
<td>333</td>
<td>133</td>
<td>114</td>
</tr>
<tr>
<td>Men</td>
<td>342</td>
<td>219</td>
<td>59</td>
<td>56</td>
</tr>
<tr>
<td>Women</td>
<td>223</td>
<td>118</td>
<td>62</td>
<td>56</td>
</tr>
</tbody>
</table>

*Event mortality rates*

Fatality rates per event for MI and ISC were developed from National Center for Vital Statistics Causes of Death² data and the event rates with some smoothing to account for annual variation in mortality and the American Heart Associations’ Heart Disease and Stroke Statistics.³ Mortality per hemorrhagic stroke was obtained from AHA statistics and estimated at 50%. GI bleed rates and mortality per serious GI bleed was derived from Rockall et al.⁸ The rates of mortality per GI bleed were 2% for people ages 40 and younger, 5% ages 41-50, 7% ages 51-60, 10% ages 61-70,
and 12% ages 70 and higher. These results were consistent with the NCVS Causes of Death data.(12)

**QALY Assessment**

To determine the effect of fatal cardiovascular events on QALYs, the time of the event was subtracted from the expected years of life remaining from NCHS Life Tables,¹ with a 3% discount rate. This uses an assumption of 1 QALY per year lost.

Calculation of the effect of a non-fatal event on QALY’s included three factors – a decrease in quality of life the year of the event, a smaller decrease in quality of life every remaining year after the initial event, and a reduction in life expectancy resulting from the non-fatal events. The effect of events on quality of life were obtained from the published literature. The effect of events on future survival was obtained from AHA statistics and prior cardiovascular modeling studies. As seen below, non-fatal MI and ICH were assumed to double expected mortality, ISC tripled mortality if it occurred at less than age 60, doubled mortality if it occurred at great than age 60, and GI bleed did not alter follow-up mortality rates. ³ We used a 3% discount rate.

\[
\text{SMR} = \text{standardized mortality rate (age, sex, race adjusted mortality rate)}
\]

\[
\text{Mortality rate per non-fatal MI}^{10, 11} = 2 \times \text{SMR}
\]

\[
\text{Mortality rate per non-fatal ISC}^{12, 13} = 3 \times \text{SMR if age < 60} \quad \text{OR}
\]

\[
= 2 \times \text{SMR if age } \geq 60
\]

\[
\text{Mortality rate per non-fatal ICH}^{14} = 2 \times \text{SMR}
\]

\[
\text{Mortality rate per non-fatal GIB}^{15} = 1 \times \text{SMR (No long term effect)}
\]
MODEL ASSESSMENT

Model calibration

To assess calibration we compared the basic epidemiology of the cohort model with National Center for Health Statistics (NCHS) data on cause of death\textsuperscript{1,2}, overall survival rates, and the patient-level meta-analysis that we used to determine the effects of aspirin.\textsuperscript{9} As seen in Appendix Table 1, the model appropriately estimated overall fatal event rates. While information on fatality from first event was unavailable, AHA data describes a similar number of first MI and first stroke (785,000/year vs. 610,000/year), but far more recurrent MI (470,000) than recurrent stroke (185,000).\textsuperscript{3} Since recurrent MI is associated with significantly higher mortality than first event, the model was calibrated to have fewer MI deaths than NCHS data. Our model also calibrated between conditions successfully. The model’s ratio of total CHD death to stroke death (slightly under 3:1) is very similar to that seen in the primary meta-analysis,\textsuperscript{9} the NCHS data,\textsuperscript{1,2} and the AHA statistics.\textsuperscript{3}

The table below demonstrates calibration of our model compared to NCHS data for specific subgroups. Compared to overall fatal event rates, we effectively differentiated the rate of event by age and gender. The lower rate of fatal events in men over age 55 was to account for mortality from recurrent MI.
Model Stratification

Since a central part of our hypothesis was that risk prediction could be utilized to guide medical decision-making, having risk models that effectively stratified by medical condition was important.

Our model did effectively stratify by risk, although the Framingham Heart Score had substantially better stratification than the other scales. People in the 75th percentile of event rate for MI had 7 times the predicted event rate than those in the 25th percentile, but 75th percentile for ISC, ICH and GIB was three times the event rate for the 25th percentile. Age was by far the most important predictor of all four event-types.

