This report is based on research conducted by an Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No.). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help clinicians, employers, policymakers, and others make informed decisions about the provision of health care services. This report is intended as a reference and not as a substitute for clinical judgment.

This report may be used, in whole or in part, as the basis for the development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Draft Methods Report

Number XX

Evaluation of the Benefit and Harm of Aspirin for Primary Prevention of Cardiovascular Events: A Comparison of Quantitative Approaches

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. [redacted]

Prepared by:
[redacted]

Investigators:
[redacted]

AHRQ Publication No. xx-EHCxxx
<Month Year>

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted for which further reproduction is prohibited without the specific permission of copyright holders.

Suggested Citation:

No investigators have any affiliations or financial involvement (e.g., employment, consultancies, honoraria, stock options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in this report.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review prior to their release.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

Comments are welcomed on this evidence report. They may be sent by mail to the Task Order Officer named below, at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D.
Director, EPC Program
Agency for Healthcare Research and Quality

Parivash Nourjah, Ph.D.
Task Order Officer
Agency for Healthcare Research and Quality

Jenae Benns,
EPC Program Coordinator

The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services of a particular drug, device, test, treatment, or other clinical service.

Acknowledgments

Abstract

Background: Prior work has described various quantitative approaches to the assessment of benefit and harm. Researchers rarely use these approaches in the context of a systematic review.

Objective: Our objectives were to evaluate how well two quantitative approaches to benefit-harm assessment work in the context of a systematic review; and determine methodologic challenges of applying these two quantitative approaches in the context of a systematic review.

Methods: We compared two quantitative approaches, 1) the number-needed-to-treat and number-needed-to-harm and 2) the Gail/National Cancer Institute approach, for assessing the benefit (myocardial infarction and ischemic stroke) and harm (hemorrhagic stroke and major gastrointestinal bleeding) of aspirin for primary prevention of cardiovascular events.

Results: The assessment of benefit and harms of aspirin for primary prevention is sensitive to the choice of the underlying data source (trials or observational studies) to inform risks of events without treatment, the treatment effect of aspirin on these outcomes, and preference weights assigned to these outcomes. This approach may be more relevant when benefit and harm outcomes are few in number, similar in importance, and similar in quality. The Gail approach, on the other hand, considers multiple outcomes of varying importance and provides a single benefit and harm comparison estimate. However this approach requires data from additional sources on baseline risks and weights of outcomes, which may not be routinely available in systematic reviews. More than one quantitative approach may be maximally informative.

Conclusion: The assessment of the benefit and harm requires careful selection and integration of data from disparate sources, including individual risks of events without treatment, the effects of treatments on various outcomes, and weights of these outcomes. Further evaluation is needed to determine whether such quantitative approaches may inform and support the assessment of benefit and harm by decisionmakers.

Contents

Evidence Report 1

Background 3

Methods 5

Results 11

Discussion 15

Limitations 17

Future Research 18

Reference List 21

Table 1. Incidence rates (per 1,000 person-years) without aspiring prevention based on Surveillance data 7
Table 2. Assumptions 8
Table 3. Number needed to treat and number needed to harm for aspirin for primary prevention in men 12
Table 4. Number needed to treat and number needed to harm for aspirin for primary prevention in women 12
Table 5. Expected number of events without and with aspirin prevention in men 13
Table 6. Expected events without and with aspirin prevention in women 13
Table 7. Benefit harm comparison estimates using the NCI approach 14
Table 8. Principles 19

Appendix A. Details of our data-sources for effect estimates, baseline risk of all 4 outcomes, and weights of outcomes 23
Appendix B. List of Acronyms/Abbreviations 25

EVIDENCE REPORT

Background

Systematic reviews conducted by Evidence-based Practice Centers assess the comparative effectiveness and safety of health care interventions, and are useful to a variety of decisionmakers. Quantitative approaches for the assessment of benefits and harms may enhance, support, and facilitate how decisionmakers use systematic reviews.

The first report of this project outlined the challenges and principles of benefit and harm assessment, the influence of values and preference, and the key characteristics of quantitative approaches.[1] That report identified 16 approaches for the quantitative assessment of benefit and harm. Researchers and methodologists developed several of these approaches using data from a single study, but these approaches could potentially also be suitable for systematic reviews. Simpler approaches, such as the ratio of the number-needed-to-treat (NNT) and number-needed-to-harm (NNH), may be suitable for relatively simple decisionmaking contexts where relevant benefit and harm outcomes are few in number, similar in importance, and similar in quality. Decisionmaking contexts that have a larger number of relevant benefit and harm outcomes may need more complex approaches. However, the understanding of the comparative strengths and limitations of these approaches is limited. There have been few comparative empirical applications of these quantitative approaches to assess benefit and harm assessment in systematic reviews.

The specific objectives of this second part of the project were:

1. To evaluate how well one or more quantitative approaches to benefit-harm assessment works in a systematic review
2. To evaluate the methodologic challenges of applying one or more of these quantitative approaches to benefit-harm assessment in a systematic review

Methods

We reviewed a sample of systematic reviews from the www.effectivehealthcare.org website and investigators' reference libraries.[2-9] These systematic reviews addressed a wide variety of clinical questions. We selected the systematic review of the benefit and harm of low-dose aspirin for primary prevention of cardiovascular events because this complex decisionmaking scenario needs to balance the multiple outcomes for benefit (myocardial infarction (MI), ischemic stroke) and harm (bleeding and hemorrhagic stroke).[9] Such a complex scenario allows one to explore how variations in gradient in baseline risk across these outcomes (i.e., without aspirin) and variations in their relative importance and relevance to patients (i.e., preferences) may affect the comparison of benefit and harm of aspirin. The availability of data for both benefit and harm outcomes from meta-analyses of randomized controlled trials (RCTs) also made this a pragmatic choice.

