Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Testing of CYP2C19 Variants and Platelet Reactivity for Guiding Antiplatelet Treatment*

Draft review available for public comment from June 19, 2012 to July 17, 2012.

Research Review Citation: Dahabreh IJ, Moorthy D, Lamont JL, Chen ML, Kent DM, Lau J. Testing of CYP2C19 Variants and Platelet Reactivity for Guiding Antiplatelet Treatment. Comparative Effectiveness Review No. 125. (Prepared by Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 13-EHC117-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2013. www.effectivehealthcare.ahrq.gov/reports/final/cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1726
Published Online: September 25, 2013

General Considerations

We thank all reviewers' for their details and constructive comments. Below we provide a point-by-point response to every comment submitted.

Before discussing the reviewers' specific comments, however, we would like to address the concept of grading the strength of evidence (SOE) and the relationship between SOE grades and associated clinical recommendations by decisionmakers. The assessment of the SOE in EPC reports follows a formal framework very similar to that used in the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system. Details on the methods we applied are provided in the Draft Report's Methods section and the cited chapters of the AHRQ Methods Guide. As in all SOE systems the final grade incorporates unavoidable subjective interpretation, however, we strive to make our criteria explicit and the process transparent. The subjectivity in assessing SOE is evident in the substantial discrepancies among the peer reviewers' assessments of the same body of evidence (as summarized in the report): some state that SOE is insufficient or low for all tests, some that there is moderate or high strength in favor of genetic testing and others in favor of testing for platelet reactivity. For formal evidence reviews, such as our Comparative Effectiveness Review, the hope is that using the same operational definitions independent teams evaluating the same body of evidence will reach similar conclusions. In this regard, the similarity of our conclusions (low SOE for prognostic effects and insufficient for decisionmaking) to those of an independent systematic review team using the GRADE system is notable (Bauer et al. BMJ 2011).

Importantly, SOE grades are not clinical or policy recommendations—and the EPC's mandate specifically precludes making such recommendations. Instead, decisionmakers (such as the FDA, or guideline-issuing bodies) are expected to use the review of the evidence and the SOE grade as one of many inputs into their decision process. Additional inputs may be the relative weighting of different clinical outcomes considered (including the relative weighting of benefits and harms) or public health considerations (accessibility to testing, burden of disease, etc). It is not uncommon that different decisionmakers reach different conclusions when interpreting the same body of evidence: for example, the 2012 American College of Chest Physician guidelines state that "there is no evidence to support the use of platelet function or genetic testing to individualize antiplatelet therapy" (Eikelboom et al. Chest 2012) whereas the FDA-approved label for clopidogrel supports genetic testing for CYP2C19 variants with treatment modification for carriers of two loss-of-function alleles. Thus, we believe that the distinction between the SOE and the clinical recommendations regarding CYP2C19 testing may explain the apparent discordance between the some of the existing recommendations for practice and the report's SOE grade.

Comments From, and Responses to, Reviewers 2, 3, 4 in the Table Below

Because the comments from these three reviewers were identical (with the exception of a single issue raised only by reviewer #2) we have attributed them to all three reviewers jointly (marked as "Reviewers 2, 3, 4" under the Commentator column). Our numbering follows AHRQ numbering of the reviews received (and is blinded in this document).

Source: www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1726
Published Online: September 25, 2013

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	General Comments	Quality of the Report: Superior Number of Hours Spent to Review the Report: 4	Thank you for your comments. Please see below for our point-by-point response.
Peer Reviewer #1	General Comments	As per agreement, only the executive summary was reviewed. I found the report to be overall excellent. The authors have to be commended for performing such a comprehensive review of a very difficult topic. The target audience and key questions are very well defined.	Thank you. No further response necessary.
Peer Reviewers 2, 3, 4	General Comments	The strength of evidence to support an association between the presence of loss of function alleles (LOF) and increased risk of events has been assessed by the reviewers low.	We believe that the evidence at the time of the Draft Report submission had low strength for some of the outcomes assessed. The SOE assessment for AHRQ reviews follows pre-specified methods (as listed in the Methods section of the report). Briefly, our assessment relies on four components: (1) ROB; (2) directness; (3) precision; and (4) consistency. Table B (Executive Summary) of the Draft Report provided a summary of the rationale for our judgment for each exposure-outcome comparison. Note that the report has been updated to include a large number of studies published after our original search strategy (including a new trial of genetic testing versus no testing and the CHARISMA trial genetics substudy). Please see the Final Report for our current assessment of the SOE.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	General Comments	The assessment of quality and bias seems subjective despite concordance between independent reviewers. For example, none of the trials or observational studies were prospectively designed to look at genotype, and consequently, it should be assumed that blinding to genotype is present since the analyses are retrospective.	ROB assessment and SOE grading indeed incorporate subjective judgments. As discussed in the "General Considerations" section we make every effort to follow a systematic and transparent approach. For observational studies, we considered 2 types of blinding (with regards to test performance): (1) blinding the genotyping assessment to clinical outcomes (ROB item Q4) and (2) blinding the outcomes assessment to the genotype results (ROB item Q7). For "repurposed RCTs" we considered 4 types of blinding: (1) blinding the genotyping assessment to clinical outcomes (ROB item Q4); (2) blinding the outcomes assessment to the genotype results (ROB item Q7); (3) blinding of the outcome assessors to treatment assignment (ROB item Q14); (4) and blinding of the patient and caregivers to the treatment assignment (ROB item Q13). All types of blinding were assessed ***as applicable*** to each particular study design. Given that in most cases the exact timing of genetic testing was not reported it is impossible to determine whether genotype ascertainment was blinded to patient outcomes. Similarly, it is impossible to adjudicate whether all genotyping was completed after all outcomes were reported. For this reason in several of the large trials we rated blinding (of "testing to outcomes" and of "outcome ascertainment to test results") as "unclear". We note that this determination does not directly increase the ROB of a study (since poor reporting is distinguished from ROB). Blinding of "outcome assessors to treatment assignment" and blinding of "patients to treatment assignment" was evaluated separately. As discussed below, most large "repurposed RCTs" were based on large well-conducted RCTs assessing treatment effects (this point had been highlighted in the main text of the Draft Report).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	General Comments	Moreover, the trials on which many of these studies are based were sufficiently rigorous to support approval of the drugs and change clinical practice guidelines (i.e., included controls, blinding, and endpoint adjudication).	We agree that the trials on which many of the pharmacogenetic studies were based were sufficiently rigorous to support approval of the drugs and change clinical practice guidelines. However, regulatory approval is of course based on "main effects" of treatment, not the potential for pharmacogenetic interactions. Again, we agree that the large repurposed RCTs included in our analyses were based on well-conducted, large trials. This is reflected in the assessment of appropriate ROB items that were scored separately (e.g., randomized sequence generation, blinding, allocation concealment, etc.), as reported in the main text of the Draft report). However, other ROB items (those relevant to the assessment of a pharmacogenetic interaction) were also scored. As the reviewers' state in their previous comments, the pharmacogenetic analyses of these trials were not pre-planned (in most cases) and were retrospectively performed on subsets of the trial populations (in all large trials the percentage was <50% of the overall trial population).
Peer Reviewers 2, 3, 4	General Comments	Intra-subject variability generally would not be expected to depend on genotype, a priori.	We agree with this point; however, we are uncertain about which statement in the report this refers to. We are happy to reconsider if the reviewers could point us to the relevant section of the text.
Peer Reviewer #6	General Comments	Quality of the Report: Good Number of Hours Spent to Review the Report: 10	Thank you for your comments. Please see below for our point-by-point response.
Peer Reviewer #6	General Comments	The authors analyzed the potential useful of genotyping and platelet function testing in patients treated with clopidogrel by including numerous studies. Based on their extensive and well performed statistical analysis, the authors found limited evidence on the validity of genetic testing for platelet reactivity, whereas a large body of evidence on the analytic validity of assays for measuring platelet reactivity. The authors suggest that no phenotypic assays can be considered a "gold standard" test. They also suggest that additional research is needed to better establish the predictive value and clinical utility for treatment decision making, both for genetic testing for CYP2C19 variants and phenotypic testing for platelet reactivity, focusing on standardizing testing methods and assessing the relative impact of testing strategies on patient-relevant clinical outcomes in large, well-conducted clinical trials.	Thank you for your comments. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General Comments	Although coronary artery related thrombotic event occurrences are “platelet-centric” , they are influenced by multiple factors. Targeting one pathway of platelet function- P2Y12 receptor blockade is just one aspect of treatment strategy. Moreover, clopidogrel is inherently associated with suboptimal pharmacodynamic effect characterized by a wide antiplatelet response variability that is associated with worsened clinical outcome in patients treated with percutaneous coronary intervention. This suboptimal pharmacodynamic effect in turn is related to variable and comparatively low levels of active metabolite generation. The latter phenomenon has been found to be influenced by multiple factors including single nucleotide polymorphisms of genes encoding proteins associated with clopidogrel metabolism particularly cytochrome P450CYP 2C19. However, clopidogrel metabolism is also influenced by drug-drug interactions such as PPIs, calcium channel blockers, smoking (?), and statins (?) etc. The cumulative effect of these these factors is the variable and suboptimal PD response. Therefore, genotyping is considered as a “piece of puzzle” and cannot be considered as a surrogate for phenotyping. While concluding their argument, the authors ignored many important things including the role of drug-drug interactions on PD effect and also on clinical outcome.	We agree that drug-drug, gene-drug, or other interactions are potentially affecting the prognostic value of the tests evaluated in our report. Some of this thinking has been incorporated in the Background section of the Final Report. However, in the Results section we summarize the published evidence on such interactions. The low number of relevant studies, the inconsistencies in their design, analyses, and reporting, and the often discrepant results (when multiple studies have evaluated the same interaction effect) lead us to consider the currently available evidence as insufficient. As we now state in the Methods section of the report, judging the evidence as insufficient does not necessarily imply that no association exists.
Peer Reviewer #7	General Comments	Quality of the Report: Fair Number of Hours Spent to Review the Report: 2	Thank you for submitting these comments. Please see below for a point-by-point response.
Peer Reviewer #7	General Comments	The review is quite comprehensive, but the level of detail is quite overwhelming, making it very hard for anyone reading the report to glean much insight. I read only the 40 page “executive summary”, not the whole 1099 page report. Even that summary is hard to make sense of. It’s not clear to me whether the problem is the very strict requirements of AHRQ in commissioning the report, thereby handcuffing the authors, or whether the authors themselves just lapsed into endless recitation of results. This would benefit greatly from rewriting and condensation into a more user-friendly document.	We regret that the Peer Reviewer felt overwhelmed with the volume of the report. The Executive Summary, at its current length of 34 pages summarizes 325 studies. Note also that some of the language in the Executive Summary is “standard” as per the current CER template. Please also note that the main report is just over 230 pages. It is the extensive appendix that pushes the total page count over 1000 pages. Although a lot of our effort was actually expended toward compiling this appendix, we think most readers only need to review the Executive Summary and main text of the report. However, the extensive appendices are useful for people finding that they need to further explore the topic area to understand the relevance of the results for their patients or themselves.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	General Comments	One point that I didn't see in the document was that the clinical significance of any of the platelet function tests is not very well established.	We have made several changes in the Introduction and Discussion section of the report. We believe that the changes address this issue.
Peer Reviewer #8	General Comments	I have reviewed the document on the testing of CYP2C19 variants and platelet reactivity for guiding antiplatelet treatment. It is a superb document. It represents the most comprehensive evaluation of the subject with the best scientific data up to the present. The document comes at an important time when the amount of data is increased greatly but the clinical questions of what and how to do things remain. Many institutions and physicians are now trying to put in place approaches to this very question that has been raised in this document.	Thank you for your kind comments. We appreciate the involvement of the ACC in the development of our protocol and the review of the completed report. We hope that the updated report will be useful to practicing clinicians and may inform clinical practice guidelines that are now in development.
Peer Reviewer #8	General Comments	There are several very important features of the document: 1) The methodology used for assessment is very robust and I agree with the approach. 2) The key questions - 1, 2, 3, and 4 - as they are set up really address all of the crucial issues for the analysis of the genetic testing as well as platelet function and its relevance and the potential for harm. 3) I agree fully with the conclusions about the strength of the data, the diverse trial designs, and the issue of heterogeneous populations. Clearly the conclusion that "the strength of evidence regarding the use of genetic or platelet reactivity testing to guide antiplatelet therapy selection is insufficient" is well founded on the data that has been analyzed and assessed with this.	Thank you for your comments. No further response necessary.
Peer Reviewer #8	General Comments	Thank you for the chance to review this very important document. We clearly need more data in terms of well-designed clinical trials.	Thank you for your comments. No further response necessary.
Peer Reviewer #9	General Comments	Quality of the Report: Superior Number of Hours Spent to Review the Report: 6	Thank you for your comments. Please see below for our point-by-point response.
Peer Reviewer #9	General Comments	The report is very timely and is of utmost clinical importance. The target population, audience and key questions are all clearly stated.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	General Comments	I am concerned that only the positive associations were reported in the Abstract (page 5). This detracts from the lack of robustness certainty surrounding these estimates given the presence of outcome reporting bias and small study bias. I would reword this so that it also reports the lack of associations with CV outcomes and more cautiously presents the positive findings, especially in light of the potential for bias and lack of evidence on effect modification (the "acid-test" of a pharmacogenetic test).	We have revised the abstract to provide additional information, as suggested.
Peer Reviewer #10	General Comments	Quality of the Report: Superior Number of Hours Spent to Review the Report: 2	Thank you for your comments. Please see below for a point –by-point response.
Peer Reviewer #10	General Comments	Yes. This appears to be a very rigorous review of platelet function testing and CYP2C19 genetic testing for assessment of efficacy of clopidogrel.	Thank you. No further response necessary.
Peer Reviewer #10	General Comments	There are two points that may be worth further mention. 1. Published work that post-dates the search performed for this manuscript supports the notion that the clinical utility of CYP2C19 genetic testing on predicting clinical response to clopidogrel may be indication specific. For those PCI patients who receive a coronary stent, CYP2C19 genotype may be a more important marker of poor on-treatment outcomes than when clopidogrel is given for other indications. Although there was an effort made to stratify by indication, e.g., ACS, stroke, afib, it would be of interest to stratify CAD/PCI patients by those who received a stent and those who have not received a stent. Although rare stent thrombosis was examined, the more common MACE endopint should also be examined in stented vs non-stented PCI patients. This information can be gleaned from some (but unfortunately) not all published studies. Indeed, some of the heterogeneity/bias observed may be due to the fact that the large studies performed earlier had fewer stented patients while the smaller more recent studies had a higher proportion of stented patients. See also Holmes et al JAMA 2011 and responses that followed (Shuldiner, et al JAMA 2012; Mega et al. JAMA 2012; Siasos et al JAMA 2012); and commentary by Johnson et al Clin Pharm Ther 2012.	We have updated our search strategy through to July 2012. As the reviewer acknowledges, the suggested regression analyses would either be susceptible to ecological bias (because only averages across studies were reported) or be based on a limited subset of the available evidence. As such, we have refrained from these analyses. Thank you for pointing us to the meta-analysis by Holmes et al. We had summarized the key findings from that work in the Discussion section of the Draft report. We are also aware of the correspondence that followed the publication of that study.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #10	General Comments	2. In addition to the common CYP2C19 *2 loss of function variant, the CYP2C19*17 gain of function variant was reviewed. An important caveat to the findings is that these two variants are not independent of one another; they are in linkage disequilibrium. Thus the effect of *17 to improve clopidogrel response (or be associated with increased bleeding) may not be independent, but rather due to the fact that in a population of *17 carriers, there will be a relative paucity of *2 carriers and conversely in a population of individuals without the *17 allele, there will be a relative excess of *2 carriers. Most studies genotyped and analyzed one of these variants and not the other while other studies have measured both. In those latter studies, through regression or stratified analyses, it is less convincing that the *17 has an independent effect on clopidogrel response. A systematic examination of whether the *17 variant is an independent predictor of on-treatment clinical outcomes would be very useful.	Assessment of whether each variant has effects independent of the other is limited by the information reported in the published studies. We have noted this in the Discussion section of the report. Note also that – given the relatively low prevalence of each of these alleles and the linkage between them – joint statistical analysis is likely to require very large sample sizes.
Peer Reviewer #11	General Comments	Quality of the Report: Superior Number of Hours Spent to Review the Report: 14	Thank you for your comments. Please see below for a point-by point-response.
Peer Reviewer #11	General Comments	The report is very well done and is clinically meaningful. Target populations is well defined and key questions are well formulated and explicitly stated.	Thank you. No further response necessary.
Peer Reviewer #12	General Comments	Quality of the Report: Good Number of Hours Spent to Review the Report: approximately 8 hours each by two reviewers.	Thank you for your comments. Please see below for our point-by-point response.
Peer Reviewer #12	General Comments	The report should be clinically useful and meaningful with a clearly defined target population and audience. The key questions are appropriate and explicitly stated.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #12	General Comments	Although the original concept for the report was to cover a wide range of uses of clopidogrel in vascular disease (afib, PVD, and ischemic heart disease), by the nature of the available literature this review is limited to testing for clopidogrel reactivity primarily in ischemic cardiovascular disease patients. That is, since the literature does not include many studies of clopidogrel use in cerebrovascular and peripheral vascular disease much of the review concentrates on individuals with heart related conditions such as acute coronary syndromes, etc. This is particularly relevant to the studies of predictive ability of the tests of platelet activity to identify who is, or who is not going to have an adverse outcome, and therefore should be treated more or less aggressively with antiplatelet agents. It is possible that for the other conditions in which clopidogrel is used that the outcome prediction could be different, either because of the nature of the disease itself or the comorbidity associated with it. The review should make it clearer upfront that results are almost exclusively applicable to ischemic cardiovascular disease.	We agree with this point. We discuss it in the "Applicability" subsection of the Discussion.
Kevin J. Croce	General	As a practicing interventional cardiologist and translational research physician who focuses on investigating the mechanisms of atherothrombosis, I was very interested in the content of the AHRQ *Testing of CYP2C19 Variants and Platelet Reactivity for Guiding Antiplatelet Treatment* document. I certainly appreciate the careful work that went into this excellent summary and would like to make several comments that relate to the analysis and the author's conclusions.	Thank you for your comments. Please see below for a point-by-point response to your suggestions.
Kevin J. Croce	General	The document does not clearly acknowledge that there currently is a knowledge gap; that the use of platelet function testing or genotype to guide treatment choice has not been evaluated well and that there are several issues with the design of the GRAVITAS study which are related to the fact that only 48% of patients treated with high dose clopidgrel achieve a PRU value below 208 which is the PRU cutoff value that appears to best predict reduced risk of MACE (*Price M, et. al. Circulation, 2011).*	We have made several changes in the Discussion section of the manuscript that address this comment. Thank you for pointing us to the article by Price et al. It has already been included in our report.

