Effective Health Care Program

Future Research Needs Paper
Number 1

Future Research Needs for the Comparison of Percutaneous Coronary Interventions With Bypass Graft Surgery in Nonacute Coronary Artery Disease

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

This report is based on research conducted by the Tufts Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10055-I). The findings and conclusions in this document are those of the author(s), who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care researchers and funders of research make well-informed decisions in designing and funding research and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of scientific judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical research and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances.

This report may be used, in whole or in part, as the basis for research design or funding opportunity announcements. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This information does not represent and should not be construed to represent a determination or policy of the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

This project was funded under Contract No. 290-2007-10055-I from the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (HHS).

Future Research Needs Paper

Number 1

Future Research Needs for the Comparison of Percutaneous Coronary Interventions with Bypass Graft Surgery in Nonacute Coronary Artery Disease

Identification of Future Research Needs from Comparative Effectiveness Review No. 9

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

Contract No. 290-2007-10055-I

Prepared by:
Tufts Evidence-based Practice Center
Boston, MA

Investigators
Thomas A. Trikalinos, M.D., Ph.D.
Issa J. Dahabreh, M.D.
John Wong, M.D.
Madhumati Rao, M.D., Ph.D.

AHRQ Publication No. 10-EHC068-EF
September 2010

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Trikalinos TA, Dahabreh IJ, Wong J, Rao M. Future Research Needs for the Comparison of Percutaneous Coronary Interventions with Bypass Graft Surgery in Nonacute Coronary Artery Disease. Future Research Needs Paper No. 1. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 10-EHC068-EF. Rockville, MD: Agency for Healthcare Research and Quality. September 2010. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the Children's Health Insurance Program (CHIP).

AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting comparative effectiveness reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see http://effectivehealthcare.ahrq.gov/reference/purpose.cfm.

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence.

As part of a new effort in 2010, AHRQ has supported EPCs to work with various stakeholders, including patients, to further develop and prioritize the future research needed by decisionmakers. The Future Research Needs products are intended to inform and support researchers and those who fund research to ultimately enhance the body of comparative effectiveness evidence so that it is useful for decisionmakers.

Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an email list to learn about new program products and opportunities for input. Comparative effectiveness reviews will be updated regularly.

AHRQ Contacts

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

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

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Center Program
Agency for Healthcare Research and Quality

Elisabeth U. Kato, M.D., M.R.P.
Task Order Officer
Evidence-based Practice Center Program
Agency for Healthcare Research and Quality

Contents

Tables

Figures

Executive Summary

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its comparative effectiveness reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Background

Percutaneous coronary interventions (PCI) with or without stents and coronary artery bypass graft surgery (CABG) are the two broad categories of interventions for mechanical revascularization of atherosclerotic coronary arteries in patients with coronary artery disease (CAD). Generally, both approaches would be clinically relevant for patients with single-vessel disease of the proximal left anterior descending artery, most types of two-vessel disease, as well as for patients with three-vessel disease that is not particularly extensive. Because PCI and CABG differ in their procedural risk and their initial and downstream costs, assessing their comparative effectiveness and safety is of great interest.

PCI and CABG have already been compared in several randomized controlled trials (RCTs) and analyses of large clinical registries. The aim of this report is to identify needs for future research in the comparison between PCI and CABG. We use as a basis the 2007 comparative effectiveness review (CER) by the Stanford–University of California at San Francisco Evidence-based Practice Center (Stanford–UCSF EPC) that assessed PCI vs. CABG for coronary artery disease (hereafter we refer to the report as the "Stanford CER" for brevity). The CER summarized evidence published through 2006. Table I lists the key questions of the CER and a qualitative synopsis of the findings:

Table I. Summary of findings of the Stanford CER

Key Question	*Summary of findings*
In patients with ischemic heart disease and angiographically proven single- or multiple- vessel disease, what is the comparative effectiveness of PCI compared to CABG in reducing the occurrence of adverse objective outcomes and improving subjective outcomes? Over what period are the comparative benefits of PCI and CABG sustained?	Procedural survival was high for both interventions and was not significantly different between PCI and CABG. Freedom from procedural strokes significantly favored PCI compared to CABG, but freedom from procedural myocardial infarction was heterogeneously defined and was not significantly different between the two treatments. Long-term survival was significantly better for CABG compared to PCI in the older trials (pre 1994, where PCI was mostly balloon angioplasty). However, long term survival was not significantly different between CABG and PCI trials in the stent era. Freedom from myocardial infarction was not statistically significant between the compared treatments. Regarding long-term subjective outcomes, freedom from angina was greater for CABG rather than PCI interventions. Quality of life outcomes favored CABG over PCI between 6 months and 3 years of followup but equalized thereafter. The degree of improvement in quality of life was correlated with relief of angina.