Although one can potentially apply several quantitative approaches to a systematic review key question, we illustrate two quantitative approaches for the assessment of benefit and harm of aspirin for primary prevention of cardiovascular events: NNT and NNH, and the Gail/National Cancer Institute (NCI) approach. We selected the NNT and NNH approach because these are frequently-used metrics to judge the balance of benefits and harms, and therefore maybe often available to decisionmakers. Also, systematic reviews conducted by the Cochrane collaboration routinely report NNT and NNH separately. They provide a natural point of reference for more complex approaches. We selected the Gail/NCI approach because it can consider multiple outcomes, it offers a benefit-harm comparison metric[*], and it can consider competing risks, and it can take into account weights for various outcomes.[10]

Specification of the Decisionmaking Context

We specified a population-based decisionmaking context from the perspective of a hypothetical guideline-maker assessing the benefits and harms of aspirin in the primary prevention of cardiovascular events. The target population was defined as the adult population aged 50 to 84 years in the United States (U.S.) without evidence of cardiovascular disease or stroke. We assessed the use of low dose (75-100 mg) aspirin compared to placebo. The outcomes evaluated were MI and ischemic stroke (benefit) and hemorrhagic stroke and major bleeding which was primarily gastrointestinal (GI) bleeding (harm) associated with the use of aspirin. The unit of number needed to treat was defined as the number of person-years of exposure to treatment needed in order to prevent one event. We considered death from any cause as a competing risk, and a time horizon of 10 years for the Gail approach.[11]

Selection of Data Sources

The data inputs needed for a quantitative approach to benefit and harm assessment include: a) the effect estimates of aspirin on the outcomes of MI, ischemic stroke, hemorrhagic stroke, and GI bleeding; b) the probability of these outcomes without treatment with aspirin; and c)

[*] Approaches that provide explicit quantitative information on the benefit and harm outcomes, such as putting benefit and harm on the same scale to provide a benefit harm comparison metric (e.g. QALYs, probability scale, risk scale, NNT/NNH ratio, etc.), resulting in a benefit-harm comparison estimate.[1]

weights for these outcomes. Certain approaches for quantitative benefit and harm assessment, such as the NNT and the NNH, do not always require an explicit weight for these outcomes and rely on decisionmakers to weigh the outcomes.

There is general scientific consensus that high-quality RCTs, and/or their meta-analysis, provide a reliable source of evidence on treatment effects of interventions, including benefit and harm. However, sometimes data on harm may not be available from RCTs, and researchers will need to retrieve data from additional sources including observational studies.

Retrieving and selecting the appropriate estimates for the probabilities of benefits and harm outcomes *without* treatment, however, often requires additional data sources beyond what RCTs report. Additional data sources may include large observational studies or national surveillance data. Occasionally, when the trial populations reflect the baseline risk (for both benefit and harm outcomes) of the target population of intended users, RCTs can provide a useful source of information on baseline risk without treatment.

Some quantitative approaches for benefit and harm assessment explicitly use weights for various outcomes.[1] RCTs do not provide selection of weights for these outcomes (i.e., how important the outcomes are relative to one another as perceived by patients). Additional evidence from surveys using conjoint analysis or other preference-eliciting techniques can provide information on weights of outcomes.

We followed a transparent selection process. We prespecified that these estimates should be applicable to the population of interest. We relied on a previously conducted meta-analysis of RCTs for the effect estimates of aspirin on the outcomes of MI, ischemic stroke, hemorrhagic stroke, and GI bleeding.[9] Additionally, we searched for data sources that described the occurrence of cardiovascular events and GI bleeding among the target population in cohort studies.[12 13]

We searched for studies that elicited patient preferences for MI, stroke (ischemic and hemorrhagic) and GI bleeding to weight the outcomes. There are no search filters to identify studies on patient preferences and this literature is not indexed consistently. We searched MEDLINE with the PubMed interface using the medical subject headings “patient preference”, “aspirin”, and “cost-benefit analysis”, and used the “related articles” function to search for potential studies. We selected a study that measured the preferences of participants that had not experienced these events (primary prevention population). We relied on a study conducted in the primary prevention population,[14] because the weights assigned by participants who have not experienced events (*ex-ante*) may be different from those who have experienced such events (*ex-post*).

Appendix A shows the details of our data-sources for effect estimates, baseline risk of all 4 outcomes, and weights of outcomes.

Effect Estimates of Aspirin on Benefit and Harm Outcomes

We used the effect estimates from an updated meta-analysis of aspirin for primary prevention for our relevant decisionmaking context.[9] This study reported a relative risk (RR) of 0.86 (95 percent CI 0.74-1.00) for myocardial infarction, RR of 0.87 (95 percent CI 0.73-1.02) for ischemic stroke, RR of 1.35 (95 percent CI 1.01-1.81) for hemorrhagic stroke, and RR of 1.62 (95 percent CI, 1.31-2.00) for aspirin compared to placebo or controls using a random effects meta-analysis. A previous meta-analysis of a smaller dataset by the same authors formed the basis of recommendations by the U.S. Preventive Services Task Force on the benefits and harms of aspirin.[15]

Estimates of Incidence Rates without Aspirin (Baseline Risks)

We used data from the Atherosclerosis Risk in Communities Study, which is a prospective epidemiologic study conducted in four U.S. communities, to obtain baseline incident rates of MI and stroke without treatment.[12]We chose Atherosclerosis Risk in Communities Study over other large cohorts such as the Framingham Study or the Cardiovascular Health Study because of its random population-based sampling. Since we were unable to find appropriate population-based estimates of GI bleeding in the U.S. population, we relied on estimates from the General Practice Research Database, from the United Kingdom, and the Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria, from Spain.[16] To estimate the risk of death, without treatment, we relied on estimates from the Center for Disease Control. [17]

Table 1 shows the incidence rates for untreated men and women. For all outcomes there is a large gradient in incidence rates across age categories, and between men and women. Incidence rates are much higher for GI bleeding than for MI and stroke, but comparable to mortality rates. Mortality was considered a competing risk in the Gail approach.

Table 1: Incidence rates (per 1,000 person-years) without aspirin based on surveillance data

	Incidence rates in men				**Incidence rates in women**			
	45-54	55-64	65-74	75-84	45-54	55-64	65-74	75-84
Myocardial infarction	4.0	6.2	9.3	14.0*	1.2	3.0	4.7	8.2
Major ischemic stroke	1.2	2.5	5.6	10.8	0.9	2.0	3.6	7.5
Major hemorrhagic stroke	0.2	0.4	0.8	1.6	0.1	0.3	0.5	1.1
Major gastrointestinal bleeding	8.0	16.0	30.0	48.0	4.0	8.0	15.0	24.0
All-cause mortality	5.0	10.0	25.0	67.0	3.0	7.0	16.0	48.0

*No reliable data from the Atherosclerosis Risk in Communities Study for men in age category 75-84 years. To estimate the incidence rate we assumed a 50 percent increase from age category 65-74 years based on similar increases in incidences in age category 65-74 to 75-84 in the Framingham Heart Study and the Cardiovascular Health Study.

Weights for Outcomes for Gail/NCI Approach

Any assumption about preferences may have limitations given the paucity of data on preferences. We identified one small study that reported on patients' preferences for various health outcomes with aspirin for primary prevention using a visual analog scale.[14] The patients were free of cardiovascular disease and represented subjects from a primary prevention setting. To assign weights for various outcomes, we used data on preferences for these outcomes from this small sample of participants.[14] We also solicited electronic input from our Technical Expert Panel on the incorporation of these weights for the outcomes identified above.