Commentator & Affiliation	Section	Comment	Response
Kevin J. Croce	General	A test and treat to target platelet function testing strategy which has the greatest potential to decrease MACE has not been tested. The AHRQ document fails to acknowledge that there is significant scientific support for the premise that high on treatment platelet reactivity (HTPR) is a modifiable risk factor that can be addressed with intensification of antiplatelet therapy. Specifically: a. HTPR identified by LTA or Verify Now predicts increased risk of MACE (acknowledged in the AHRQ document). b. Patients with HTPR defined as a Verify Now PRU >230 have 50% more CV events compared to PRU<230 (Brar JACC 2011). c. Patients with HTPR can have their platelet reactivity reduced by intensifying therapy with more potent agents such as prasugrel or ticagrelor.	We reviewed several studies comparing testing versus no-testing strategies for tests measuring platelet reactivity. For a summary of their findings, please see the relevant sections under Key Questions 3 and 4. Thank you for pointing us to the individual patient data meta-analysis by Brar et al. Data from this work have been summarized in the Table of previously published meta-analyses (in the Discussion section of the report).
Kevin J. Croce	General	Although the AHRQ document comments on the "need to standardize PFT in order to assess the relative impact of testing strategies" the analysis pools several different types of platelet function testing that have markedly variable clinical performance (Breet N, JAMA 2010).	We respectfully disagree with this point. We did not pool (quantitatively) or qualitatively synthesize evidence across different assays. In fact, we note that our findings regarding analytic validity (Key Question 2a) and the analyses reported regarding prognostic effects (Key Question 2b) lead us to not combine different types of platelet function testing. Thank you for pointing us to the paper by Breet et al. It had already been included in our analyses.

Commentator & Affiliation	Section	Comment	Response
Kevin J. Croce	General	The analysis and summary document fail to consider several new published or presented studies that show favorable performance of platelet function testing in identifying patients at risk for MACE. Querying VerifyNow in pub med, 88 studies have been published since the AHRQ document cutoff date of August 24, 2011. These published or presented studies include: a. ADAPT DES (~8500 patients) b. GRAVITAS (~2800 patients) c. ARMYDA PROVE (~730 patients) d. CROSS VERIFY (~800 patients)	We have updated our search strategy through to June 2012. Please note that our protocol pre-specified that analyses not published in full (e.g., data available only in abstract form) would not be included. Of the suggested studies, those that have been published in full have been included in the report. Including data available only in abstract form has inherent risks such as numerical errors and misinterpretation. This can occur when trying to submit findings in a very short time frame or due to severe word count limitations. AHRQ has an updating process so that trials that are not available in full text for inclusion this time around will be available around the time of the update.
Kevin J. Croce	General	Several new studies are due to be presented 30-60 days after the AHRQ document is finalized. These studies are likely to shed important light on the clinical utility of genetic and/or platelet function testing in clinical practice and may dramatically alter the conclusions of the Testing of CYP2C19 Variants and Platelet Reactivity for Guiding Antiplatelet Treatment document. These studies include (but are not limited to): a. ADAPT-DES 1 year follow-up (~8500 patients) b. TRILOGY ACS platelet function substudy (~3000 patients) c. ARCTIC (~2500 patients)	Please see our reply to the preceding comment.
Kevin J. Croce	General	I strongly believe that the conclusions in the AHRQ document need to be modified to a. Carefully outline the ongoing uncertainty in this field given that study designs have not adequately addressed the efficacy of a test and treat to target strategy that employs LTA or Verify Now. b. Acknowledge that it is mechanistically plausible that intensification of antiplatelet therapy could benefit HTPR patients even though randomized clinical trial data is currently lacking.	We have made several changes to the Discussion section of the report. We believe that they address the reviewers' point (a), to the extent justified by the data and our methodological approach. The mechanisms underlying the use of phenotypic tests for platelet reactivity are described in detail in the reports' Background section. Please see the revised strength of evidence section for our interpretation of the available studies on the use of phenotypic testing for guiding antiplatelet treatment.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	General	We read with great interest the draft Comparative Effectiveness Review titled "Testing of CYP2C19 Variants and Platelet Reactivity for Guiding Antiplatelet Treatment". The report states that there is "evidence that high on-clopidogrel platelet reactivity is associated with an increased risk of adverse cardiovascular outcomes for at least some of the available assays" but concludes that the "strength of evidence regarding these prognostic effects is low because of concerns regarding selective outcome reporting and the relatively small number of studies reporting clinical outcomes". The rationale for this conclusion also appears to include concerns about the analytic validity of platelet function testing. We feel there is sufficient evidence to draw alternative conclusions, in particular, that the strength of evidence for prognostic utility of platelet function testing is in fact higher than the designation proposed in the draft report. Our comments supporting this conclusion are detailed below.	NOTE: These comments were submitted jointly by 4 public reviewers. We thank the reviewers for their constructive comments. We appreciate the input from investigators who have co-authored many of the primary publications included in the report. Please see below for a point-by-point response to the reviewers' specific suggestions.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	General	There is clear, established utility of the predictive ability of platelet function testing in high-risk patients. Although there is no "gold standard" for PFT, there are "practical standards" available that are well characterized methods that have been associated with outcomes through multiple independent studies. Measurements using a practical standard should be specific for P2Y12 inhibitor effect in a physiologically relevant sample matrix, approved for clinical use, and possible to integrate into the patient workflow.	Please see our other replies to similar commnets from these reviewers regarding "gold standard" tests and their role in assessments of analytic validity. Please also see our extensive replies regarding SOE and the general remarks at the beginning of this document. Key Question 2a of the revised reports offers details on issues related to the absence of a gold standard test and the revised Discussion section addresses issues related to SOE.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	General	Clopidogrel is now generic and will continue to be front-line therapy in the United States due to its overall effectiveness and new economic pressures to keep patients on clopidogrel. In addition, the generic clopidogrel available in the United States has been evaluated only through assessments of bioavailability and not in clinical outcome trials. However, within-individual differences in the pharmacodynamic effect of these "similar" medications have been reported (Jeong et al Korean J Intern Med 2010), so it is even more important to confirm that there is a measureable pharmacodynamic effect, especially if switching between generics or between branded and generic clopidogrel.	Thank you for this information. No further response needed.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	General	Above all, it is important to remember that the intent of P2Y12 inhibitors is to reduce reactivity to ADP. If there's no evidence of a pharmacodynamic effect of these life-saving medications, is it reasonable to expect a benefit? There has been a call for a large randomized trial to evaluate the utility of platelet function testing to guide therapeutic decisions. However, such a trial would require randomizing patients with high platelet reactivity to continue taking clopidogrel. Based on the consistent associations that have been reported between high platelet reactivity and increased risk for thrombosis, it may no longer be possible to conduct such a trial due to the reluctance of physicians to have their patient with high platelet reactivity risk continuing to take clopidogrel. This was evident in the TRIGGER-PCI trial, which saw approximately 30% of patients drop out prior to randomization after it was know that they had high platelet reactivity (Trenk et al, J Am Coll Cardiol 2012).	We have adopted some of this thinking in the Discussion of Future Research Needs in the Final Report. However, the results and strength of evidence cannot be altered just because something is logistically difficult to do. The strength of evidence is determined based on the confidence that the results seen in the CER will not change with future research.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	General	Finally, research in the area of genotyping and platelet function testing is continuing to evolve. The ADAPT-DES study is a large, 8500 patient multicenter registry that has reported a significant association between high platelet reactivity and 30-day stent thrombosis at a major international conference. However, the study results have not been published and were not included in the analysis described in the draft report. In addition, the scientific community is anticipating the results of the ARCTIC trial, which is a large, 2500 patient trial evaluating the effectiveness of treatment guided by platelet function testing. We recognize that a cutoff date for including studies in the analysis described in the draft report is necessary. Considering the amount of significant data that either has been reported since the cutoff or is expected to be reported in the next 12-18 months, we urge the authors to update their analysis once the new data are available.	We have updated the report to cover published evidence though to July 2012. We agree that this is a rapidly evolving topic. AHRQ has a process for updating CERs and trials not being available for inclusion in this report will have the opportunity to be included in a subsequent update.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	General	Thank you for your consideration of these comments in your preparation of the final report.	Thank you for your extensive and very constructive comments. Indeed, we found them helpful while preparing the Final Report.