Appendix Table 2 shows the range of values that the clinical prediction scales were able to predict, in events per 10,000 years of use. For example, someone in the 10th percentile of risk for MI has a 2.3 per 10,000 chance of having an MI in the next year, whereas someone in the 90th percentile has a 103/10,000 chance.

Appendix Table 2: Rate of first cardiovascular event per 10,000 years of aspirin use.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>MI</th>
<th>ISC</th>
<th>ICH</th>
<th>GIB</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>2.3</td>
<td>10.6</td>
<td>0.7</td>
<td>3.8</td>
</tr>
<tr>
<td>25%</td>
<td>6.7</td>
<td>10.7</td>
<td>1.0</td>
<td>5.3</td>
</tr>
<tr>
<td>50%</td>
<td>22</td>
<td>10.8</td>
<td>1.6</td>
<td>8.5</td>
</tr>
<tr>
<td>75%</td>
<td>47</td>
<td>29</td>
<td>3.4</td>
<td>17</td>
</tr>
<tr>
<td>90%</td>
<td>103</td>
<td>64</td>
<td>7.2</td>
<td>31</td>
</tr>
</tbody>
</table>
Note that if risk stratification techniques were improved (as with new techniques like Coronary Artery Calcium scoring, etc) that overall relationship between event risk and treatment benefit would become even more clinically important.

*Power calculations for appropriate studies*

We made a rough estimate of the size of study that would be necessary to test our models estimate of the amount of benefit for the “small net benefit” and “substantial benefit” groups, to demonstrate the inherent difficulty of determining benefit in small benefit groups by RCT. The following table (Appendix Table 3) summarizes the event rates and the study-years that would be needed to assess the benefit of this treatment with an alpha of 0.05 and beta of 0.1.

**Appendix Table 3: Power calculations for hypothetical risk-based randomized trials**

<table>
<thead>
<tr>
<th>Event Rate, per 10,000 person-years</th>
<th>Study-years needed, total</th>
</tr>
</thead>
<tbody>
<tr>
<td>On aspirin</td>
<td>Not on aspirin</td>
</tr>
<tr>
<td>&gt;10.6%</td>
<td>142</td>
</tr>
<tr>
<td>6.1-10.6%</td>
<td>49</td>
</tr>
<tr>
<td>&lt;5.5%</td>
<td>21</td>
</tr>
</tbody>
</table>

While these studies would be large and expensive, underused but robust statistical methods exist to test for evidence of heterogeneity of treatment effects as a function of baseline patient risk, and there have been calls for journals and regulatory agencies to be more proactive in insisting that such analyses be done.16, 17 In this way, more reliable information could be provided to those with moderate baseline risk regarding the pros and cons of them receiving a treatment.
Appendix 2: Justification for basing our model around cardiac risk

*Recent history of Clinical Practice Guidelines for aspirin in primary prevention*

In the United States over the last 10 years, multiple groups have provided different recommendations for primary prevention with aspirin. The current American Heart Association guidelines, first published in 2002, recommend treating anyone with a ten-year CAD risk greater than 10%, but provided no detailed analysis justifying this particular cutoff or justification for not also including stroke risk. Until recently, the US Preventive Task Force (USPSTF) recommended the slightly more-aggressive guideline of treating people with a five-year CAD risk greater than 3%. Since these guidelines were published there has been substantial new research.

In 2009, new USPSTF guidelines not only changed who would get treated, but created a new algorithm for determining treatment eligibility. This algorithm was based on an assumption that three separate interactions in aspirin’s effectiveness were true for primary prevention: that it prevented ischemic strokes (ISC) but not heart attacks in women; that it prevented myocardial infarction (MI) but not strokes in men; and that it prevented nonfatal but not fatal events overall. To account for this, they developed a treatment plan that made decision-making for men based on MI and women based on ISC.

Below, we provide details on why these interactions are hypothesized, why we feel that the evidence to support these interactions is weak, and describe the implications of this discrepancy in interpretation. Ultimately, while we agree that this gender-interaction hypothesis should be
investigated further in the future, we feel that this hypothesis is untested, and our study, therefore, was based on the assumption that aspirin is equally effective at preventing cardiovascular events in men and women and at preventing fatal and nonfatal events.