There are valid criticisms of these weights for outcomes. Participants are likely to assign higher weight to irreversible outcomes such as stroke than reversible outcomes such as GI bleeding. Alternative preference elicitation techniques may lead to alternative valuations, depending on the patient population (primary vs. secondary prevention), the techniques used (time tradeoff vs. standard gamble vs. conjoint analysis), and the outcomes evaluated. Another study, among patients with atrial fibrillation reported a similar gradient across weights of outcomes with severe stroke being the worst outcome, and with GI bleeding being the least

severe outcome (mild stroke was not elicited).[18] Recognizing the paucity of data about preferences for the outcomes relevant to aspirin in primary prevention settings, we also conducted sensitivity analyses using equal weights for outcomes and different weights for outcomes.

Assumptions that Apply to both Quantitative Approaches to Benefit and Harm Assessment

The implementation of number needed to treat and number needed to harm and the Gail/NCI approach required some assumptions. Table 2 describes assumptions that apply to both approaches, except weights are not required for NNT and NNH.

Table 2. Assumptions

Subject	Assumptions*
Heterogeneity of Treatment Effects	We assumed no heterogeneity of treatment effects on the relative scale of aspirin on benefits and harm outcomes based on the results of the 2011 Berger meta-analysis.[9]
Effects by Gender	We assumed that the effects of aspirin on benefits and harm outcomes did not vary by gender based on the result of the 2011 Berger meta-analysis.[9]
Effects of Aspirin over Time	We assumed that the relative risk reductions or relative risk increases of aspirin on benefits and harm outcomes did not change over time.
Incidence Rates over Time	We assumed the incidence rate of all outcomes without treatment did not change over time.
Measures of Severity	We assumed that all MIs were of the same severity, that all strokes were severe and that all major bleedings were major upper or lower GI bleedings because severity was not consistently reported across the trials.
Risk Profiles	We considered different risk profiles of the population, based on characteristics including age and sex. Additional characteristics such as race, blood pressure, or cholesterol levels were not considered because all outcome incidence rates were unavailable for such profiles.
Applicability of Effects	We assumed relative risk reductions in benefits or relative risk increase in harms, from the trials were applicable to the source populations from which estimates of the baseline risks originated.
Weighting of Outcomes (applies only to Gail/NCI approach)	To assign weights for various outcomes, we used data on preferences for these outcomes from a small sample of participants considering aspirin for primary prevention.[14]

*The assumptions above are common to both approaches, except for weighting of outcomes, which does not apply to the NNT and NNH approach.

Number Needed to Treat and Number Needed to Harm

We calculated the NNT for MI and ischemic stroke and NNH for hemorrhagic stroke and GI bleeding (and 95 percent confidence interval [CI]) with aspirin, by applying the (RR) for ischemic stroke, hemorrhagic stroke, and GI bleeding from the above meta-analysis[9] to the population event rates for these respective outcomes from the observational studies [12,16]**(Table 1)** using Visual Rx, version 3.0.[19, 20] We estimated NNTs and NNHs for various age and sex specific categories.

We estimated the NNT and NNH with aspirin in patients with different baseline risks of cardiovascular events and bleeding, as the NNT and NNH varies when aspirin is used in a general population outside highly-selected trial participants.[21] To better account for time at risk, the unit of NNT or NNH was defined as the number of person-years of exposure to treatment needed in order to prevent one event. The NNT is the number of person-years of exposure to treatment with aspirin, rather than with comparator, in order for one patient to be protected from MI or ischemic stroke. The NNH is the number of person-years of exposure to treatment with aspirin, rather than with placebo or comparators, for one additional patient to be harmed by an adverse bleeding event.

Gail/National Cancer Institute Approach

The Gail approach considers multiple patient-important outcomes of a medical intervention and provides profile-specific estimates of the benefit-harm balance. For example, for a patient of a certain age, sex, and with presence or absence of risk factors for the patient-important outcomes, the NCI approach provides a benefit-harm comparison estimate that can inform decisionmakers (patients, health care providers, policymakers, payers) whether treatment will increase or decrease patient-important outcomes, as compared to without treatment, over a certain period of time.

In a first step we calculated the number of events for each of the four outcomes per 1,000 subjects over 10 years, based on the incidence rates described above, and stratified for age and sex. We calculated the number using equation 1:

(1) $$N_{x,p} = 10{,}000*\{(I_x /(I_x+ M_x)\}*[1-\exp\{-10(I_x+ M_x)\}]$$

where Nx, p is the number of events Nx.p is the number of events per 1,000 subjects over 10 years in subjects without aspirin, Ix is the incidence of the event and Mx is all cause mortality, which we treated as a competing risk.

We then calculated the number of events with aspirin for each of the four outcomes per 1,000 subjects over 10 years, again stratified for age and sex. We used equation 2:

(2) $$N_{x,t} = 10000*\{(R_x*I_x /(R_x*I_x+ M_x)\}*[1-\exp\{-10(R_x*I_x+ M_x)\}]$$

where Rx represents the relative risk of aspirin on the outcome (derived from the Berger meta-analyses).

We calculated the difference, Nx, in the number of events per 1,000 over 10 years in subjects with a certain profile between aspirin users and non-users using equation 3:

(3) $$\text{Difference } N_x = N_{x,p} - N_{x,t}$$

Finally, we put all four outcomes on a single scale using equation 4:

(4) $$\text{Index}(W1, W2, W3) = W1 \Sigma Nx + W2 \Sigma Nx + W3 \Sigma Nx + \ldots \quad (4)$$

Where the index represents the benefit-harm comparison metric as the sum (Σ) of differences in events (Nx) for each outcome, using relative weights W, which represent the importance patients attach to certain outcomes in relation to one another. The number of weights depends on how many different weights are used and does not need to equal the number of outcomes. In the original paper about the Gail/NCI approach, for example, three different weights for used: 1.0 for very important outcomes, 0.5 for important outcomes and 0.0 for less important outcomes, which means that latter did not enter the benefit harm analyses. The benefit-harm comparison metric has a positive value if more events are expected with a treatment (in the case of aspirin it would mean more harm than benefit), compared to no treatment or an alternative treatment. The metric has a negative value if the treatment reduces the number of events (e.g., more benefit than harm in the case of aspirin). In the main analysis we used the weights as described above, followed by sensitivity analyses.