Commentator & Affiliation	Section	Comment	Response
Daniel L. Simon	General	I read, with great interest, the Comparative Effectiveness Review entitled *Testing of CYP2C19 Variants and Platelet Reactivity for Guiding Antiplatelet Treatment*. As someone who actively follows and performs research within the individualized antiplatelet therapy space, as well as incorporating some of the lessons in my clinical practice, the choice of AHRQ to perform this analysis is very timely. I also commend the writing committee for producing a very thorough review in a rapidly evolving area that has a constant stream of new data and publications, from which we can glean insight to improve care for our cardiovascular patients. With that said, I have a few comments and concerns regarding the draft report.	Thank you. Please see below for a point-by-point response to your comments.
Daniel L. Simon	General	The purpose of platelet function testing is to measure the effect of antiplatelet therapy. Antiplatelet drugs, such as clopidogrel, provide benefit by reducing the risk of thrombosis by reducing platelet reactivity to ADP. High platelet reactivity has been conclusively shown to be associated with increased rate of major adverse cardiovascular events (MACE) in several independent studies, meta-analyses, and this was even confirmed in the draft AHRQ comparative effectiveness review.	We agree that the evidence reviewed in our report supports an association between high on-clopidogrel reactivity and adverse cardiovascular outcomes. Our assessment of the strength of evidence is provided in the updated report. Please note that we use a formal approach to judging how conclusive the available evidence is. Details about our approach are provided in the Methods section of the report and citations therein.
Daniel L. Simon	General	After adjustment for clinical and procedural risk factors, platelet reactivity is an independent risk factor, as shown in the GRAVITAS trial published in Circulation (2011) by Price et al, the POPular trial published in JAMA (2010) by Breet et al, the patient level meta-analysis published in JACC (2011) by Brar et al, and in the ADAPT-DES study presented at TCT 2011 (Stone, GW). Therefore, there is significant value in using platelet function testing to confirm that clopidogrel is having the expected effect. In particular, it is even more important to measure platelet reactivity in patients with higher clinical risk (e.g., patients with heart failure, obesity, diabetes, acute coronary syndrome presentation, and renal disease) because the MACE rates are higher in these individuals. Because VASP is not widely available and is not approved in the United States, the choices should likely be limited to light transmittance aggregometry and the VerifyNow test.	Thank you for pointing us to these studies that were only available as conference abstracts. The GRAVITAS and POPular studies were reviewed in the report. The individual patient data meta-analysis by Brar et al. did not meet our criteria (we did not review meta-analyses to answer our Key Questions), however it was identified by our searches and key findings from it were presented in the Discussion section of the Draft report (the study was summarized in a table and discussed in the text). Please note that, by mandate, Evidence-based Practice Center reports do not make clinical practice recommendations. We review, summarize and assess the availble evidence, determine the strength of evidence, and the applicability of evidence.

Commentator & Affiliation	Section	Comment	Response
Daniel L. Simon	General	Evaluations and considerations of the benefit of platelet function testing should not continue to be performed in a vacuum separate from everyday, real-world decision making. The clinician is practicing under 2 FDA black box warnings with respect to clopidogrel drug-frug interactions and poor metabolizers. In addition to its value as a predictive tool, platelet function testing can also provide useful information with respect to the conomics/affordability of drug treatment. Now that clopidogrel is available in the United States as a generic and is a Tier 1 medication, insurance companies are starting to require prior authorization, including demonstration of high platelet reactivity on clopidogrel (and therefore increased risk of MACE and stent thrombosis) before authorizing reimbursement for more potent and more expensive P2Y12 inhibitors. Platelet function testing is the only way to confirm that the antiplatelet medication is having the intended effect.	We realize the need for decisionmaking under uncertainty and that factors other than strength and applicability of evidence sometimes come into play. Please note that we do not in any case make clinical practice recommendations regarding testing. That is outside the scope of the Effective Healthcare Program and the work of the Evidence-based Practice Centers. However, we hope that our comprehensive assessment of the evidence can inform decisionmaking.
Joseph Brent Muhlestein	General	By the way, thank you for your careful review of this important subject.	Thank you. No further response necessary.
Sandra Zelman Lewis (on behalf of The American College of Chest Physicians)	General	The American College of Chest Physicians appreciates the opportunity to provide comments on this draft report on "Testing of CYP2C19 Variants and Platelet Reactivity for Guiding Antiplatelet Treatment." This document was reviewed by 4 members of the panel that developed the Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines [Chest 2012;141(2) (suppl) - http://chestjournal.chestpubs.org/content/141/2_suppl]. Their comments are collated below.	We appreciate the College's involvement in this review (including earlier stages) and thank the 4 members for taking the time to submit these helpful comments.
Sandra Zelman Lewis (on behalf of The American College of Chest Physicians)	General	Overall, the reviewers commented that the report was both thorough and well done and they primarily agreed with the conclusions, which seem to be wellbalanced and mostly supported by the evidence.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Sandra Zelman Lewis (on behalf of The American College of Chest Physicians)	General	The questions (and sub-questions) guiding the review were appropriate, a thorough review of the literature was conducted, and the strength of the evidence was assessed. It is of note that for every question/sub-question the quality of evidence was either NA, insufficient, or at least low. Arguably the review of studies evaluating use of genotyping or phenotypic testing to guide platelet therapy is the most clinically relevant. All other evidence pertains to prognostic ability of genotypic or phenotypic testing or to surrogate outcomes. In this section, it is stated: "studies generally indicated that patients with test-based monitoring had better outcomes... but the differences were often not statistically significant." Even this is an overstatement. Only one of the studies showed a significant effect. All the others had very wide confidence intervals suggesting no conclusion could be drawn. At present the abstract and conclusion suggests that the strength of the evidence regarding genotypic and phenotypic testing as a guide to platelet therapy is "insufficient." This is correct but a stronger statement stating that such testing should not be routinely used would be appropriate (given limited health care resources, poor test-to-test reproducibility, potential for harm with increasing anti-platelet therapy strength based on this testing, etc.).	We agree with the interpretation of the evidence (with regards to evidentiary strength). However, by mandate EPCs do not make clinical practice recommendations; as such we have refrained from making any recommendation in favor or against testing. We review, summarize and assess the available evidence, determine the strength of evidence, and the applicability of evidence. We also agree with the suggestion that the effect of CYP2C19 is mediated through its effect on platelet reactivity. We have cited papers relevant to the proportion of "variability explained" by CYP2C19 genotyping. Nonetheless, please note that this was not one of the outcomes of the review. As such we have not systematically searched for relevant evidence in health or diseased individuals. Finally, thank you for pointing us to the RAPID GENE trial. As you know, this was published after the final search date covered by the Draft report. Our updated search identified this study and it was included in our analyses. Our evaluation of the evidence from this trial is generally in line with that of the reviewers. Please see the full text of the revised report for additional details.

Commentator & Affiliation	Section	Comment	Response
		Indeed, poor response to clopidogrel has been consistently associated with recurrence of cardiovascular events, at least in acute settings, and platelet reactivity is therefore considered as a surrogate marker of cardiovascular risk in patients on clopidogrel. It should be clearly mentioned that the CYP2C19 genotype is in fact a surrogate marker of the latter surrogate marker (onclopidogrel platelet reactivity) that is thus an even more indirect marker regarding the evaluation of the risk related to clopidogrel response. The contribution of the CYP2C19*2 allele in the variability of clopidogrel responsiveness in healthy individuals is mentioned on page 2 and the work of Shuldiner et al is quoted (ref#25). In this latter study (GWAS study), this polymorphism accounted for about 12% of the total variability. It should be stressed that in cardiovascular patients, the variability explained by the CYP2C19*2 is even smaller (around 5%) (PMID: 21628721, PMID:20510210, PMID:21692977). It is therefore not surprising that there is a large variation of clopidogrel response in both carriers and noncarriers of the CYP2C19*2 allele; between 20 and 40% of CYP2C19*2 of non-carriers still display high platelet reactivity (PMID:22615340, PMID:22088980). An important study, the RAPIDGENE study (PMID:22464343) addressed the issue of tailoring antiplatelet drug treatment according to CYP2C19*2 genotype in 187 patients. The authors found that after a 7-day course of treatment (prasugrel for carriers of the CYP2C19*2 allele and clopidogrel for the CYP2C19*2 non carriers), none of the patients allocated to the genotype-guided strategy and carrier of the 2C19*2 allele had high platelet reactivity compared with 30% in the standard treatment group. The company manufacturing the SPARTAN device is pushing for the generalization of these findings to the entire population, which does not seem appropriate.	[This is a continuation of the comment in the preceding row. The text was split in 2 table rows for technical reasons.]