*The evidence for the aspirin-gender interaction is most likely due to chance*

Central to the USPSTF model are three separate interactions—aspirin prevents primary MI but not stroke in men, primary stroke but not MI in women, but secondary stroke and MI in both.9,18 We believe this hypothesis is biologically and epidemiologically unlikely. Biologically, to our knowledge there remains no strong physiologic or animal-science reason for these disparate findings.

Epidemiologically, we believe the consistent effects in secondary prevention are useful evidence against a gender interaction and the evidence itself conflate hypothesis-generating data with hypothesis-demonstration. From what we can determine, the first major publications to hypothesize these interactions strongly appeared after the publication of the Women’s Health Study (WHS),19 a large, high-quality randomized trial of aspirin for the primary prevention of heart disease in women that unexpectedly found a benefit in stroke reduction but no benefit for MI. Since even large studies can have chance findings, scientific standards insist that confirmatory evidence be found to support an unexpected finding. Although a subsequent meta-analysis might appear to have confirmed the interaction hypothesis,9 it did not, since its results are driven by the same study that initially generated the hypothesis. The Women’s Health Study was the source data for 198 of 235 women’s MIs and 221 of 281 women’s strokes in this paper.9,19 Therefore, although both studies were well-conducted, the meta-analysis results for women
was mainly a re-publication of the Women’s Health Study’s unexpected finding, and the results for men had inadequate statistical power to rule out a stroke benefit.

In summary, for the aspirin-gender interaction to be true, there would need to be some unknown phenomenon related to aspirin’s affect on platelet aggregation that results in: 1) similar bleeding risks in men and women, 2) differential effects for men vs. women on MIs and CVAs for primary prevention, but 3) no such interaction for men and women for secondary prevention of MIs and CVAs. Certainly, this hypothesis warrants additional research, but we think these interactions are likely a chance finding and lack adequate evidence upon which to base our model.

**Analysis of gender interaction assumptions**

**Methods**

To assess the implications of the gender assumption, we utilized the same Markov Model under two different circumstances. In the first we used the same assumptions in our primary model. In the second we used the relative risk estimates found in the ATT model\(^9\) separated by gender. These are a relative risk of MI of 1.01 in women, 0.68 in men, ISC 0.76 in women but 1.0 in men, ICH of 1.07 in women and 1.69 in men, and GIB of 2.0 in men and women.

We then estimated the net population benefit of aspirin use at different cardiac risks for the ‘no gender interaction’ assumptions used in the primary results and the ‘gender interaction’ assumptions described above. By this standard, a higher treatment cut-off (such as a FHS of 20%) would treat fewer Americans and have fewer QALYs saved, but more efficiently. This is
demonstrated by the red and blue curves in Appendix Table 2. We also assessed the implications of the new USPSTF guidelines, also for both sets of interaction assumptions.

Results: Implications of the gender interaction assumptions

Although we feel that there is insufficient evidence for the aspirin-gender interaction, our model showed there was little population benefit to be gained from assuming that it is true. As seen in the Appendix Figure, basing clinical decision-making around cardiac risk provides useful, efficient increase in population benefit with decreasing marginal benefit per new patient treated whether the interaction is real or not. In contrast, the costs and risks of basing care around the interaction if it were not to be true proved to be clinically substantial. If the gender interaction is not a true phenomenon, then we estimate that the USPSTF guideline would underperform a simpler Framingham Heart Score (FHS) approach by almost 11% (e.g., treat if FHS > 9%, which treats about the same proportion of the population as the USPSTF guideline). However, even if the gender assumption is true, the benefit of the USPSTF over the simple FHS approach by less than 2%.
Using our model, we assessed the implication of different assumptions of the benefit of treatment with aspirin. The x-axis represents the percentage of the US population treated and the y-axis represents the QALY’s gained nationwide by that system. The curves demonstrate the population benefit of treatment by Framingham Heart Scores at any individual point. So, for example, if all Americans whose 10-year risk was greater than 10.6% (those left of the orange vertical line), 35% of eligible Americans would be treated. The red curve and circle indicate the effects of aspirin use if the gender interaction proves to not be correct—the USPSTF guideline would inefficiently treat 20% more people with the same benefit. However, the blue curve and circle demonstrate that even if the gender assumption is true, the USPSTF model is less efficient than a FHS-based model, in spite of being substantially more-complicated to use. By any of these models, aspirin is under-utilized in people with no opposition to treatment.
Appendix 3: Individual examples