Sensitivity Analysis for Gail Approach

One critique of quantitative approaches using weights is that their selection is arbitrary to some extent, unless there is strong evidence on weights from various relevant studies (e.g., from primary prevention populations). To address this, we used alternative weights. We compare the results using the alternative weights to the results incorporating weights derived from a study of benefit and harm outcomes of aspirin for primary prevention.[14]

In the first sensitivity analysis, we assigned equal weights to all outcomes (i.e., "equipreference" for all four outcomes).

In the second sensitivity analysis, we assigned weights just as the NCI did in their example on tamoxifen for the prevention of breast cancer. We assigned a weight of 1 for very important outcomes (severe stroke), 0.5 for important outcomes (MI), and 0 for less important outcomes (GI bleeding). Thus the second sensitivity analysis means that we ignored GI bleeding as a harm outcome.

In the third sensitivity analysis, we varied the source of data for baseline risk and used incidence rates from the trials for MI, stroke, and GI bleeding instead of surveillance data. In this analysis, we used weights from the literature[14] just as we did for the main analysis.

Results

Number Needed to Treat and Number Needed to Harm

Table 3 and Table 4 show the NNT and NNH for men and women respectively.

The NNT is the number of number of person-years of treatment with aspirin, rather than with comparator, in order for one patient to be protected from MI or ischemic stroke. The NNH is the number of number of person-years of treatment with aspirin, rather than with placebo or comparators, for one additional patient to be harmed by an adverse bleeding event.

Both the NNT and NNH for aspirin consistently decline with increasing age because of the increase in incidence rates for all outcomes across age categories. Apart from the treatment effect of aspirin on outcomes (which is assumed constant on the relative scale here), the underlying increasing baseline risk across increasing age categories affect estimates of NNT and NNH. Among men in the oldest age category, a modest 35 percent relative increase in the risk of hemorrhagic stroke with aspirin, (Appendix A) results in very low absolute risk of such events (NNH 1,786), because of the low underlying baseline risk of hemorrhagic stroke.

In contrast, among women aged 45-54, a 62 percent relative increase in the risk of major gastrointestinal bleeding with aspirin, (Appendix A) results in a high absolute risk for that outcome (NNH 404) given the high baseline risk of gastrointestinal bleeding. The NNH for major gastrointestinal bleeding for other age and sex specific categories are even smaller given the relatively higher estimates of baseline risk of GI bleeding across these categories.

Since the estimates of some of the rare harmful effects, such as hemorrhagic stroke, are imprecise, there is much more uncertainty around these estimates. The 95 percent CIs for ischemic stroke include the possibilities of both benefit and harm (RR 0.87, 95 percent CI 0.73-1.02) and thus are reported as NNT as well as NNH, respectively. This illustrates the challenge of conveying sampling uncertainty to end-users.

Our estimates for the NNT and NNH in person-years must be distinguished from that reported by Berger et al. for the outcomes of major adverse cardiovascular events and bleeding in the trials.[9] They reported a NNT of 253 (95 percent CI 163-568) to prevent one major cardiovascular event and NNH of 261 (95 percent CI 182-476) to cause one major bleed with aspirin over the mean duration of the trials (6.9 years). Thus they estimated that with 1,000 people treated with aspirin for 5 years, approximately three excess cardiovascular events would be prevented whereas there would be approximately three excess major bleeds. However we chose to report separate NNTs for MI and ischemic stroke, as these were deemed to be dissimilar outcomes of unequal importance to participants. We also considered that the trial event rates did not reflect the underlying risk of our target population.

Table 3. Number needed to treat and number needed to harm for aspirin for primary prevention in men in person-years #

Age	45-54	55-64	65-74	75-84
Number Needed to Treat (95 percent Confidence Intervals)				
Myocardial infarction	1786 [962-NA]	1153[621-NA]	769 [414-NA]	511 [275-NA]
Major ischemic * stroke	6411[NNT3087-NNH41667]	3077[NNT1482-NNH 20000]	1374[NNT 662-NNH 8929]	713[NNT343-NNH4630]
Number Needed to Harm (95 percent Confidence Intervals)				
Major hemorrhagic stroke	14286[6173-500001]	7143[3087-250001]	3572[1544-125001	1786[772-62501]
Major gastrointestinal bleeding	202 [125-404]	101[63-202]	54[34- 108]	34[21-68]

Relative risk for MI, ischemic stroke, hemorrhagic stroke and GI bleed from Berger et al 2011

Baseline risk of MI, ischemic stroke, hemorrhagic stroke and GI bleed for age and sex specific categories from ARIC and observational studies in Table 1

***** The 95 percent CIs for ischemic stroke (RR 0.87, 95% CI, 073 -1.02) include the possibilities of both benefit and harm and thus are reported as Number Needed to Treat (NNT) as well as Number Needed to Harm (NNH) respectively The NNT is lower because it reflects the lower limit of the CI and NNH is higher because it reflects the upper limit of CI

NA= Not applicable because confidence intervals for relative risk approximate 1

Table 4. Number needed to treat and number needed to harm for aspirin for primary prevention in women in person-years #

Age	45-54	55-64	65-74	75-84
Number Needed to Treat (95 percent Confidence Intervals)				
Myocardial infarction	5953 [3206-NA]	2381[1283-NA]	1520 [819-NA]	872[470-NA]
Major ischemic stroke*	8548[NNT4116-NNH55556]	3487[NNT1852-NNH 25000]	2137[NNT1029-NNH13889]	1026[NNT494-NNH 6667]
Number Needed to Harm[95 percent Confidence Intervals]				
Major hemorrhagic stroke	28572[12346-1000001]	9524[4116-333334]	5715[2470-200000]	2598[1123-90910]
Major gastrointestinal bleeding	404[250-807]	202 [125-404]	108 [67-216]	68[42-135]

Relative risk for MI, ischemic stroke, hemorrhagic stroke and GI bleed from Berger et al 2011

Baseline risk of MI ischemic stroke, hemorrhagic stroke and GI bleed for age and sex specific categories from ARIC and observational studies in Table 1

***** The 95 percent CIs for ischemic stroke include the possibilities of both benefit and harm and thus are reported as NNT as well as NNH respectively

NA= Not applicable because confidence intervals for relative risk approximate 1

Gail/NCI Approach

Tables 5 and 6 show the absolute number of expected events for 1,000 men (Table 5) or women (Table 6) over 10 years without and with aspirin prevention, respectively. Take for example, men age 55 to 64 years. The expected number of MIs per 1,000 untreated men over 10 years is 57. This is a little less than 62 (Table 1, 10 times the incidence rates per 1,000 person-

years), because we considered death as a competing risk that prevents MIs from occurring in those who die early. As age increases, the impact of death becomes larger. For example, the expected number of GI bleedings in untreated men age 75 to 84 years is 285 over 10 years, considerably lower than the 480 that would be expected without considering death as a competing risk.