Commentator & Affiliation	Section	Comment	Response
Sandra Zelman Lewis (on behalf of The American College of Chest Physicians)	General	Surprisingly, there were only few non-CYP2C19*2 carriers with high platelet reactivity (9.6%, 95%CI[5.8-14.8]) in that study, which is at odds with other studies (PMID:22615340, PMID:22088980). If all patients are to be tested for response to clopidogrel (or other platelet inhibitor), it seems more prudent to do point of care testing of platelet function (reactivity), than the presence of a genetic mutation, which may have either a significant effect or minimal effect on platelet function. In so doing, we can identify the specific patients in whom platelet inhibition is achieving its goal.	We agree with these suggestions in principle, however, we can only evaluate the evidence that is available through published studies.
Sandra Zelman Lewis (on behalf of The American College of Chest Physicians)	General	It is hoped that these comments will be useful for AHRQ and the EPC conducting this review. Again, the ACCP appreciates the opportunity to respond.	Again, thank you for your comments. No further response necessary.
Peer Reviewer #6	Abstract	This comparative effectiveness review evaluated the analytic validity, predictive value, and comparative effectiveness of two types of medical tests (genetic testing for CYP2C19 variants and phenotypic testing to measure platelet reactivity) to identify patients who are most likely to benefit from clopidogrel-based antiplatelet therapy and to guide antiplatelet therapy in patient populations who are eligible for clopidogrel treatment. - The main focus of this review is the usefulness genotyping and platelet function testing in patients who are already on clopidogrel and who may need alternative strategy/agents. Therefore, we believe that the authors should update this sentence, particularly “to identify patients who are most likely to benefit from clopidogrel-based antiplatelet therapy and to guide antiplatelet therapy in patient populations who are eligible for clopidogrel treatment”	We have revised the sentence to read “to identify patients who are most likely to benefit from clopidogrel-based antiplatelet therapy and to guide antiplatelet therapy in patient populations who are eligible for (or are already receiving) clopidogrel treatment”.
Peer Reviewer #6	Abstract	The majority of studies were conducted in populations with ischemic heart disease. Patients with high platelet reactivity at baseline were more likely to be clopidogrel nonresponders during follow up. There is limited evidence to support this observation and it should be noted in the manuscript.	This has been noted in the Results section of the Final Report.
Peer Reviewer #6	Abstract	The ability to predict clinical outcomes was reported for various assays; the most commonly assessed were light-transmission aggregometry (46 studies); VerifyNow P2Y12 (28 studies); the vasodilator-stimulated phosphoprotein (VASP) assay (14 studies); Multiplate analyzer (12 studies); and Platelet Function Analyzer-100 (8 studies).	No response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Abstract	It should be noted that each assay is based on different principles- LTA= platelet-platelet binding via fibrinogen in PRP, VerifyNow= platelet binding to fibrinogen coated beads in whole blood, VASP= analysis of phosphorylated state of VASP protein that is associated with downstream signaling of P2Y12 receptor, Multiplatelet analyzer= binding of activated pplatelets to electrode and PFA= platelet plug formation in the presence of an agonist and shear. None of these assays are surrogate for in vivo clinical event occurrence- ie thrombosis. These functional assays may be useful in evaluating drug response, but in vivo thrombotic event is influenced by multiple factors. Therefore these assays have low positive predictive value and high negative predictive values. The authors should address this in the manuscript. Moreover, the authors should also highlight the studies that used receiver operator characteristic curve analysis to demonstrate the relation of platelet function to clinical outcome. This is superior to studies that used upper quartile measurements.	We have emphasized the differences in the principles of measurement across assays. We have also provided citations to numerous reviews that provide detailed information on the analytical aspects of the use of each of the assays in common use. We have already extracted information on the methods for determining cut-offs for reactivity in each individual study. Although arguably better than arbitrary dichotomization, ROC analysis – as applied in the reviewed studies – has its own limitations. We mention these briefly in the Discussion section of the Final Report.
Peer Reviewer #1	Introduction	I do not have any major comment. As a suggestion, have the authors considered a strategy based on standard therapy directly with subsequent adjustment based on genetic results? This would alleviate the concern about treatment being delayed.	We have mentioned that several possibilities exist for incorporating testing in clinical practice, in the Background, Methods, and Discussion section of the report. In the Results section we have reviewed all strategies that had been applied in at least one publication.
Peer Reviewer #6	Introduction/ Background	ES-1 With some patients showing no platelet response to clopidogrel administration (“nonresponsiveness” or “resistance”). This should be changed to “no or minimal platelet response”	Thank you. We have adopted the suggested wording.
Peer Reviewer #6	Introduction/ Background	Alternatives to standard clopidogrel treatment include higher-dose clopidogrel regimens and the use of other antiplatelet agents, such as prasugrel or ticagrelor, which are not metabolized through the same pathways as clopidogrel. This is wrong- prasugrel is also metabolized by same CYP Isoenzymes and ticagrelor is also metabolized by CYP3A4. The important point here is prasugrel is effectively metabolized to an active metabolite and on molar basis has same PD effect as clopidogrel. Ticagrelor is a direct drug although the active metabolite is as potent as native clopidogrel.	Thank you. We have revised this sentence in the Executive Summary and the main text of the Final Report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Introduction/ Background	Prasugrel and ticagrelor have efficacy similar or superior to clopidogrel for preventing major adverse cardiovascular events (MACE). However, these drugs may increase the risk of bleeding complications. Clinical efficacy of prasugrel and ticagrelor is superior to clopidogrel in various subgroups of ACS not similar to clopidogrel (TRITON and PLATO studies).	Thank you. We have revised this sentence in the Executive Summary and the main text of the Final Report.
Peer Reviewer #6	Introduction/ Background	ES-1 The question of identifying the optimal antiplatelet therapy may also carry substantial cost implications because generic clopidogrel products will soon be available. _ Generic clopidogrel is already in the market. Please update.	Thank you. We have revised this sentence in the Executive Summary and the main text of the Final Report.
Peer Reviewer #6	Introduction/ Background	ES-2. "The functional status of some combinations of genotypes (usually combination of LoF and GoF) is currently unknown…" Moreover, the PD effect in patients with one LoF allele is also unknown and found to be variable. It is only in poor metabolizers with two LOF alleles, the PD effect of clopidogrel is poor.	To address this concern we have conducted analyses using alternative genetic models. Please see the Methods and Results section of the Final Report.
Peer Reviewer #6	Introduction/ Background	ES-3-(2) direct testing of the patient's blood while the patient is taking clopidogrel to see whether the platelets actually have become less prone to aggregate in response to specific agonists (phenotypic testing for platelet reactivity). It is not specific agonist- it is just ADP. P2Y12 is otherwise known as ADP receptor and clopidogrel effect is only confined to the inhibition of this receptor, nothing else.	We prefer our original wording since it can account for agonist combinations, used in some of the devices that we reviewed.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Introduction/ Background	Regarding treatment decision making, we conceptualized the analytic framework as a decision problem, wherein patients' disease can be managed with one of the following approaches (depicted from top to bottom in the flow diagram): 1. Undergo genetic testing and then base the treatment decision on the test results. 2. Undergo genetic testing and then base the treatment decision on the test results. After receiving therapy for an adequate period of time, undergo phenotypic testing for platelet reactivity and use the results to decide whether the treatment strategy should be modified. 3. Receive standard treatment directly and, after an appropriate amount of time, undergo phenotypic testing for platelet reactivity and use the test results to decide whether the treatment strategy should be modified. Use of phenotypic testing (but not genetic testing) as a monitoring test can be considered a variation of this strategy in which the test is repeatedly performed. 4. Receive antiplatelet therapy without undergoing any testing (the current standard of care). The authors should consider the following option which is more practical 1. In patients who are already on clopidogrel- platelet function testing should be the first choice and genotyping may be complementary. 2. In high patient ACS who are not on clopidogrel, but need to undergo primary PCI, genotyping can be performed to rule of the chances of poor metabolizer status (only 2% in Caucasian population).	The strategies listed originally where those that we anticipated would be the most prevalent based on preliminary reviews of the literature. To account for the additional strategies identified by the comprehensive review (now presented in the Final Report) we have added the following wording: *"The above strategies were identified as the most prevalent in published studies by preliminary searches conducted in preparation of this review. Given the clinical complexity of managing patients who are candidates for antiplatelet therapy, we anticipated that variations of these strategies may be uncovered by the full evidence review."*
Peer Reviewer #7	Introduction	ES-2, lines 5-6 - the sentence is quite confusingly written. It makes it sound like CYP2C19 binds to the P2Y12 receptor	The sentence has been rephrased.
Peer Reviewer #9	Introduction	Well written and concise.	Thank you. No further response necessary.
Peer Reviewer #10	Introduction	Reads well. Well done.	Thank you. No further response necessary.
Peer Reviewer #11	Introduction	The introduction is informative and well-written. The authors may want to consider adding information (if available) on how many are tested by genotyping/phenotyping methods each year. In addition, direct-to-consumer genotyping is an issue that may be worth mentioning for background/contextual purposes.	Thank you. Regarding the extent of test use in practice, we could not find data from adequately representative populations; thus, we did not add this information in the Final Report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	I realize this is only the executive summary, but I wasn't entirely clear how the 226 articles were chosen out of the 842 reviewed in full. Perhaps including a flowchart would be useful. Also, a Table outlining specific criteria used for assessing bias and applicability. I am aware this is available in the full report, but having quick summaries might be helpful to the ES and perhaps to the full report as well.	Thank you for realizing that avoided providing this information in the Executive Summary to conserve space. A flowchart with all requested information is available in the Results section of the manuscript. The Appendix presents detailed reasons for exclusion for each excluded study that was reviewed in full text.
Peer Reviewers 2, 3, 4	Methods	All analyses combine carriers of either one or two alleles. The review does not address the issue of poor metabolism, that is, carriage of two loss-of-function alleles. This population is the most susceptible to adverse cardiovascular outcomes based on pharmacokinetic data which is the basis for the US-approved clopidogrel labeling recommendation to consider other therapies in patients with two LOF The assumption of dominance in this scenario is not correct.	We agree that the genetic model for CYP2C19 effects is important. We have explicitly stated the reasons for using a dominant genetic model in the Draft report. We note that all previously published systematic reviews have used the same genetic model in their analyses (e.g., Bauer BMJ 2011; Holmes JAMA 2011; Zabalza Heart 2012). Pharmacokinetic studies suggest that the appropriate model is probably dominant or additive (Scholz et al. Br J Clin Pharmacol 2009; Yoo et al. Br J Clin Pharmacol 2010; Yasui-Furukori et al. Br J Clin Pharmacol 2004). Our selection of a dominant model follows similar arguments as previous analyses (e.g., Bauer et al. BMJ 2011) and is driven by the availability of data in the primary studies. We have attempted to perform sensitivity analyses based on *additive* and *recessive* genetic models in the Final Report. We note that in a large number of studies data were not extractable for comparisons under these genetic models, raising additional concerns about selective reporting bias.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	Methods	In their review of the studies, the authors should address other weaknesses of these data, including knowledge of whether samples were collected at baseline (which for a cohort would introduce null bias and decrease the effect size), or whether the substudies are representative of the overall trial populations (i.e., bias in sampling).	Information on the timing of testing and information related to the sampling methods has been extracted and was provided in the Draft Report (please see Appendix Tables on study design and test-related information corresponding to each Key Question). In order to conserve space, we could not include this level of detail in the executive summary or the main body of the report. We understand that given its length, you may not have seen it. It is not clear why collection of samples at baseline in a cohort study would introduce "null bias". If anything, exposure ascertainment using samples obtained at baseline is the preferred approach. Further, in genetic association studies ascertainment of genotypes after some followup time has elapsed may introduce survival bias (largely of unpredictable magnitude or direction).
Peer Reviewers 2, 3, 4	Methods	Active metabolite pharmacokinetics should be included as an intermediate outcome like platelet function (e.g., Varenhorst, et al. EHJ 2009 would meet inclusion criteria).	Outcomes for this review were determined after extensive discussion with Key Informants and Technical Experts with expertise in genetics, genetic epidemiology, internal medicine and cardiology (including frontline clinicians), clinical trial design, and health technology evaluation. In each of these discussions the EPC went over the list of outcomes listed in the draft protocol to ensure that no outcomes of importance were ignored. Active clopidogrel metabolite concentration was not proposed for consideration in any of the calls. In fact, several Key Informants and Technical Experts suggested that the report should only consider clinical outcomes (i.e., to not review platelet reactivity as an intermediate outcome).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	Methods	To assess clinical utility, metrics such as sensitivity, specificity, positive and negative predictive values, likelihood ratios, and numbers needed to genotype should be included. Additionally the question of utility should be framed against the magnitude of the public health issue, including the size of the affected population (number of stents deployed each year), the number of stent thromboses and deaths that would be averted by a genotyping strategy vs. standard of care, or the incurred cost vs. benefit, for example.	Time-to-event metrics are more appropriate for the study designs reviewed (all but two studies of CYP2C19 variants had a longitudinal design). No study provided information on prognostic performance accounting for the time-to-event nature of the data or for censoring (such analyses – e.g. modified ROC analyses – are generally rarely reported). To accommodate the reviewers' comment, in the Final Report, we performed analyses of predictive sensitivity and specificity for the two outcomes with the largest number of studies (MACE and stent thrombosis). Results from these analyses have been provided in the revised Final Report. With regards to the "number-needed-to-genotype" (NNG), no studies included in the draft report were comparative studies of genotyping versus no genotyping. Thus, the NNG cannot be calculated unless one uses a decision model or other extrapolation method. The single directly comparative study of genotyping versus no genotyping was published after our last search (Roberts et al. Lancet 2012) and is currently being included in the report's update. This study reported no clinical events in both arms, again making the NNG uninformative. We generally avoid calculating NNG in a meta-analysis setting (across studies) because of their reliance on absolute risk calculations (Engels et al. Stat Med 2000), suboptimal statistical properties (as suggested in the Cochrane Handbook of Systematic Reviews), and potential to mislead in the presence of heterogeneity (Ebrahim, Eval Health Prof. 2001). The Introduction section of the report provides information on the magnitude of the public health issues. We provide information on metrics of disease burden, provided that they are available from representative samples of the US population.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	Methods	Despite imprecision, the lower bound of the relative risk for stent thrombosis, an event with high-case fatality, was 15%. The authors do not specify what a clinically relevant effect of CYP2C19 genotype would be.	There is ongoing discussion of whether such minimum clinically important effects should be considered in EPC reports and how to obtain them. We opted to leave this judgment to readers. We note, however, that selective outcome reporting, publication bias, other biases, or chance could very well account for a 15% increase in risk.
Peer Reviewers 2, 3, 4	Methods	The manner of presenting results for GOF variants may be misleading because the comparison is actually against patients without any LOF alleles and patients with one or two LOF alleles (essentially the inverse comparison of LOF alleles). Consider eliminating this analysis as the *17 variant does not reproducibly correspond to higher active metabolite concentrations.	We realize that this analysis also assumes a dominant genetic model. However, it is the analysis that can be performed across the largest possible number of available studies. We note that all other meta-analyses that have assessed this association have assumed the same genetic model: for example Holmes et al. JAMA 2011; Bauer et al. BMJ 2011; Zabalza et al. Heart 2011. The number of studies reporting data that allowed the evaluation of alternative genetic models was very limited for *17 variants, as such statistical analyses could not be undertaken.
Peer Reviewers 2, 3, 4	Methods	Undue emphasis is given to “potential harms of testing” based on GINA protections and, more importantly, the “higher-risk treatments” also have greater efficacy irrespective of genotype (including a survival benefit).	They Key Questions of the report, including those on the potential harms of testing were determined a priori, following an extensive process involving Key Informants and Technical Experts, following standard EPC procedures. We cannot make post hoc changes to the Key Questions at this point. Further, our report highlights the lack of evidence on potential harms from test-directed treatment and will hopefully motivate further research on this topic, both for genetic and phenotypic tests.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	The authors should refer to HuGE guidelines on the conduct of systematic reviews in genetic epidemiology and ensure that they have covered all relevant aspects of examining the validity of a pharmacogenetic interaction.	Thank you for this suggestion. We believe that our review is consistent with current recommendations by the Human Genome Epidemiology Network. We have consulted the methods guide of the HuGe Network. In developing our protocol we also relied heavily on the CDC's EGAPP guidelines (available here: www.egappreviews.org/workingrp/methods.htm). Further, a CDC EGAPP representative was involved in early steps of this review (including the specification of the research questions and the development of the research protocol and methods). An additional CDC representative and another EGAPP member provided comments on the draft research protocol that was posted online for public comment. We incorporated all suggestions in the final review protocol that guided our review. Finally, please note that EPC Reports are subject to additional reporting standards as required by AHRQ policy or suggested in the EPC Methods Guide.
Peer Reviewer #7	Methods	No comments	No response necessary.
Peer Reviewer #9	Methods	Inclusion and exclusion criteria are appropriate and justifiable. Search strategy is logical.	Thank you. Please note that, although our original search was quite comprehensive, we have expanded it further in preparation of the Final Report.
Peer Reviewer #9	Methods	I have some concerns about the statistical methods as outlined below: I am concerned with the use of random effects modeling to present the major findings from meta analysis as this gives more weight to smaller studies (and this is particularly important in this review, as there is strong evidence for small study bias for many of the outcomes reported). Also, there is a misconception that random effects meta analysis is by default appropriate when the I2 is high – this is not true, and in fact it depends on whether one considers there to be "one true effect" or whether there is an "average treatment effect". For small study bias, the Harbord test for small study effects is considered to be superior to Eggers test (www.ncbi.nlm.nih.gov/pubmed/16345038) as it has fewer false positives with retained power.	We used fixed effects analyses as sensitivity analyses. We agree that the choice between random and fixed effects analyses is determined by the analyst's belief about the underlying model – not data driven observations on the extent of heterogeneity. We are aware of the somewhat superior properties of the Harbord test (compared to the Egger test). However, the former is not applicable to cases where studies report adjusted estimates of effect. Several large studies included in the review reported information exclusively in this form.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	Methods	I could not see where the authors described dealing with "0" values in meta-analysis. Looking at the meta plots, presumably the authors used the continuity correction function in Stata, which assumes a value of 0.5. However this can introduce bias, and it is preferred to exclude studies that have zero counts from meta-analysis.	We have provided this information in the Methods section of the Final Report.
Peer Reviewer #9	Methods	Finally, how come there are no funnel plots?	We think that the interpretation of funnel plots is not reliable when the number of studies is small-to-moderate and practically impossible in the presence of between-study heterogeneity. We have expanded on our rational for not using funnel plots in the Methods section of the Final report.
Peer Reviewer #11	Methods	Inclusion and exclusion criteria are justifiable. Search strategies are logical and explicitly stated, with exact search terms included in an appendix. Definitions, outcome measures and statistical methods are appropriate. Very high marks for rigor, systematic approach, transparency and reproducibility of methods described in this review.	Thank you. No further response necessary.
Peer Reviewer #12	Methods	The inclusion and exclusion criteria are justifiable and the search strategies are explicitly stated and logical. The outcome measures used for this topic are quite complicated with several of them relying on epidemiological and statistical techniques to detect variable interaction, but generally the ways of assessing the value of testing for platelet reactivity are appropriate and reflect the available literature.	Thank you. No further response necessary.
Peer Reviewer #12	Methods	Two specific areas could be clarified. First, the discussion of harms from clopidogrel use (primarily bleeding outcomes) should be framed more explicitly in terms of the balance of harms and benefits (bleeding events versus reduced MACE events). Clinically, the question of interest is whether clopidogrel should be used based on a test result; in a way that creates a more favorable balance of cardiovascular risk reduction versus increased bleeding risk. Although the studies available may not address and therefore may not permit such a risk-benefit calculation, this question should be recognized as the principle issue (to patients and clinicians) and not directly informed by this review. It should be identified as a priority for future research.	We have explicitly referred to the need to consider benefits and harms.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #12	Methods	Regarding the use of statistical methods for meta-analysis, there seems to have been a considerable amount of subjective judgment exercised about whether or not it was appropriate to do a formal meta-analysis on a given set of studies. Perhaps more times a meta-analysis with the presentation of summary statistics should have been done in this report (as discussed further below). However, in other cases provision of more detail on the reason and justification for proceeding one way or another in the body of the report and in the executive summary may be all that is needed. For example, in the section on LTA use in ischemic heart disease, no quantitative meta-analysis is done in spite of a relatively large number of studies (ES-15). The main text (page 117) suggests that the reason for not doing so was the use of different positivity thresholds in different studies. If this is indeed the issue, this should be made clearer and there should be a brief discussion about why it was not possible to synthesize these data using a summary ROC curve or similar meta-analytic technique.	We agree that there is unavoidable subjectivity in deciding whether studies should be pooled or not. This is because similarity (exchangeability) is not quantitatively assessed and because the decision on whether statistical pooling is useful is also based on judgment. To address the reviewers' concern, we have provided additional details on our rationale for each decision. Note that when we refer to differences in the criteria for positivity we do not refer to the use of different cut-off values on the same metric (in which case sROC analysis would indeed be useful). Instead, we refer to differences in the actual measurement (e.g. maximal reactivity vs. early reactivity or late reactivity; or absolute reactivity vs. change in reactivity from baseline). We believe that such differences cannot be accounted for by simple sROC analysis. This point has been clarified in the text of the Final Report.
Peer Reviewer #12	Methods	In addition, while the issue of study heterogeneity and how to consider it in making decision about the appropriateness of meta-analysis is addressed in the main body of the report (page 77), we recommend clarifying the way statistical tests for study heterogeneity are used in the executive summary. We agree that number of studies and sample size of each can affect the performance of heterogeneity statistics (and thereby influence whether it is appropriate to quantitatively summarize the studies) and would add that Cochrane's Q statistic is subject to excessive power when many studies are included (BMJ. 2003 Sep 6;327(7414):557-60),while the I2 statistic is not. The executive summary should be clearer about how these statistics were used to guide the decision about whether to conduct a meta-analysis.	Please note that we did not base decisions to perform meta-analysis on tests for heterogeneity. It was based on the clinical and methodological heterogeneity between trials and studies we found when reviewing the literature available to answer each Key Question. This has been emphasized in the Final Report.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	**Variability among methods for measuring platelet reactivity is expected.** In the analysis of analytic validity for phenotypic measurements, the authors report that the level of correlation between methods for measuring platelet reactivity is low to moderate. These observations were due, in part, to the inclusion of studies reporting the correlation of one method against another method that was treated as a "gold standard". The authors correctly report that in these analyses, the "gold standard" for comparison was considered to have no measurement error, which is known to be an incorrect assumption. In fact, there is no "gold standard" for platelet function testing because these assays rely on biological platelet function that is not possible to create in vitro or prepare as a reference standard. Further, the assays evaluated in the draft review, while similar, have fundamental differences in their fundamental scientific principle.	We agree with the reviewers' interpretation of the evidence on analytic validity. Please note that this is reflected in our decision not to combine evidence from studies assessing platelet reactivity using assays based on different principles of measurement. Note that this is a distinctly different approach from most existing meta-analyses of these tests which have combined disparate analysis methods. We also agree that currently there is no test that can be considered as having no measurement error (i.e. there is no gold standard test).
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	Light Transmittance Aggregometry Light transmittance aggregometry (LTA) is a long-standing historic method to measure platelet activator (agonist) mediated aggregation. It is the most common method that used for comparison of newer methods for measuring platelet reactivity and is the predicate method cited for most modern commercially-available methods for measuring platelet reactivity to evaluate the effect of antiplatelet medications. LTA is a time-consuming process with several steps that requires specialized training to minimize variability. LTA is based on the principle that platelets will aggregate in the presence of an agonist. The agonist activates the platelets and stimulates aggregation. Aggregation produces clumps of platelets, which results in an increase in light transmittance. The amount of light transmittance is proportional to the level of aggregation. Citrated whole blood samples for measurement by LTA are prepared by first centrifuging the sample to isolate platelet rich plasma (PRP). The blood sample is further centrifuged at higher speed to isolate platelet poor plasma (PPP). Centrifugation is a source of pre-analytic error in the LTA assay because the process itself can stimulate platelets and affect assay results. Some laboratories still use a consistent platelet count for LTA measurements, so the platelet count in PRP is adjusted using PPP.	The reviewers provide a detailed description of the laboratory procedures and factors affecting analytic performance of light transmission aggregometry (LTA) methods. This is consistent with our understanding of this test. However, a detailed description of the analytical aspects of performing LTA was considered out of the scope of the review (which focuses on summarizing the results of studies fulfilling our inclusion criteria). To provide relevant background to interested readers, we have cited a number of informative reviews for the tests considered in the report. We note that some of these reviews were authored by the reviewers, and thus reflect their thinking on the analytic validity of LTA. Finally, please note that we have extracted information on modifiers of analytic test performance whenever reported in the primary studies.