The individual examples below (Appendix Table 6) are designed to show ways that such calculations could be used to aid personalized decision-making by showing the benefit that could be obtained by giving aspirin to realistic patients. As can be seen, the benefit of aspirin use is primarily found among the rare patient with especially high risk. Even at FHS \( \sim 15\% \), we estimate that 32 patients would require five years of aspirin therapy to save a single QALY.
Appendix Table 4: Illustrative examples. Examples are per five years of treatment

<table>
<thead>
<tr>
<th>FHS, %</th>
<th>NNT per QALY saved*</th>
<th>NNT to prevent an event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Net Fatal Events</td>
</tr>
<tr>
<td>Patient A: 74 y.o. man with diabetes, BP 124/64, TC/HDL 230/52, non-smoker**</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>Patient B: 62 y.o. year old man with diabetes, BP 154/73, TC/HDL 254/68, non-smoker</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>67 y.o. man, non-smoker, no diabetes, BP 127/71, TC/HDL 137/54</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>Patient C: 63 y.o. woman with BP 145/78, TC/HDL 230/52, non-smoker, no diabetes</td>
<td>10</td>
<td>42</td>
</tr>
<tr>
<td>Patient D: 51 y.o. women with BP 152/67, TC/HDL 193/60, non-smoker, no diabetes</td>
<td>5</td>
<td>63</td>
</tr>
</tbody>
</table>

* Assumes 0.001 QALY per year of treatment
**BP = blood pressure, TC/HDL = total cholesterol/high density cholesterol
Appendix Table 5: Population-wide implications of different treatment cut-points

<table>
<thead>
<tr>
<th></th>
<th>Number treated*</th>
<th>QALYs saved / year, nationwide†</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 10% Predicted (AHA guidelines)</td>
<td>54</td>
<td>670,000</td>
</tr>
<tr>
<td>&gt; 10.6% Predicted (substantial benefit)</td>
<td>50</td>
<td>650,000</td>
</tr>
<tr>
<td>&gt; 6.1% (smaller benefit)</td>
<td>89</td>
<td>840,000</td>
</tr>
</tbody>
</table>

* In millions, out of 176 million possible
† Assumes 0.001 QALY per year of treatment
Appendix Table 6: Median number needed to treat (NNT) over five years of treatment to save 1 QALY or prevent 1 event of each type.*

<table>
<thead>
<tr>
<th>NNT per QALY saved**</th>
<th>NNT to prevent an event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MI</td>
</tr>
<tr>
<td>Substantial benefit</td>
<td>20</td>
</tr>
<tr>
<td>(&gt; 10.6% 10-year risk, AHA guidelines)</td>
<td></td>
</tr>
<tr>
<td>Smaller benefit (6.1-10.6% 10-year risk)</td>
<td>43</td>
</tr>
<tr>
<td>Small Benefit (&lt; 6.5%)</td>
<td>132</td>
</tr>
</tbody>
</table>

* Assumes 0.001 QALY per year of treatment
** NNT per QALY is the median NNT for that group shown
Appendix Table 7: Summary recommendations. Although the benefit of therapy increases across risk, these dividing lines are chosen to emphasize the varying benefit.

<table>
<thead>
<tr>
<th>Effect</th>
<th>10-year risk</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substantial benefit</td>
<td>&gt; 10.6%</td>
<td>Strongly recommend use</td>
</tr>
<tr>
<td>Small benefit</td>
<td>6.1-10.6%</td>
<td>Recommend use, but tempered by patient preferences</td>
</tr>
<tr>
<td>Small Benefit</td>
<td>&lt; 6.1%</td>
<td>Entirely by patient interest, likely harms and benefits are quite low</td>
</tr>
</tbody>
</table>
Appendix Works Cited