Table 5. Expected number of events without and with aspirin prevention in men

	Number of expected events over 10 years per 1,000 men							
	Without aspirin				**With aspirin**			
	45-54	55-64	65-74	75-84	45-54	55-64	65-74	75-84
Myocardial infarction	38	57	79	96	33	49	68	83
Major ischemic stroke	12	24	48	74	11	21	42	64
Major hemorrhagic stroke	2	4	7	11	2	5	10	15
Major gastrointestinal bleeding	75	141	231	285	119	218	344	411

All-cause mortality is considered as a competing risk

Tables 5 and 6 also allow for a comparison between the number of expected events without and with aspirin. For example, the difference in the number of MIs for men age 65 to 74 years is nine (74-68) over 10 years. This means that nine MIs are prevented by aspirin, based on a relative risk reduction of 14 percent as derived from the meta-analysis by Berger et al.[9] Aspirin reduces major ischemic strokes to a similar extent (-13 percent relative risk reduction). At the same time, aspirin increases the number of major hemorrhagic strokes (+35 percent relative risk increase, e.g., excess of three hemorrhagic strokes for men age 65 to 74 years) and the number of major GI bleedings (+62 percent relative risk increase, e.g., excess of 113 major GI bleedings for men age 65 to 74 years) over 10 years.

Table 6. Expected events without and with aspirin prevention in women

	Number of expected events over 10 years per 1,000 women							
	Without aspirin				**With aspirin**			
	45-54	55-64	65-74	75-84	45-54	55-64	65-74	75-84
Myocardial infarction	12	29	42	63	10	25	37	54
Major ischemic stroke	9	20	33	58	7	17	29	51
Major hemorrhagic stroke	1	3	5	9	2	4	7	12
Major gastrointestinal bleeding	39	74	129	171	62	117	200	260

All-cause mortality is considered as a competing risk

Table 7 shows the benefit-harm comparison estimates for men and women for the main analysis and all 3 sensitivity analyses. When we use empirically derived preferences[14](e.g., major stroke is weighted about twice as much as MI and about 8 times as much as GI bleeding), aspirin causes more harm than benefit in all age categories of men and women. For example, if 1,000 women with age 65-74 years receive aspirin over 10 years, we can expect 15 harm events after taking all prevented MI and ischemic stroke into consideration. If we do not weight the outcomes (equipreference, sensitivity analysis I), the harm from aspirin is much greater than the benefit. Higher incidence rates for GI bleedings and a larger effect estimate (as compared to the other outcomes) drives this result. In the sensitivity analysis II, where we weight strokes with 1 (very important outcome), MI with 0.5 (important outcome), and GI bleeding with 0 (minor outcome), aspirin provides benefit for all sex and age categories.

Finally, when we base the analyses on incidence rates of the placebo groups in the trials, there is more benefit than harm from aspirin. Importantly, we can make no age-specific benefit-harm comparisons because incidence rates from trials are only available for the entire trial populations. For a participant with a mean age of 53 to 65 years, the incidence rates in the trials were comparable to those from surveillance data, with the important exception of GI bleeding (see next paragraph). The incidence rate in the trials for MI was 5.6 per 1,000 person-years for men and 1.0 per 1,000 person-years for women. Compared to the incidence rates in Table 1, the trial-based incidence rates compare well to those observed in the same age categories (45 to 54 and 55 to 64 years). We observed similar age-specific incidence rates for major ischemic strokes (2.1 per 1,000 person-years for men and 1.1 per 1,000 person-years for women), major hemorrhagic strokes (0.3 per 1,000 person-years for men and 0.2 per 1,000 person-years for women), and for death (7.9 per 1,000 person-years for men and 3.5 per 1,000 person-years for women). Thus, the difference in the expected number of MI, major strokes, and death for 1,000 men and women with or without aspirin treatment was similar, regardless of whether we based incidence rates on surveillance data or trial data.

However, there was a major difference in the incidence rates for GI bleeding, comparing trial data to population-based surveillance data. In the trials the incidence of GI bleeding was much lower (1 per 1,000 person-years for men and 0.5 per 1,000 person-years for women) than in the population-based surveillance data (by a factor of about 15-20, which is very large). As a consequence, using the incidence rates of GI bleeding from trials, the number of GI bleedings caused by aspirin is much lower than found in population-based surveillance data; and, as a result, the benefit-harm comparison estimates are all negative and thus favor aspirin.

Table 7. Benefit harm comparison estimates using the NCI approach

	Men				**Women**			
	45-54	**55-64**	**65-74**	**75-84**	**45-54**	**55-64**	**65-74**	**75-84**
Main analysis using empirically derived utilities	9[1]	17	24	24	5	9	15	17
Sensitivity analysis I, equipreference of outcomes	37[2]	67	99	108	21	38	63	76
Sensitivity analysis II, only very important and important outcomes	-6[3]	-9	-14	-19	-2	-6	-9	-14
Sensitivity analysis based on trial event rates III	-6				-1			

[1]Positive values = excess events = aspirin harmful; Negative values = prevented events = aspirin beneficial

[2] Number of events were standardized to account for the different sum of weights (sum weights is 2.43 for main analysis and sensitivity analysis III, 4 for sensitivity analysis I and 2.5 for sensitivity analysis II) so that the main and sensitivity analyses are comparable

[3] Weight of 1 for very important outcomes such as major stroke, 0.5 for important outcomes such as MI and 0 for GI bleeding

[4] Mean age across trials 53 to 65 years, no age-specific incidence rates and benefit-harm comparison estimates available

Discussion

The conduct of a quantitative approach to an assessment of the benefit and harm requires integration of data from disparate sources, including individual risks of events without treatment, the effects of treatments on various outcomes, and the preference weights of these outcomes. The results illustrate how the combined effect of incidence rates, treatment effects, and the weights of these outcomes, drives the analysis. If incidence and treatment effect is large for a given outcome, this outcome will largely drive the benefit-harm assessment, even if we weight the outcome very little. For example, aspirin is beneficial only when we ignore GI bleeding or when we obtain incidence rates for GI bleedings from the placebo groups of trials. The discrepancy between the main results and sensitivity analyses of the Gail/NCI approach illustrate how the sources of evidence on incidence rates and weights can have a major impact on results. For instance, if we want to evaluate the balance of benefit and harm for a target population that is quite different from the trial participants, then the incidence rate from the placebo arm might not be appropriate. For the sake of simplicity, we assumed that the treatment effects from trials applied to our target populations. However, one must be careful that, in the presence of heterogeneity of treatment effects, the average treatment effect estimate for the trials may not be entirely applicable to the entire target population. Therefore, a crucial step in the planning of a quantitative approach to a benefit-harm assessment is careful consideration of data sources as well as the planning of sensitivity analyses. This would have an impact on the protocol for a systematic review used to inform such a benefit-harm assessment and has both time and resource implications for those who may wish to consider such an approach.