Commentator & Affiliation	Section	Comment	Response
		The LTA procedure is performed by first preparing a sample of PPP for use as a background measurement. The LTA assay proceeds by pipetting the (adjusted) PRP sample into a cuvette containing a stir bar. Care must be taken to avoid introducing bubbles and stimulating platelets during all pipetting steps. The agonist is added to initiate platelet aggregation. The assay is allowed to run for a period of time, usually 5-8 minutes, and a tracing of % aggregation vs. time is created. Results are reported either as maximum platelet aggregation (MPA), which is proportional to the highest light transmission observed during the assay, or final platelet aggregation (FPA), which is related to light transmission measured at a specific time point near the end of the assay. P2Y12 receptor blockade by P2Y12 inhibitors is evident when the % aggregation result is lower than the reference range of baseline ADP-induced aggregation in a P2Y12 inhibitor-naïve state. The reference range for LTA is variable and depends on the ADP concentration used in the assay, the clinical disease state, and concomitant drug therapy (Paniccia et al, Thromb Haemost 2010). There is no consensus on the reference range of % aggregation results obtained using LTA and different LTA analyzers can have different calibration that influences the results obtained. Furthermore, dietary fat intake produces lipemia that can affect the relative difference in light transmittance between PRP and PPP. There also is no consensus on the concentration of ADP used in the assay, nor is there consensus on whether MPA or FPA should be reported. Therefore, there is no standardization of LTA measurements between laboratories	
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	**Vasodilator-stimulated Phosphoprotein (VASP) Analysis** Platelet P2Y12 inhibitors such as clopidogrel, prasugrel and ticagrelor inhibit thrombosis by directly interfering with platelet cell signaling after binding of ADP to the P2Y12 platelet cell surface receptor. Binding of ADP to the P2Y12 receptor leads to intracellular signaling that involves phosphorylation of vasodilator-stimulated phosphoprotein (VASP), an intracellular platelet phosphoprotein. VASP phosphorylation is regulated by the cyclic adenosine monophosphate (cAMP) cascade. Prostaglandin E1 (PGE1) activates this cascade, and the cascade is inhibited by activation of the platelet P2Y12 receptors by ADP. By triggering PGE1 receptor-mediated	The reviewers' provide a detailed description of the laboratory procedures and factors affecting analytic performance of the Vasodilator-stimulated Phosphoprotein (VASP) assay. This is consistent with our understanding of this test. However, a detailed description of the analytical aspects of performing VASP analysis was considered out of the scope of the review (which focuses on summarizing the results of studies fulfilling our inclusion criteria). To provide relevant background to interested readers, we have cited a number of informative reviews for the tests considered in the report. We note that some of these reviews were

Commentator & Affiliation	Section	Comment	Response
		signaling, binding of PGE1 alone leads to phosphorylation of VASP and GPIIb/IIIa inactivation, while incubating platelets with both ADP and PGE1 has the opposite effect. VASP phosphorylation correlates with the level of blocked P2Y12 receptors, whereas its non-phosphorylation state correlates with the the level of unblocked P2Y12 receptors. Current VASP assays are based on the measurement technique described by Schwarz et al (Thromb Haemost 1999). The VASP assay uses a citrate-anticoagulated whole blood specimen. The specimen is incubated with a high concentration of PGE1 in the presence and absence of ADP. The high concentration of PGE1 is used to stimulate maximum phosphorylation of VASP. The incubation period is immediately followed by fixation and permeabilization. Fixed, permeabilized cells are first labeled with a primary antibody against the phosphorylated Ser-239 residue of VASP (mouse monoclonal anti-VASP-P), followed by staining with a fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse polyclonal antibody and counterstaining with antibody against the platelet surface marker CD61. Background fluorescence is assessed by simultaneously running a negative antibody control. Flow cytometric analysis of sample fluorescence generates a statistic known as the Platelet Reactivity Index (PRI), which is inversely correlated to the level of responsiveness to the anti-P2Y12 drug tested. In other words, the PRI result is related to the percent of active, unblocked platelet P2Y12 receptors on the platelet surface that are available for ADP binding. Results are considered to be related to absolute platelet reactivity to ADP. However, the VASP assay is insensitive to low-tomoderate levels of P2Y12 receptor blockade and so good agreement is not expected with other assays that have a more linear relationship with P2Y12 receptor blockade Judge et al (Thromb Haemost 2010). When evaluating results from the VASP assay in comparison to LTA and VerifyNow results, it is also critical to consider that both platelet activation and platelet aggregation contribute to thrombus formation in vivo. Therefore, the limitation of the VASP assay in comparison to LTA and VerifyNow testing is that there is no measurement of platelet aggregation. VASP results are only indicative of the potential for platelet activation by ADP binding to the P2Y12 receptor.	authored by the reviewers, and thus reflect their thinking on the analytic validity of the VASP assay. Finally, please note that we have extracted information on modifiers of analytic test performance whenever reported in the primary studies.

Commentator & Affiliation	Section	Comment	Response
		P2Y12 receptor blockade by P2Y12 inhibitors is evident when the PRI result is less than the reference range of baseline PRI results in a P2Y12 inhibitor-naïve state. VASP assay PRI results less than 50% are generally accepted to be related to a measurable antiplatelet effect of a P2Y12 inhibitor, though baseline reference ranges have not been established and the test manufacturer makes no conclusive recommendation. Furthermore, the VASP assay is not FDA-cleared and available for clinical use in the United States.	