Key Question	*Summary of findings*
Is there evidence that the comparative effectiveness of PCI and CABG varies based on (predefined) patient and procedure related factors?	The Stanford CER evaluated the evidence for comparative effectiveness of PCI and CABG based on (predefined) patient-, provider-, disease- and intervention- related factors. However, the report noted that in contrast to the fairly robust evidence concerning overall clinical outcomes, there was much less evidence from randomized trials to gauge whether and how comparative effectiveness varied across these factors. Most clinical trials reported only on survival and not on other outcomes in key patient subpopulations.

CABG=coronary artery bypass surgery; CER=comparative effectiveness review PCI=percutaneous coronary intervention

The Stanford CER includes almost all large RCTs that are available as of this writing. It is therefore a relatively current report. However two limitations should be noted. Medical treatment (no revascularization) is now considered a clinically valid option for many patients with nonacute CAD, but it was not in the scope of the Stanford CER. Therefore the Stanford CER does not represent a full evaluation of the actual clinical options that are available today. Second, several RCTs have published their results after the completion of the Stanford CER and several ongoing RCTs are expected to report results in the next 2 or 3 years. This means that the following evidence gaps identified in the Stanford CER should be assessed in the light of the newer evidence. The identified gaps were:

1. Paucity of published analyses of PCI and CABG outcomes according to patient characteristics. Specifically highlighted subgroups include women, the elderly, patients with chronic kidney disease, and patients with left ventricular dysfunction.
2. Paucity of data on the comparative effectiveness and safety of drug eluting stents (DES) in particular, and especially in the long term.
3. Paucity of data on the relationship between procedural volume and outcomes of minimally invasive approaches to CABG. It cannot be assumed that the relationships defined for standard CABG will apply to minimally invasive CABG. For example, the volume thresholds of minimally invasive CABG may differ, given the differences in the technique.
4. Paucity of data on metrics for quality of care for PCI and CABG procedures.

Methods

We aimed for a practical approach to identifying evidence gaps and prioritizing them into research needs. This approach was based on combining limited stakeholder input, focused searches of the literature, and results of simple modeling to derive a prioritized list of research needs which could then be fitted to possible research designs. We focused on the comparison of PCI interventions and CABG in patients with nonacute CAD in whom revascularization is warranted because of their symptoms or because of the extent of their disease.

We followed an iterative process to compile a list of important evidence gaps. First, we generated an initial list of evidence gaps based on the Stanford CER and initial feedback from a group of key informants. Subsequently, we searched the literature to identify trials that have been published after the Stanford CER, and ClinicalTrials.gov to identify ongoing studies in the field. Taking into account the results of the literature searches, we proceeded to further refine the initial list of evidence gaps in one-to-one interviews with key informants. This resulted in an interim expanded list of evidence gaps, which was pruned according to *post hoc* criteria, but without direct and explicit feedback from stakeholders. This pruning step was not anticipated in

the beginning of the pilot project but it was considered necessary in view of the expansion of the evidence gaps list at the prior steps.

We classified evidence gaps into four thematic areas of future research that are not overlapping, could be pursued independently, and are amenable to different types of study designs (and therefore do not all require extensive resources to pursue):

A. The comparative effectiveness and safety of PCI vs. CABG
B. The role of testing to inform choice of revascularization procedure
C. Methods for enhancing patient participation
D. Methods for assessing performance

Only the first thematic area was directly related to the scope of the Stanford CER. The other three thematic areas were added based on key informant input.

We then recommended specific research designs for the most important evidence gaps within each thematic area. To this end we assessed the importance of each evidence gap, estimated the feasibility of different research designs, asked for additional key informant feedback, and performed focused modeling analyses to pinpoint for which parameters it is most important to obtain additional information through future research.

Specifically, we considered the following *ad hoc* criteria to prioritize research needs: feasibility in terms of research costs, feasibility in terms of projected study duration, likelihood that the study will have nontrivial findings, likelihood that the study will provide unbiased results (to inform clinical practice), and likelihood that ongoing research will address the evidence gap. We considered that a new study would be infeasible if it were too expensive or complex to conduct, if it required too long a followup (beyond 5-7 years), or if it relied on information or data that is not yet available. Generally, RCTs are among the most expensive research designs. Some of the largest recent efficacy RCTs in nonacute CAD have sample sizes in the neighborhood of 2,500 patients. Using this as a reference, we commented on the feasibility of other research designs. We performed sample size calculations using standard formulae for a two-sided, chi-squared test at the 0.05 level of significance.

For research needs falling in the first thematic area (comparative effectiveness of PCI vs. CABG) we also performed focused modeling to pinpoint which parameters are most important to study in future research.[a] We developed simple decision models that compare different revascularization options in typical RCT participants, elderly patients (older than 75 years), and diabetics. Estimates for model parameters were derived from published sources. We used network meta-analysis to obtain consistent estimates for all treatment effects. We used different types of analyses (decision analysis, cost-effectiveness analysis, and value of information analysis) to identify the most influential parameters in the models.