We considered these quantitative approaches in light of principles that are potentially relevant to the conduct of quantitative approaches to benefit harm assessment in the context of systematic reviews. These are summarized in Table 8.

Our results demonstrate several drawbacks of the NNT and NNH approach. In considering aspirin for primary prevention, one would evaluate two beneficial outcomes (MI and stroke), and two harmful outcomes (hemorrhagic stroke and GI bleeding) with their respective NNTs and NNHs and confidence intervals. The communication of uncertainty, such as the confidence intervals around ischemic stroke which reflect the possibility of both benefit and harm, is challenging. Some have argued that NNT or NNH should not be reported when there is uncertainty around the possibility of both benefit and harm. Weighting multiple outcomes of varying importance may also be cognitively challenging for decision makers. It is potentially possible to use NNT and NNH to incorporate weights from different stakeholders to generate relative value NNTs and relative value NNHs.

Our results also demonstrate the trade-offs between more simpler and complex quantitative approaches for benefit and harm assessment in systematic review. Although the Gail approach provided a single benefit and harm comparison metric, it was unable to convey the uncertainty around the comparison metric. Further methodological modifications could incorporate uncertainty.[10] Simpler approaches can be more easily computed directly from the results generated from meta-analysis. However these simpler approaches may also have strong assumptions, and postpone the task of weighting of outcomes for decision makers. In contrast, more complex approaches which generate a benefit-harm comparison metric require additional data on baseline risk and weights. Both complex and simpler approaches may include the inherent assumption about the applicability of these data to the intended target population.

Our findings demonstrate the feasibility and utility of such methods, but are not clinically directive. Earlier analysis by the U.S. Preventive Services Task Force in 2009 showed a benefit of aspirin for people with moderate to high cardiovascular disease risk.[22] However, the analysis assumed a relative risk reduction of about 32 percent in MI from a 2006 meta-analysis[22] of six primary prevention trials using a Peto odds ratio approach.[15] However the updated meta-analysis reported a smaller relative risk reduction for MI of about 14 percent[9] using the relative risk method. Although the USPSTF analysis used similar baseline risk estimates, it did not consider death as a competing risk, or incorporate relative weights and could not explore the effect of a gradient in weights across outcomes.

Limitations

A quantitative approach to the assessment of benefit and harm is limited by the quality of reported data. We obtained our estimates for the treatment effects for both benefit and harm outcomes from a meta-analysis of large RCTs of aspirin for primary prevention, which increases our confidence in these estimates. Trial incidence rates were comparable for myocardial infarction and stroke but not for GI bleeding. The observational studies reported several-fold higher risk of gastrointestinal bleeding than the trials. The possible reasons for this include less rigorous ascertainment of GI bleeding in the trials because it was not a primary outcome in the trials. Heterogeneity in GI bleeding definitions is possible across trials. Finally, the trials excluded participants at high risk for GI bleeding (e.g., previous GI bleedings due to aspirin or NSAIDs) but the surveillance data included such patients.

We did not investigate the possibility that benefit-harm balance varies over time, due to lack of individual-level data. Although methods for estimating NNT and NNH from survival analysis data have been proposed, access to only summary data from the meta-analysis precluded such estimation.[23] We could not consider joint dependence of benefit and harm without individual-level data or reporting of the joint dependence of benefit and harm. Although we evaluated low-dose aspirin, we included studies of higher doses that were tested in large primary prevention trials. The updated meta-analysis included some patients with diabetes and vascular disease, who may be at higher risk of cardiovascular events.

We used weights obtained from a sample of participants considering aspirin for primary prevention. We assumed that these perceptions of the importance of MI, stroke, and GI bleedings, as expressed on a visual analog scale, reflected their relative weights for these health states. Preference elicitation using either the analytic hierarchy process or conjoint analysis with patients assess relative weights may be suitable for exploring the heterogeneity of treatment preferences, but was beyond the scope of this project. Benefit and harm estimates remain sensitive to assumptions about the heterogeneity of treatment preferences in the population.[24] Real heterogeneity in patient preferences would be important to inform the decisionmaking process, but frequently we do not know how much variability among preferences exists. Thus sensitivity analyses are always warranted.

Future Research

While the literature clearly defines methods for selection of RCTs for inclusion in a systematic review, the appropriate selection of additional data sources, beyond what is typically included in systematic reviews, needs further development. Future research should address how to select the most valid and applicable incidence data,[25] and how to select and rate the quality of preference assessment studies.[26] The reporting and assessment of heterogeneity of treatment effects in systematic reviews could be further improved.[27]

Although we considered estimates stratified by age and sex, we did not consider additional characteristics such as race, blood pressure, or cholesterol levels, because stratified outcome incidence rates were unavailable for these variables. The challenge will be to balance the need for finely granulated data (e.g. incidence rates stratified for four to five variables) and accurate estimates of incidence. One could consider reporting incidence rates from surveillance stratified for these variables. However, limited sample sizes may become a challenge even in large studies such as Atherosclerosis Risk in Communities. An alternative would be to estimate outcome risks based on risk prediction models such as the Framingham Risk Index. However, outcome predictions may be poorly calibrated if the models were not developed or updated in the population of interest.

It is uncertain how these approaches will perform when outcome data on harms are more sparse and heterogenous, as is typical in many systematic reviews. For this illustrative example, we did not have a situation where robust evidence is available for benefit outcomes but little evidence is available for harms, which is typical for many other clinical questions. Such a situation would add a layer of complexity to quantitative approaches to benefit-harm assessments, and is beyond the scope of this report. An evaluation of these quantitative approaches in the context of intermediate outcomes (such as forced expiratory volume in 1 second in chronic obstructive pulmonary disease or glycated hemoglobin in the context of type 2 diabetes mellitus) is needed. In these situations, systematic reviewers and investigators will either have to elicit data on preferences for intermediate outcomes (surrogate outcomes) or make assumptions about linkages between intermediate outcomes to health outcomes (patient-important outcomes). Future research on these and other quantitative approaches (such as probabilistic simulation and multicriteria decision analysis) should consider appropriate methods of capturing and conveying the uncertainty around the benefit-harm assessment and assessing and reporting the linkages between surrogates and clinical outcomes.