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	**VerifyNow P2Y12 Test** The VerifyNow P2Y12 Test requires no sample manipulation prior to measurement. Direct, unprocessed citrated whole blood samples are spiked onto a needle in the sample port of the VerifyNow test device. The VerifyNow Instrument controls all steps of the assay automatically and does not require (nor does it allow) any assay calibration by the user. The VerifyNow P2Y12 Test scientific principle is based upon light transmittance and the ability of activated platelets to bind fibrinogen. Fibrinogen-coated microparticles aggregate in whole blood in proportion to the number of expressed platelet GP IIb/IIIa receptors. The rate of microbead aggregation is more rapid and reproducible if platelets are activated; therefore, the reagent ADP is incorporated into the assay channel to induce platelet activation without fibrin formation. The measurement is specific for P2Y12-mediated platelet aggregation through the inclusion of PGE1. Light transmittance increases as activated platelets bind and aggregate fibrinogen-coated beads. Though there are differences in the sample matrix and reagents, the VerifyNow P2Y12 Test has been shown to significantly correlate with LTA. This significant correlation is likely due to the fact that both methods are based on the scientific principle of light transmittance related to platelet aggregation. Both VerifyNow and LTA also measure the combination of platelet activation and aggregation, which produces an ex vivo measurement that is more indicative of the in vivo condition. The VerifyNow P2Y12 Test reports results as P2Y12 Reaction Units (PRU). The PRU result is a measure of absolute platelet reactivity to ADP and is specific for activation via the P2Y12 receptor. Increasing levels of P2Y12 receptor blockade produce decreases in PRU results. P2Y12 receptor blockade by P2Y12 inhibitors is evident when the PRU result is lower than the reference range of baseline platelet reactivity to ADP. A PRU reference range of 194-418 has been established and is reported in the VerifyNow P2Y12 Test package insert.	The reviewers' provide a detailed description of the laboratory procedures and factors affecting analytic performance of the VerifyNow P2Y12 assay. This is consistent with our understanding of this test. However, a detailed description of the analytical aspects of performing analysis using the VerifyNow P2Y12 assay was considered out of the scope of the review (which focuses on summarizing the results of studies fulfilling our inclusion criteria). To provide relevant background to interested readers, we have cited a number of informative reviews for the tests considered in the report. We note that some of these reviews were authored by the reviewers, and thus reflect their thinking on the analytic validity of the VerifyNow P2Y12 assay. We did not collect information on reference ranges or the methods used to establish them. Finally, please note that we have extracted information on modifiers of analytic test performance whenever reported in the primary studies.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	The following table summarizes the similarities and differences between the methods on which the draft report is focused. [Table summarizes the similarities and difference between LTA, VASP, and VerifyNow P2Y12 Test. Parameters: FDA cleared; Specificity for P2Y12 Receptor-mediated Platelet Activation and Aggregation; Platelet Agonist; Sample Manipulation Required Prior to Measurement; Sample Matrix; Materials Required; Pre-analystic Sample Stability; Time to Result After Blood Collection; Quality Control; Established Reference Range; User Involvement in Obtaining Results.]	Thank you for this informative table. We have provided numerous citations to studies that present similar information. We have decided not to reproduce this information in what already is a fairly long review. We have to balance background completeness with report readability.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	The correlations between the methods are statistically significant, though there is variability in the reported correlation coefficients. Several factors can influence the reported correlation between methods, including: • Inconsistent use of Pearson or Spearman correlation coefficient. • Differences in sample matrix. LTA uses plasma; VASP and VerifyNow use whole blood. • Inter-operator variability (greater influence in LTA and VASP measurements). • Test method repeatability. • Differences in scientific principle of the assay. • Differences in reagents, specifically, the concentration of ADP used to induce platelet aggregation and the presence or absence and concentration of PGE1. • Sample transport (the VerifyNow test is point-of-care and can be performed at the location of sample collection; LTA and VASP require testing at a location that is typically separate from the location of sample collection). • Differences in LTA result reported; results can be reported as FPA or MPA. As described above, there are significant fundamental differences between the test methods evaluated in the draft review. LTA and the VerifyNow tests are based on light transmittance, but the VASP assay is not. The VerifyNow test and the VASP assay use whole blood, but LTA uses plateletrich plasma. The VerifyNow test and the VASP assay are more specific for the P2Y12 receptor, but the specificity of LTA is affected by P2Y1 receptor activation as well as the operator's choice of reporting MPA or FPA. LTA and the VASP assay require user involvement in obtaining results, but the VerifyNow test is operator-independent. These differences result in a lack of standardization among platelet function tests. Furthermore, there is a lack of standardization even within some methods, as there is no consensus for the optimal concentration of ADP used in LTA measurements. Taken together, there is no established "gold standard" method for measuring platelet reactivity. Each assay should be evaluated independently on the basis of 1) the reproducibility of results, 2) the ability to specifically measure the effect of the P2Y12 inhibitor, 3) the ability to identify patients at risk, and 4) the ease of integrating the test into the patient care pathway.	We agree that these are some of the potential reasons that may account for the observed difference in analytic performance between methods. Unfortunately, few studies provided direct evidence on the relative effect (or magnitude of effect) for these factors. When available in studies otherwise fulfilling our criteria, this information has been extracted in the report's Evidence Tables. Again, we agree with the reviewers' assertion that there exists no gold standard for the assessment of platelet reactivity; this is also noted in the Final Report.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	**Some analytical performance metrics are less applicable to platelet function testing.** Table 35 in the draft report states "few studies reported information on analytic sensitivity and specificity, possibly reflecting the research community's belief that there is no good reference standard assay for platelet reactivity." As stated in the draft report, studies that do describe analytical sensitivity and specificity report results from a comparison of one method to another method for measuring platelet reactivity. However, as stated in the draft report, there is no "gold standard" for measuring platelet reactivity and as such, reliance on these data to describe the analytical performance of a single method for platelet function testing is limited. Analytic sensitivity and specificity are used in the characterization of tests that measure levels of an analyte. Analytical sensitivity refers to the minimum detectable level of the analyte and analytical specificity describes the resistance of a test method to substances that might cross-react or interfere with the assay and affect results. The term "analytical sensitivity" is not well-suited to describing platelet function testing because the tests are not measuring levels of an analyte; instead, they are measuring platelet function. Low levels of platelet aggregation are consistent with a substantial effect of the drug – these levels are very far away from any clinical decision point described for predictive ability or as evidence of a measurable drug effect. Analytical specificity can be used to describe platelet function tests, and the evaluation of potential cross-reactants and interferents is typically described in the manufacturer's instructions for use. However, because platelet function tests measure platelet reactivity and not levels of an analyte, an important distinction must be made. Substances that adversely affect the assay are classified as interferents or cross-reactants, but there is an important distinction between substances that affect the assay and substances that affect platelet reactivity without adversely affecting the assay. Substances that affect platelet reactivity to ADP, but not the assay itself, should not be considered assay interferents or cross-reactants because the assay is still performing as expected as an *ex vivo* measurement that is reflective of the *in*	We believe that the reviewers definition of analytic validity is more narrow that what we considered for this review. As is stated in the report we have followed the ACCE framework. Under this framework, when studies consider measurements obtained from an assay as a reference standard (Note: a reference standard is *not* necessarily a gold standard) then it is possible to calculate the sensitivity and specificity of a second assay (the "index" test) against the reference standard. To clarify the terminology, the target quantity measured by both the index and reference test is "true" platelet reactivity or on-clopidogrel reactivity status (which – we agree with the reviewers – is a non-measurable latent variable). Note that the same situation appears in many laboratory tests; in all these cases the latent variable approach is useful despite being somewhat non-intuitive. The ACCE framework also includes reliability as a potential domain for the assessment of analytic validity (also covered in our report).