Conclusion: The assessment of the benefit and harm requires careful selection and integration of data from disparate sources, including individual risks of events without treatment, the effects of treatments on various outcomes, and weights of these outcomes, but are potentially feasible in the context of systematic reviews. Decisionmakers need to evaluate whether such quantitative approaches can facilitate, support and inform their evidence-based decisions regarding treatment.

Table 8. Principles

Principles[1]	**Adherence with number needed to treat and number needed to harm approach**	**Adherence with National Cancer Institute approach**
Present benefits and harms in such a way as to enable the assessment of the balance of the benefits and harms in different decisionmaking contexts	Results stratified for age and sex Data considered that inform a primary prevention decisionmaking context	Results stratified for age and sex Data considered that inform a primary prevention decisionmaking context
Identify the key potential harms in the assessment and the key (usually clinical) outcomes that will be considered as benefits	Benefit and harm outcomes identified from an updated systematic review Considered all outcomes commonly considered important in guidelines, trials, and observational studies We did not interview subjects from a primary prevention setting to ask about outcomes	Benefit and harm outcomes identified from an updated systematic review Considered all outcomes commonly considered important in guidelines, trials, and observational studies We did not interview subjects from a primary prevention setting to ask about outcomes
Consider the different characteristics and assumptions of the existing quantitative approaches to assess the balance of benefit and harm to determine the appropriate approach for a specific decisionmaking context	Assumptions stated on Page 12	Assumptions stated on Page 12
State whose preferences were considered, how these were ascertained, and how variation in preferences would affect the results of an assessment of the balance of benefits and harms	NNTs and NNHs allow end users to put respective weights on outcomes	We used weights from a study for aspirin among primary prevention and considered alternative weights (equal weights)
Convey all convey all major sources of uncertainty, uncertainty that arises from sampling as well as that due to quality of the evidence, in the assessment of the balance of benefit and harm.	Conveys statistical uncertainty on outcomes through 95 percent CI in Table 2 and Table 3	Sensitivity analyses done No estimates of uncertainty (such as 95 percent CI), but probability of index being positive or negative could be done using Markov modeling[10]
Report information fundamental to the assessment of the balance of benefits and harms ("information	Unable to account for joint distribution of benefit and harm	Does not consider joint distributions of benefit and harm, but considers competing

preserving”)		risk of mortality
State how decisions about comparisons, outcomes, baseline risks, and time horizons were made to increase transparency and traceability	See decisionmaking context on page 5	See decisionmaking context on page 5

Reference List

1 Boyd CM, Singh S, Varadhan R, Weiss C, Sharma R, Puhan M. Methods for Benefit and Harm Assessment in Systematic Reviews. Reviews of Evidence. Evidence Report/Technology Assessment No. XXX. (Prepared by Johns Hopkins University Evidence-based Practice Center under contract No. **290-2007-10061-I**). AHRQ Publication No. XX-XXXX-X. Rockville, MD Agency for Healthcare Research and Quality. In press.

2 Bennett WL, Wilson LM, Bolen S *et al.* Oral Diabetes Medications for Adults With Type 2 Diabetes: An Update [internet]. Agency for Healthcare Research and Quality (US); 2011 Mar. Report No.: 11-EHC038-EF 2011.

3 Matchar DB, McCrory DC, Orlando LA *et al*. Comparative Effectiveness of Angiotensin-Converting Enzyme Inhibitors (ACEIs) and Angiotensin II Receptor Antagonists (ARBs) for Treating Essential Hypertension [Internet]. Agency for Healthcare Research and Quality (US); 2007 Nov. Report No.: 08-EHC003-EF. 2007.

4 Bravata DM, McDonald KM, Gienger AL *et al*. Comparative Effectiveness of Percutaneous Coronary Interventions and Coronary Artery Bypass Grafting for Coronary Artery Disease. Agency for Healthcare Research and Quality (US); 2007 Oct. Report No.: 08-EHC002-EF 2007.

5 Coleman CI, Baker WL, Kluger J *et al*. Comparative Effectiveness of Angiotensin Converting Enzyme Inhibitors or Angiotensin II Receptor Blockers Added to Standard Medical Therapy for Treating Stable Ischemic Heart Disease. Agency for Healthcare Research and Quality (US); 2009 Oct. Report No.: 10-EHC002-EF 2009.

6 Bolen S, Wilson L, Vassy J *et al*. Comparative Effectiveness and Safety of Oral Diabetes Medications for Adults With Type 2 Diabetes . Agency for Healthcare Research and Quality (US); 2007 Jul. Report No.: 07-EHC010-EF 2007.

7 Bennett WL, Maruthur NM, Singh S *Et Al.* Comparative Effectiveness and Safety of Medications for Type 2 Diabetes: an Update Including New Drugs and 2-Drug Combinations. Ann Intern Med 2011; 154(9):602-13.

8 Chong J, Poole P, Leung B, Black PN.Phosphodiesterase 4 Inhibitors for Chronic Obstructive Pulmonary Disease.Cochrane Database Syst Rev. 2011 May 11;(5):CD002309.

9 Berger JS, Lala A, Krantz MJ, Baker GS, Hiatt WR.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.

10 Gail, M. H.; Costantino, J. P.; Bryant, J.; Croyle, R.; Freedman, L.; Helzlsouer, K., and Vogel, V. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. J Natl Cancer Inst. 1999 Nov 3; 91(21):1829-46.

11 Varadhan R, Weiss CO, Segal JB, Wu AW, Scharfstein D, Boyd C.Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications.Med Care. 2010 Jun;48(6 Suppl):S96-105.

12 Incidence and Prevalence: 2006 Chart Book on Cardiovascular and Lung Diseases. National Institutes of Health: National Heart, Lung, and Blood Institute. Available: Http://Www.Nhlbi.Nih.Gov/Resources/Docs/06_Ip_Chtbk.Pdf .

13 Garcia Rodriguez, L. A., Perez Gutthann, S. (1998). Use of the UK general practice research database for pharmacoepidemiology. British Journal of Clinical Pharmacology, 45(5), 419-425.

14 Man-Son-Hing M, Laupacis A, O'Connor AM, Coyle D, Berquist R, McAlister F. Patient preference-based treatment thresholds and recommendations: a comparison of decision-analytic modeling with the probability-tradeoff technique.Med Decis Making. 2000 Oct-Dec;20(4):394-403.

15 Berger JS, Roncaglioni MC, Avanzini F, Pangrazzi I, Tognoni G, Brown DL. Aspirin for the primary prevention of cardiovascular events in women and men: a sex-specific meta-analysis of randomized controlled trials.JAMA. 2006 Jan 18;295(3):306-13.