Commentator & Affiliation	Section	Comment	Response
		vivo condition.	
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	The draft report also describes inadequate reporting of analytical performance characteristics in publications as a limitation in the strength of evidence for platelet function testing. However, it is important to note that the bulk of the literature reports on studies that were conducted after the tests in question were approved by regulatory authorities, and the performance characteristics are typically described in the manufacturer's instructions for use.	We reviewed the information that was available to documents submitted to regulatory authorities when available (e.g. through the FDA website). Unfortunately, this information is usually based on studies of healthy volunteers, or studies that were otherwise not eligible (based on our predefined criteria). Detailed reasons on why each identified regulatory document was not included in the report is provided in the Final Report's appendices.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	**Studies of analytic performance in healthy volunteers can provide useful information about platelet function testing.** Studies establishing and validating the analytic performance of tests are commonly performed using specimens from healthy volunteers. We note that studies describing analytic performance using blood samples from healthy normal donors were not considered in the analysis. The platelet function tests as described in the draft review are not used for diagnosing a disease or condition in the context of non-specific symptoms; rather, they are measuring the activity of platelets. Because these phenotypic tests are measuring platelet reactivity and the effect of antiplatelet medications, studies describing analytical performance in healthy volunteers are suitable for characterizing the analytical performance of the method for measuring platelet reactivity. Pharmacokinetic and pharmacodynamic testing (such as platelet function testing) in healthy volunteers plays an important role in optimizing dose selection in the early phases of drug development. We also note the paradoxical comment in Table 35 regarding genotyping, which states "however, based on data on healthy volunteers (not reviewed in this report), the analytic validity of genotyping assays can be considered robust", yet the report also describes finding limited and insufficient information on the analytic validity of genetic testing.	We agree with this point; however, feasibility constraints precluded the expansion of the review scope to studies of healthy volunteers. This decision was discussed with our Key Informants and was pre-specified in the review protocol. Furthermore, the report currently covers more than 120 studies on aspects of analytic validity – we think it is unlikely that studies of healthy volunteers (which provide only indirect information regarding the populations of interest to the report) would substantially affect our conclusions on the analytic validity of tests for platelet reactivity. In addition, there is some evidence (also mentioned by the reviewers' in preceding comments) that the analytic performance of the assays of interest differs across levels of underlying platelet reactivity. It is well established that patients with vascular disease have higher levels of reactivity compared to healthy volunteers, suggesting that assessments of analytic validity using samples from the latter populations may be less relevant to patients who receive antiplatelet treatment in clinical practice. We also appreciate the suggestion about pharmacodynamics endpoints. Again, after consultation with our Key Informants, we did not review data on such outcomes. This point has been noted in the revised report's Methods section.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	**Phenotypic tests allow physicians to determine if there is a measurable antiplatelet effect.** The use of clopidogrel in addition to aspirin has been conclusively proven to reduce the risk for thrombosis compared to aspirin alone. This benefit is a result of the action of clopidogrel in reducing platelet reactivity to ADP. If there is no reduction in platelet reactivity to ADP, there is no associated reduction in the risk for thrombosis. Therefore, **there is value in confirming that patients taking clopidogrel are exhibiting a reduction in platelet reactivity to ADP, and only platelet function testing can objectively provide this information**.	We cannot assess the "value" of specific interventions or clinical practices. Such considerations are beyond the scope of the evidence report.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	The heterogeneity in response to clopidogrel has been well established using platelet function testing, and an inadequate response to the drug has been linked to an increased risk for major adverse cardiovascular events (MACE). This is expected, because all of the benefit and risk associated with taking an antiplatelet medication is solely due to its phenotypic antiplatelet effect. Confirmation of a significant, measurable antiplatelet effect of a P2Y12 inhibitor provides the physician with evidence that the medication is having its desired effect. Based on the presence of a measurable drug effect, one would expect that the antiplatelet effect would confer benefit to the patient.	Please see our reply to the preceding comment. Note that the critical question is not only the direction of expected benefit, but also the magnitude of the effect and the assessment of the strength of the relevant evidence. We believe that this issues were addressed by the report.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	Platelet function test results that are below the reference range of baseline platelet reactivity to ADP provide specific evidence of an antiplatelet effect of a P2Y12 inhibitor. Therefore, it is important for measurement of the drug effect to be highly specific. The VASP assay and the VerifyNow P2Y12 Test are the most specific measurements of platelet reactivity described in the draft report due to their use of PGE1 in the assay. The VerifyNow P2Y12 Test has an established, validated reference range of baseline platelet reactivity to ADP of 194-418 PRU. Because the test is specific for P2Y12 receptor blockade, values less than 194 are highly specific evidence of the antiplatelet effect of a P2Y12 inhibitor. There is widespread variability in agonist concentrations and reported results from LTA; based on this variability there is no established, validated reference range for LTA measurements. The VASP method is not available for clinical use in the United States, but there also is no established, validated reference range for VASP measurements.	Please see our responses to the two preceding comments.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Methods	Describing the sensitivity of the phenotypic test's ability to detect a P2Y12 inhibiting effect of clopidogrel is challenging because there are several factors that can affect the analysis. Various factors, including the CYP2C19 genotype, concomitant medications, comorbidities, and other factors such as compliance have all been shown to be predictors of a poor phenotypic effect of clopidogrel. When evaluating platelet function tests for "diagnostic" sensitivity for detecting an antiplatelet effect, on-treatment platelet function test results that are within the reference range of baseline platelet reactivity to ADP can be viewed as a "false negative" results. Various factors, including the CYP2C19 genotype, concomitant medications, comorbidities, and other factors such as compliance are all predictors of a poor phenotypic effect of clopidogrel. Other factors such as the elapsed time between the most recent drug dose and testing, as well as the potency of the dose/medication can produce "false negative" results. However, these results are more consistent with inter-individual variability in the response to the drug and should be interpreted as evidence of high on-treatment platelet reactivity, not as false negative results that suggest poor performance of the test method.	Thank you for these comments. To the extent allowed by the data reported in published studies, we believe that our review has addressed these issues adequately.
Peer Reviewer #1	Results	Again, overall excellent and very complete review. Perhaps a Figure illustrating a Forest plot of the various meta-analyses done would help. A few more specific comments: 1) ES-10: The ACTIVE-A results for bleeding are the other way around, that is more bleeding in LOF allele carriers (in the clopidogrel group). 2) It seems to me the list of genotyping techniques is not exhausitive. These SNPs have (probably) been geneotyped as part of more extensive panels in some studies, such as Affymetrix or Illumina chips, or Sequenom panels. 3) ES-22: In CURE, we did find an effect modification for GOF alleles. 4) Why no meta-analysis of randomized studies for genetics? See PMID: 22203539 for instance. I do agree it is more difficult for platelet reactivity.	Thank you. 1) We have verified all data presented in the tables or figures. 2) The list in the Introduction was not meant to be exhaustive. The exact genotyping methods used in each study have been extracted and are presented in our Appendix tables. 3) We have provided this information in the revised report. Effect modification was observed only for a sub-analysis of the CURE data. 4) We provide our rationale for not meta-analyzing these studies in the Executive Summary and the main Text of the Final Report. Briefly, the studies were considered substantially dissimilar (not exchangeable) on the basis of differences in populations included and treatments compared.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	Results	A diagram showing all steps at which studies were included or excluded from analysis should be included for transparency in the number of studies contributing to each analysis.	This flow diagram was included in the Results section of the Draft Report (please see page 21 of the original submission; Figure 2 in the main report text). We have prepared a revised version of this flow diagram for the Final Report.
Peer Reviewers 2, 3, 4	Results	Funding sources and year of trial conduct (important with the evolution of cardiac care) should be addressed in the meta-analyses.	Thank you for this suggestion. The relevant information has been extracted from all studies. We did not consider this information in meta-analyses because of poor or inconsistent reporting (of funding information or actual years of study conduct) across studies.
Peer Reviewer #7	Results	More figures summarizing the findings would be useful.	We assume that this refers to the executive summary where we have made effort to conserve space. The main report includes more than 40 figures, which we consider to be adequate for summarizing the available data.
Peer Reviewer #9	Results	The results are solid.	Thank you. No further response necessary.
Peer Reviewer #9	Results	I would like to see the totality of events and individuals stated more often.	We have provided this information in the Final Report, in graphs presenting the meta-analysis results.
Peer Reviewer #9	Results	Throughout the results, the authors should present the number of individuals and events contributing towards each meta analysis (especially for key tables such as Table 3 on p86, and throughout the Results). This will help contextualise what are described in some places as “significant” findings, e.g. for cardiovascular mortality, there are only a total of 24 events which is a very small number to draw any conclusions from.	We have provided this information in the Final Report, in graphs presenting the meta-analysis results.
Peer Reviewer #10	Results	Results are quite detailed, especially the extensive tables or reported studies.	Thank you. No further response necessary.
Peer Reviewer #11	Results	The results section is long, but the amount of detail is appropriate. Tables and figures are especially helpful.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #12	Results	The distinction between analytical validity versus clinical validity/clinical utility for genotype testing needs to be more clearly defined and the methods to assess these different parameters described more fully. The report glosses over these approaches, not presenting reports of genotype testing accuracy or reliability beyond a brief summary of test-retest studies with no quantitative synthesis of this information and only a short discussion of limitations (ES-9). The authors appear to believe that the analytical validity of genotyping technology is so well established that it does not require further measurement and verification, with this view supported by the 501(k) submissions to the FDA in Appendix C. If so, they should make the case that genotype testing is completely accurate both in the executive summary and in the Appendix (which does not appear to contain more information than a list of the relevant 501(k) submissions).	We have provided additional information in the Methods section of the Final Report. Regarding the FDA documents that we reviewed, the Appendix provides detailed reasons on why they were not considered eligible for our analyses.
Peer Reviewer #12	Results	In addition, the report identifies 40 studies that examined the association of genotype results with platelet reactivity measures (ES-11). However these studies are not discussed in sufficient detail or a synthesis provided in the executive summary. These studies provide a type of evidence that is probably best classified as clinical validity. Including a quantitative synthesis of this literature would be useful for several reasons: it would provide additional reassurance that genotype test results are consistently associated with a clinically relevant intermediate outcome, and it would provide insights into how genotype test results relate to platelet reactivity – important since much of the stronger evidence for clinical utility in this field applies only to phenotype testing thereby leaving open the question about the clinical role of genetic testing.	We avoid discussing these studies more extensively in the Executive Summary to conserve space. We have detailed the reasons for not performing meta-analyses for this outcome in the Results section of the Executive Summary and the main text of the Revised Report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #12	Results	As already mentioned, in several cases (an example given above is the LTA testing) if a quantitative synthesis was not done, the justification for not doing so needs to be clearly explained. Even where available literature is not synthesized, it would be helpful to show more forest plots in the executive summary. In particular, for 3a there were several direct comparisons of phenotype test vs. no-test strategies. However, no meta-analysis was performed and forest plots are not included in the executive summary, and do not appear until page 120, figure 32 of the full report. However, as this set of studies comes closest to evaluating the clinical impact of testing, it is potentially the most important data available in the report! In addition, although no meta-analysis is done, it looks to the naked eye as if a meta-analysis would likely give a positive result. These data should be presented graphically in the executive summary and the possibility of meta-analyzing them should be revisited.	Given the space limitations that preclude inclusion of these forest plots and all of the caveats one would have to understand when assessing them, we did not include them. We disagree the suggested additional forest plots would be helpful in their raw form, because they would be encouraging comparisons of the magnitude and precision of effect sizes that are based on very different measurements of platelet reactivity. As you note, they do reside in the report but do so for those interested in a deeper read into the results and in places where caveats regarding their interpretation can be provided.
Peer Reviewer #12	Results	Regarding the association studies (1b&2b and 3a), were the results of the genetic or platelet studies available to the clinicians and/or included in the treatment protocols? If so, there is a risk of this literature being biased, probably toward the null, since treatment and treatment adjustments would have likely lessened the risk of poor outcomes. In essence, the utility of these tests to infer whether testing is useful is limited by not knowing what the effect of the test results were in modifying therapy. This seems to be a major limitation of using these types of studies to infer the potential value of testing in improving outcomes. If this type of bias is likely to have been present, this limitation should be stated more clearly in the report.	Information on blinding has been extracted and is discussed in the Results section of the Final Report. It was also considered in evaluating the risk of bias of individual studies.
Peer Reviewer #12	Results	The interpretation of study designs in which patients are selected on the basis of test results and then randomized to treatments needs clarification (ES-22). Is this different from or the same as random assignment to test/no test strategies? (Essentially, please clarify the difference between study types 2 and 3 as described in ES-21 and ES-22) If so, how? Were such studies actually identified? And if so, then where do the later type of studies fit in?	We have clarified the difference between the two designs in the Executive Summary, as well as the Methods and Results sections.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	**The association of high platelet reactivity with increased risk for MACE has been well established and independently validated.** **We disagree with the draft report's assessment that the strength of evidence regarding prognostic effects of high on-clopidogrel platelet reactivity is low.** The three tests described in the draft report as being predictive of increased risk for thrombosis (LTA, VASP, and VerifyNow P2Y12) have been demonstrated to be predictive of increased risk for thrombosis in several independent studies. Taken together, these tests have been the subject of meta-analyses that have concluded that there is a significant association between high platelet reactivity and the incidence of thrombotic events. When considered individually, the VerifyNow P2Y12 Test has been shown through a published metaanalysis of patient-level data to be significantly predictive of thrombotic events. The tests, when considered individually, have also been shown through the meta-analysis presented in the draft report to be predictive of thrombotic events.	We agree that MACE are relevant to the assessment of tests for platelet reactivity. This is the reason we considered MACE as an outcome of interest in our assessment of individual studies, as well as in meta-analyses (when enough studies were available). Thank you for pointing us to the individual patient data meta-analysis on VerifyNow. Information from this study was summarized in our table of systematic reviews and meta-analyses (in the Discussion section of the Final Report). Our analyses agree with the authors' assertion that there is evidence of a positive association between on-clopidogrel platelet reactivity and adverse clinical outcomes. However, we have reached different conclusions regarding the strength of evidence of this association. Please see the general introductory note at the beginning of this Comment Disposition Document for a detailed discussion of this issue.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	The draft report confirmed previous observations about the association between high platelet reactivity and increased risk for thrombotic events. High platelet reactivity by LTA was reported to be associated with all-cause mortality, cardiovascular mortality, acute coronary syndromes, stent thrombosis, and MACE. However, the methods were heterogeneous. High platelet reactivity by the VerifyNow P2Y12 Test was reported to be associated with cardiovascular mortality, peri-procedural and longer follow-up acute coronary syndromes, stent thrombosis, and MACE. The VASP assay was reported to be predictive of stent thrombosis and MACE, though there was reported heterogeneity among studies for MACE.	No response necessary.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	**The consistency in observations of the association between high platelet reactivity and increased risk for thrombosis for each of the methods is consistent with more than a "low" strength of evidence as concluded in the draft report.**	Please see the general introductory note at the beginning of this Comment Disposition Document for a detailed discussion regarding grading the strength of evidence.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	**There is a high degree of consistency in cutoffs that have been identified as predictive of increased risk.** Inconsistency in the definition of high platelet reactivity was described as heterogeneity among studies and led to a decision not to evaluate the association between high platelet reactivity by the VerifyNow P2Y12 Test and the clinical outcomes of stent thrombosis and stroke. However, the stated heterogeneity did not preclude similar analyses with the other methods described in the draft report.	We agree that this is true for some of the tests we reviewed (e.g., several outcomes for the VerifyNow assay mentioned by the reviewers). In such cases we actually performed meta-analyses. However in some cases there was heterogeneity in the metrics (not cut-offs) used in the studies. For example some studies used change-from-baseline reactivity whereas others used absolute reactivity. This has been clarified in the results section of the Final Report.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	The majority of published reports evaluating the association between high platelet reactivity measured by the VerifyNow P2Y12 Test and thrombotic events describe “optimal” cutoffs that are between 208-240 PRU, with most being between 230-240 PRU. This range represents less than a 5% difference in test results, which is well within the manufacturer’s stated precision claim of < 10%.	Please see above for a response regarding heterogeneity in the metrics of reactivity.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	The draft report also cites the lack of using a training dataset to establish a proposed cutoff followed by validation of the proposed cutoff in a separate dataset. While we agree that this is a preferred approach, there are other methods to evaluate the robustness of cutoff selection. An analysis to determine the robustness of cutoff selection was performed in a meta-analysis of patient-level data reported by Brar et al (J Am Coll Cardiol 2011). As described in the publication, “he cohort was randomly divided into a derivation and validation dataset, with 50% of the sample distributed to each dataset. In the derivation dataset, bootstrap estimates (sampling with replacement) of the PRU threshold were calculated for 100 iterations, yielding the best average cutoff and 95% confidence interval (CI). For estimates of standard errors and normal approximation CI, 100 bootstrap replications are generally adequate. Next, Kaplan-Meier failure estimates and hazard ratios (HR) were calculated using the PRU threshold in the derivation and validation cohorts.” The results of this analysis led the authors to conclude that the selection of the PRU = 230 cutoff was appropriate and robust. Taken together with the individual studies reporting similar “optimal” cutoffs, there is ample evidence for consistency in the cutoffs used to define high platelet reactivity by the VerifyNow P2Y12 Test and the association with increased risk for thrombosis.	We agree that resampling methods (e.g. cross-validation, jackknifing, etc.) can provide valid estimates of test error rates. This has been noted in the Discussion section of the report. Thank you for pointing us to the individual patient data meta-analysis by Brar et al. The results from this work have been included in the Discussion section of the report (in the table summarizing meta-analysis results and in the text).

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	**MACE outcome reporting is acceptable for evaluating the predictive ability of platelet function tests.** The draft report cites selective outcome reporting as a limitation in the strength of evidence for the predictive ability of platelet function testing. This conclusion was reached because some studies reported total MACE without reporting individual component endpoints. Contrary to the conclusions of the authors, reporting of total MACE events is not a limitation in this area. Rather, it is very consistent with how antiplatelet therapies are evaluated for efficacy. For example, clopidogrel was evaluated in the CURE trial and prasugrel was evaluated in the TRITON-TIMI 38 trial using a composite endpoint of cardiovascular death, myocardial infarction and stroke. P2Y12 inhibitors are given to reduce reactivity of the platelet P2Y12 receptor to ADP, and these drugs reduce the risk of thrombosis through their pharmacodynamic effect. They are not given to selectively reduce only one of the component endpoints. Thrombosis is a cause of each of the individual components of a MACE endpoint (which can include such individual components as cardiovascular death, myocardial infarction, stroke, and stent thrombosis), so it is reasonable and completely appropriate to report the predictive ability of a platelet function test using a composite MACE endpoint.	We agree that MACE represents an outcome of interest for the tests evaluated in the report. This is why we have extracted information on this outcome and performed meta-analyses, whenever possible. We do not consider the reporting of MACE outcomes as a limitation. However, we think that when MACE is used as an outcome, the component outcomes also need to be reported. This is consistent with the recommendations for many independent methodologists.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	**The PCI patient population is ideal for evaluating the predictive ability of platelet function tests.** The purpose of prescribing P2Y12-inhibiting antiplatelet medications is to reduce the risk for thrombosis by reducing platelet reactivity to ADP. This is an effect on the patient's platelets, not on the underlying disease. Therefore, patients without evidence of a measurable drug effect should be considered to be at increased risk for thrombosis. The draft report cites the lack of data in nonischemic heart disease patient populations as a limitation in the evidence for genotyping and platelet function testing. Most studies of the association between high platelet reactivity and increased risk for thrombotic events have been performed in patients with ischemic heart disease undergoing percutaneous coronary interventions (PCI). This is an ideal population for these studies because 1) it is a population of convenience for measuring the effect since the antiplatelet therapy in question is prescribed as the standard of care in this population, and 2) it is a well-characterized population that is the subject of evaluations of P2Y12 inhibitors in pivotal phase III trials, so there is population consistency in the evaluation of the antiplatelet therapy and platelet function testing. The PCI patient population is the most reasonable population to evaluate the utility of platelet function testing because it is a well-characterized and widely available population where P2Y12 inhibitors are prescribed as the standard of care. The PCI population is a population at high risk for thrombotic events. Therefore, the association between high platelet reactivity and increased risk for thrombosis that has been established through studies of the PCI population provides further evidence that the **predictive ability of platelet function tests is clinically important.**	We agree with this point. This is why we extracted data and performed analyses in PCI populations. However, we maintain that it is not straightforward to extrapolate from PCI populations to other populations who are candidates for clopidogrel treatment (e.g. patients with chronic CAD not undergoing PCI or patients atrial fibrillation and contraindications to warfarin treatment). As such, the applicability of the findings of studies on PCI populations to other patient populations may be limited.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Results	**Platelet reactivity should be considered a viable surrogate outcome measure.** The draft report states that the authors did not assess the strength of evidence for studies exclusively assessing platelet reactivity as an outcome. Studies using platelet reactivity as an outcome provide additional support for the importance of platelet function testing because they are focused on the platelet hypothesis – if platelet reactivity is reduced, then the risk for thrombosis is reduced (Gurbel et al., Expert Rev Cardiovasc Ther. 2004). As previously stated, the purpose of prescribing P2Y12-inhibiting antiplatelet medications is to reduce the risk for thrombosis by reducing platelet reactivity to ADP. Platelet function testing was used to support dose selection in the development of the platelet P2Y12 inhibitors clopidogrel, prasugrel and ticagrelor to confirm that a targeted level of on-treatment platelet reactivity was achieved. Furthermore, the "optimal" platelet reactivity cutoffs reported in the literature are consistent with the baseline reference range of reactivity to ADP, consistent with the absence of a measurable pharmacodynamic effect of the drug. If there's a measureable pharmacodynamic effect, we would expect to see benefit (efficacy). Considering that measurements of platelet reactivity were used to support dose selection during the drug development process and that high ontreatment platelet reactivity has been consistently reported to be associated with increased risk for thrombosis, we urge the authors to re-consider their decision not to assess the strength of evidence for platelet reactivity as an outcome measure. Platelet function testing is already being used as the primary outcome measure for studies evaluating the utility of CYP2C19 genotyping as well as studies involving pharmaceutical agents.	We agree with this point. For this reason we reviewed studies reporting on pharmacodynamics outcomes. Data from these studies have been extracted and are presented in the report.
Peer Reviewer #1	Discussion/ Conclusion	I am in agreement with the authors' discussion/conclusions. The future research section is clear and logical given results of the comprehensive review.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	Discussion	It is unclear why the stent thrombosis-CYP 2C19 association strength of evidence is assessed as low.	We assessed the SOE as low for the following reasons: (1) of the more than 100 studies reporting on clinical outcomes only a minority had data on stent thrombosis indicating that reporting bias is likely; (2) studies had moderate ROB as detailed in our assessment; (3) larger (more precise) and smaller (less precise) studies produced discrepant results (this also resulted in a statistically significant test for small study effects); (4) as for all comparisons in the report, we were concerned about exposure heterogeneity. These points were explicitly stated in the full text of the report and the SOE assessment (Discussion section). We note that our assessment is concordant with the independently performed meta-analyses by Bauer (BMJ 2011) and Holmes (JAMA 2011). Notably, the work by Bauer reached almost identical conclusions based on the GRADE framework (a framework for the assessment of SOE that is very similar to that used by the EPCs).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	Discussion	There seems to be a lack of distinction between those associations judged as "low" strength of evidence and "insufficient". The rating of each study for bias, consistency, directness, and precision should be tabulated so it is clear how each study fits into the aggregate assessment.	The key distinction between topics with "low" and "insufficient" SOE was the number of available studies reporting information on each outcome (which affect our assessment of precision and risk of reporting bias). The rating of *each individual study for ROB* was provided in the Draft report (please see the corresponding Supplementary Tables by Key Question). Some components of ROB are only applicable across studies (e.g., publication bias). Consistency, directness and precision are generally evaluated across studies. For example, *consistency* is meaningful only when results are evaluated across independent estimates of a parameter. *Directness* refers to a judgment of how the evidence fits the analytic framework. *Precision* typically refers to a summary (meta-analytic estimate). In case where no meta-analysis is performed we have evaluated precision at the study level (this was generally not encountered in the review of CYP2C19 variants). A summary of the reasons on each of our SOE determinations is provided in the SOE table in the Discussion section of the report. Additional elaboration regarding SOE is provided in the text of the same section of the Final Report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewers 2, 3, 4	Discussion	The authors assert that "studies provided limited information on the value added by these tests over ascertainment of conventional risk factors in the populations of interest (e.g., clinical or laboratory information or disease-specific predictive scores)." The authors should include a summary of the comparative risks of genetic vs. nongenetic factors routinely considered in the clinic where multivariable analyses are reported (as has been published for genetic risk factors type 2 diabetes vs. family history). Additionally, it would be appropriate to frame the relative risks between CYP2C19 genotypes against the effect of clopidogrel over placebo.	We generally refrain from such informal indirect comparisons between prognostic factors for clinical outcomes. Such comparisons rely on the assumption that the populations assessed in different studies have the same distribution of risk factors and the same event rate (in the baseline group used for each comparison). Second they require that outcome definitions are shared between studies and outcome ascertainment is similar enough. Third, they ignore whether the factors compared are modifiable (smoking is; CYP2C19 genotype is not). Fourth, they ignore the potential correlation between the compared exposures (e.g., CYP2C19 genotype is correlated with platelet reactivity; diabetes is correlated with increased reactivity, etc.).
Peer Reviewers 2, 3, 4	Discussion	Comparative effectiveness of a pharmacogenetic strategy cannot directly be evaluated in the absence of data. The review should discuss the design of trial needed to adequately evaluate the utility of a pharmacogenetic strategy. Power calculations based on the anticipated effect sizes (given available data) should accompany this discussion to put the available data in a feasibility context. This will facilitate appropriate interpretation of ongoing trials to evaluate the efficacy of pharmacogenetic strategies.	Extensive power calculations are usually not performed for Future Research Needs sections of Comparative Effectiveness Reviews. We believe that conducting analyses of appropriate breadth (e.g., under different assumptions for the underlying population risk or magnitude of prognostic effect) are out of the scope of the current report. The preparation of a stand-alone Future Research Needs document (which would include analyses such as those suggested) typically takes over two months and includes an extensive effort to elicit stakeholder input. We were not tasked with preparing such a document.
Peer Reviewers 2, 3, 4	Discussion	The authors assert that the "studies differ in the alleles genotyped and the genotype groupings used, leading to heterogeneity in the exposure definition." This would introduce a null bias if anything. Throughout the document, the directionality of any bias introduced should be acknowledged so the results are given the appropriate context.	We agree that it is important to describe the direction and magnitude of suspected bias, and made a conscious effort to do so, whenever possible. However, we respectfully disagree that heterogeneity of exposure ascertainment will invariably bias toward the null. The direction and magnitude of the bias will depend on the actual groupings used across studies and the specific reporting patterns. For example, some authors may report only results from grouping of genotypes that produce statistically significant results.
Peer Reviewer #7	Discussion/ Conclusion	I would prefer to see more emphasis on the studies that look at genetic tests as effect modifiers - here it certainly appears that there is not a lot of evidence that clopidogrel has different clinical effects based on genotype.	We have expanded this information in the Final Report and supplemented it with information from additional studies uncovered in our update.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	Discussion/ Conclusion	The statement on page 217 "Although our review of observational (mostly cohort) studies identified some evidence to support the association between loss-of-function CYP2C19 variants and increased rates of cardiovascular events" I would add the word "weak" – i.e. "…identified weak evidence to support" based on the data presented. Also, with the statement "We found some evidence supporting a significant association between loss-of-function CYP2C19 variants and increased risk of stent thrombosis and cardiovascular mortality." This does not reflect the uncertainty surrounding these values. As already stated the number of events for cardiovascular mortality was tiny. And stent thrombosis showed strong evidence for small study effects using the Egger test. Therefore I think this is placing too much emphasis on the summary estimate per se rather than actually describing how robust this is.	We have used phrasing similar to that suggested in the SOE assessment to convey this information to readers. Please see the Methods section for a discussion on why do not place an overly strong emphasis on the results of publication bias tests.
Peer Reviewer #9	Discussion/ Conclusion	The future research section is clear and pertinent.	Thank you. No further response necessary.
Peer Reviewer #10	Discussion/ Conclusion	Yes. Especially the need for a prospective randomized trial to assess whether alternative therapy based on genotype improves outcomes.	Thank you. No further response necessary.
Peer Reviewer #11	Discussion/ Conclusion	Implications of major findings are clearly presented, as are the conclusions, evidence gaps and future research.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #12	Discussion/ Conclusion	The recommendations of this report are largely to not support nor endorse testing for platelet reactivity/response to clopidogrel. This recommendation derives principally from a lack of evidence in the literature showing that the tests have value. However, if there was enough literature and if the tests were found to have a favorable effect on specific outcomes of interests then one would have to balance these benefits against the harms associated with the treatment. The review did not include nearly as much attention to the harms side of the equation, nor did it propose an approach to balancing benefits versus harms in order to determine "on net" whether the benefits would outweigh the detriments. The issue of bleeding as the harm likely to enter into treatment decisions based on these tests deserves more attention. For example highlighting the data showing a borderline-significant increase in risk of bleeding in gain-of-function carriers (page 37) combined with the results of the active A trial (page 39) as a potential signal. Admittedly, this was not a pre-specified focus of the review; however some attention should be given to this issue in order for the review to provide a starting point and template for future report updates.	Please note that we make no clinical practice recommendations at all, as is mandated for EPC reports. Instead only perform a comprehensive review of the literature and assess the strength of evidence of the available body of evidence. Recommendations about incorporating (or not incorporating) the tests covered by our review are left to clinicians, policymakers, guideline-issuing bodies, and patients. We have explicitly mentioned the need to consider the tradeoff between harms and benefits of testing in the Final Report's Introduction.

Commentator & Affiliation	Section	Comment	Response
Joseph Brent Muhlestein	Discussion	In the discussion section, page 157, under "Phenotypic Testing For Platelet Reactivity", a statement from the recent consensus document from the Working Group on High On-Treatment Platelet Reactivity was quoted as such: There are "limited data to support that alteration of therapy based on platelet function measurements actually improves outcomes." This statement is quoted in a way to imply that the authors of the concensus document do not recommend the clinical use of platelet function testing. To more fairly represent the opinions of that group, I recommend that their concluding statement also be included: "Currently, platelet function testing may be considered in determining an antiplatelet strategy in patients with a history of stent thrombosis and in patients prior to undergoing high-risk PCI. However, until the results of large-scale trials of personalized antiplatelet therapy are available, the routine use of platelet function measurements in the care of patients with cardiovascular disease cannot be recommended." Additionally, I recommend that the 2012 ACCF/AHA guidelines statement regarding the potential clinical use of both platelet function and genetic testing be also referenced: "Class IIb 1. Platelet function testing to determine platelet inhibitory response in patients with UA/NSTEMI (or, after ACS and PCI) on P2Y12 receptor inhibitor therapy may be considered if results of testing may alter management.(Level of Evidence: B) 2. Genotyping for a CYP2C19 loss of function variant in patients with UA/NSTEMI (or, after ACS and with PCI) on P2Y12 receptor inhibitor therapy might be considered if results of testing may alter management.(Level of Evidence: C)" By including these statements as well, I believe that you will produce a more balanced summary of the current state of affairs which is that although definitive trials have not yet been performed that demonstrate significant clinical utility for platelet function or genotype testing, similarly, definitive trials have not yet been performed that demonstrate lack of utility. Certainly avilable evidence suggests that there is a likelyhood that some groups of patients might benefit clinically from having these tests performed. The ACCF/AHA National Guidelines have attempted to address that possibility and I recommend that something like that be included in this document as well.	We agree that our quotation from the Working Group's recommendation could be misinterpreted. We have revised the text accordingly. We agree that the 2012 ACCF/AHA guidelines are relevant and have mentioned them in the Discussion section of the revised report.

Commentator & Affiliation	Section	Comment	Response
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Discussion	Table 38 cites "combinations of tests (e.g., genetic and phenotypic testing) or combinations of genetic tests were rarely assessed." Genetic risk factors elicit their effect only through influence on phenotype, and the CYP2C19 loss-of-function genotype is a risk factor for high platelet reactivity, but is not an absolute predictor of poor phenotype. Data published from the ELEVATE-TIMI 56 trial (Mega et al JAMA 2011) suggest that approximately 25% of patients with the wild-type allele have high platelet reactivity on a standard clopidogrel maintenance dose and approximately 50% of the carriers of the loss-of-function allele do not have high platelet reactivity a standard clopidogrel maintenance dose.	We agree that none of the tests considered in this review are "perfect" predictors. This actually supports the need for comparing the tests directly in large, prospectively designed studies.
Paul A. Gurbel, Matthew J. Price, Robert F. Storey, and Paul S. Teirstein	Discussion	Table 38 cites "no studies reporting valid direct comparisons the predictive value of different tests were available" as an evidence gap, but this was done in the POPular study (Breet et al JAMA 2010).	The Results section of the Report provides details on what constitutes a valid comparison between alternative tests. We maintain that the POPular study did not perform such comparisons, not did it report adequate data for us to perform the required analyses.
Peer Reviewer #1	Clarity and Usability	The ES could be shortened. This would help increase emphasis on important "take home" messages. For instance, many results could be summarized in a few lines supported by Tables and/or Figures (e.g. ES-15 to ES-18).	We have made every effort to streamline the Executive Summary.
Peer Reviewer #7	Clarity and Usability	The report is well organized, yet deadly dull. It would be greatly improved by clearer writing.	We have made every effort to make our text clear to readers. That said, this remains a technical document reviewing evidence on a large number of complex laboratory tests.
Peer Reviewer #9	Clarity and Usability	Yes the structure is good.	Thank you. No further response necessary.
Peer Reviewer #10	Clarity and Usability	This is a very detailed analysis containing statistical concepts that may not be obvious to some readers. Perhaps a glossary of terms would be helpful.	Thank you. Instead of providing a glossary we have provided detailed references for interested readers. It is unclear if a glossary could offer concise and at the same time accurate descriptions of all technical aspects of the report.
Peer Reviewer #11	Clarity and Usability	The report is logically organized and well structured. Main points are presented clearly in text and tables. The report should be useful towards informing both policy and practice decisions.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #12	Clarity and Usability	The report is structured and organized appropriately for an AHRQ review. The main points are clearly presented. Our principal suggestion is that more detail about the methodological issues mentioned above be included in the executive summary. Specifically, we recommend that a more focused discussion of the phenotype test versus no-test strategies be included with a more complete discussion of benefits versus harms. Both of these points are described in more detail above. We believe that the conclusions of this report can, and should be used to inform current clinical policy and practice decisions. However, including more discussion of the limitations of the available literature should make the recommendations more, rather than less, relevant and applicable.	Thank you for your comments. Please see above for our detailed responses to the issues summarized in your concluding remark.

Abbreviations: SOE = strength of evidence; ROB = risk of bias.