16 Hernandez-Diaz S GRL. Cardioprotective aspirin users and their excess risk of upper gastrointestinal complications. BMC Medicine 2006; 4(22).

17 Jiaquan Xu, M.D.; Kenneth D. Kochanek, M.A.; Sherry L. Murphy, B.S.; and Betzaida Tejada-Vera. Deaths: Final Data for 2007. National Vital Statistics Reports. 2010; 58(19).

18 Pignone M, Earnshaw S, Pletcher MJ, Tice JA. Aspirin for the primary prevention of cardiovascular disease in women: a cost-utility analysis. Arch

Intern Med. 2007 Feb 12;167(3):290-5.

19 Cates CJ. Simpson's paradox and calculation of NNT from meta-analysis. BMC Med Res Methodol. 2002;2:1.

20 Cates CJ. Dr Chris Cates' EBM Web site [Web Page]. Available at http://www.nntonline.net . (Accessed 12 February 2012).

21 McQuayHJ, Moore RA. Using numerical results from systematic reviews in clinical practice.*Ann Intern Med.*1997;126(9):712.

22 US Preventive Services Task Force. Aspirin for the prevention of cardiovascular disease: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009 Mar 17;150(6):396-404.

23 Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. BMJ1999;319 (7223): 1492-5.

24 Guyatt GH, Oxman AD, Vist GE *et al*. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336(7650):924-6.

25 Shamliyan T, Kane RL, Ansari MT, Raman G, Berkman ND, Grant M, Janes G, Maglione M, Moher D, Nasser M, Robinson KA, Segal JB, Tsouros S. Development of the Quality Criteria To Evaluate Nontherapeutic Studies of Incidence, Prevalence, or Risk Factors of Chronic Diseases: Pilot Study of New Checklists. Agency for Healthcare Research and Quality; January 2011. Methods Research Report. AHRQ Publication No. 11-EHC008-EF. Available at http://effectivehealthcare.ahrq.gov/.

26 MacLean S, Mulla S, Akl EA, Jankowski M, Vandvik PO, Ebrahim S, McLeod S, Bhatnagar N, Guyatt GH; American College of Chest Physicians.Patient values and preferences in decision making for antithrombotic therapy: a systematic review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines.Chest. 2012 Feb;141(2 Suppl):e1S-23S.

27 West SL, Gartlehner G, Mansfield AJ, Poole C, Tant E, Lenfestey N, Lux LJ, Amoozegar J, Morton SC, Carey TC, Viswanathan M, Lohr KN. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010 Sep. Report No.: 10-EHC070-EF.Comparative Effectiveness Review Methods: Clinical Heterogeneity [Internet]. AHRQ Methods for Effective Health Care.

Appendix A: Details of our data-sources for effect estimates, baseline risk of all 4 outcomes, and weights of outcomes

Table 1. Treatment Effects of Aspirin for the Primary Prevention of Cardiovascular Events

Treatment Effect	Risk Ratio [95% CI][1*]
Myocardial infarction	0.86 [0.74,1.00]
Ischemic stroke	0.87 [0.73,1.02]
Hemorrhagic stroke	1.35 [1.01,1.81]
Major bleeding	1.62 [1.31,2.00]

***** This meta-analysis included the nine major primary prevention trials: British Doctors' Trial (BMD) , Physicians' Health Study (PHS), Thrombosis Prevention Trial (TPT), Hypertension Optimal Treatment (HOT) study, Primary Prevention Project (PPP), Women's Health Study (WHS), Atherosclerosis Trial [AAAT], Japanese Primary Prevention of Atherosclerosis with ASA for Diabetes, Prevention of progression of arterial disease and diabetes (POPADAD) trial. The mean age of the participants ranged from 54.6 years in the Women's Health Study to 65 in the JPAD trial. Most of the trials used low dose daily Aspirin except for the British Doctors study which evaluated 500 mg of daily aspirin and PHS which evaluated 325 mg of aspirin every alternate day and WHS which evaluated 100 mg every alternate day. The mean followup duration ranged from 3.6 years in the Primary Prevention Project to 10.1 years in the Womens Health Study.)

Table 2. Baseline Risk

Indicator (Incidence Rate)	Source (Year)
Myocardial infarction	• ARIC[2]Community Surveillance Component (1987-89, 1990-92, 1996-98) o Baseline *n*=4,000 aged 45-64 (total sample: 15,792). o Hospitalized MI incidence currently documented in men and women aged 35-84. • If population-based estimates are unavailable, data extracted from primary care setting
Major ischemic stroke	• ARIC Cohort Component[2]
Major hemorrhagic stroke	• ARIC Cohort Component[2]
Major gastrointestinal bleeding[4]	• GPRD: population-based database in the United Kingdom (approx. 3 million patients)[3]
All-cause mortality	• U.S. Vital Statistics[5]

Table 3. Weights for Outcomes[6]

Weights (descending severity)	Perceived severity	Complement (Weights used)
Major stroke	0.11	0.89
Myocardial infarction	0.55	0.45
Major gastrointestinal bleeding	0.80	0.20

References

1. Berger, J. S., Lala, A., Krantz, M. J., Baker, G. S., & Hiatt, W. R. (2011). Aspirin for the prevention of cardiovascular events in patients without clinical cardiovascular disease: A meta-analysis of randomized trials. American Heart Journal, 162(1), 115-24.e2. (

2. ARIC/NHLBI surveillance data, Incidence and Prevalence: 2006 Chart Book on Cardiovascular and Lung Diseases, Table 2.1, p.6 and Ibid, Table 4.6, p. 42.

3. Garcia Rodriguez, L. A., & Perez Gutthann, S. (1998). Use of the UK general practice research database for pharmacoepidemiology. British Journal of Clinical Pharmacology, 45(5), 419-425.

4. Hernández-Díaz S, García Rodríguez LA. Cardioprotective aspirin users and their excess risk of upper gastrointestinal complications. BMC Med. 2006 Sep 20;4:22.

5. National Vital Statistics System. (2012). Retrieved from Centers for Disease Control and Prevention website: http://www.cdc.gov/nchs/nvss.htm

6. Man-Son-Hing M, Laupacis A, O'Connor AM, Coyle D, Berquist R, McAlister F. Patient preference-based treatment thresholds and recommendations: a comparison of decision-analytic modeling with the probability-tradeoff technique.Med Decis Making. 2000 ;20(4):394-403

Appendix B. List of Acronyms/Abbreviations

Acronym	Definition
AHRQ	Agency for Healthcare Research and Quality
CI	Confidence interval
GI	Gastrointestinal
NNH	Number-needed-to-harm
NNT	Number-needed-to-treat
RCTs	Randomized controlled trials
RR	Risk ratio