Results

We present our recommendations for future research in each thematic area (Table II). Only the first thematic area was in the scope of the Stanford CER, and, thus, there are limitations

[a] Assuming a reasonably simple specification of this decisional problem, such parameters may be the prevalence of procedural deaths or procedural strokes; the relative effects for procedural deaths or strokes across the compared interventions; the frequency of long term deaths, myocardial infarctions, strokes, or repeat revascularizations; the corresponding relative effects across revascularization options; patient preferences (utilities); and immediate and downstream health care costs. This has practical implications, because different parameters are naturally amenable to different research designs.

to our research recommendations for the last three thematic areas. This is because we did not have the benefit of a thorough evaluation of the literature, and it is possible that the identified evidence gaps have been addressed.

Comparative Effectiveness of PCI vs. CABG (First Thematic Area)

We distinguish three tiers of pressing research needs in the first thematic area (Table II).

1. The first tier is studies assessing the comparative effectiveness of PCI vs. CABG among the elderly, e.g., patients older than 75 years. Up to 40 percent of people who received revascularization in large registries were 75 years or older, but elderly patients are grossly underrepresented in the existing RCTs. In addition, there is evidence of strong age-by-treatment interaction. In a recent meta-analysis of individual patient data, survival was better with GABG compared with PCI for older patients (p-value for trend was 0.002 across the age categories of younger than 55, between 55 and 65, and older than 65 years). The magnitude of the treatment effect modification was clinically significant, but it is unclear if it generalizes to older patients.
2. The second tier of future research pertains to further study of patients with heart failure and patients with stage 3 or stage 4 renal disease. Although a recent meta-analysis of individual patient data found no significant interaction between heart failure and treatment effect for overall survival (p=0.46), patients with heart failure were excluded from many randomized trials that formed the basis of the meta-analysis. Thus we deem that the interaction of congestive heart failure and treatment choice remains unclear. We identified no analysis of randomized data on patients with chronic kidney disease.
3. The third tier pertains to studying the interaction of diabetic patients and treatment effects. There is already suggestive evidence that survival is better with CABG compared with PCI in diabetics (p=0.014 for treatment by diabetes interaction); such an interaction may be of substantial clinical significance if confirmed by further studies. There are at least two large ongoing trials comparing PCI vs. CABG [FREEDOM (n=2400, NCT00086540) and VACARDS (n=790, NCT00326196)] in diabetic patients whose results should be available within the next 2–3 years. This makes obtaining additional data for diabetics less pressing than for the elderly, heart failure patients, or chronic kidney disease patients.

Finally, we deemed that the subpopulation of women and people who have received PCI in the past represent less pressing priorities than the above, and do not recommend them as priority areas for future research. The previously mentioned meta-analysis of individual patient data found no interaction between sex and treatment effect (p=0.25) for overall survival. The meta-analysis did not report analyses with respect to having a prior revascularization, but key informants suggested that there is little if any evidence for an interaction of treatment with history of revascularization with PCI.

For each of these subpopulations, DES represent a more pressing research need compared to BMS, because they are used more often than BMS in everyday practice, and this trend will most likely continue in the midterm. The above does not mean that future studies should not use BMS. If numbers allow, an option would be to further randomize patients in the PCI arm to BMS vs. DES.

Based on our modeling, new research should inform on the relative effects of treatments on the clinical outcomes of death, myocardial infarction, stroke, and secondarily

revascularization in the mid and long term. This thematic area of research is already a mature research field, where many RCTs have been performed already, but not in the subpopulations of interest. In terms of research design, it is probably prudent to first perform reanalyses of existing data, and consider performing new RCTs only if existing data are not applicable to the population of interest, or suggest a subpopulation-by-treatment interaction that is clinically important. Alternatively, one could perform efficient *de novo* nonrandomized comparative studies and proceed to a *de novo* RCT only if there is suggestive evidence of an important treatment-by-subgroup interaction. This is because, based on our sample size and power calculations, a *de novo* RCT would likely require resources comparable to recent large multimillion dollar RCTs. Reanalyses of existing data could be collaborative meta-analyses of individual patient data from prior RCTs or from large observational studies, if the populations of interest are underrepresented in prior RCTs.

Testing To Inform Treatment Choice (Second Thematic Area)

There is currently not enough data to determine which baseline test or combination of tests may be the best for predicting which patients will have a better response with CABG than with PCI, or vice versa. However, if baseline (pre-revascularization) testing were able to predict long term differential response to PCI or CABG, its effects on patient health would be substantial. For this reason, further elaboration of the ability of testing to identify the optimal intervention for each patient is highly desirable. Based on key informant input, we listed four testing options, either invasive (arteriography) or noninvasive (magnetic resonance or computerized tomography angiography; resting or exercise single photon emission computerized tomography; exercise treadmill testing with or without echocardiography).

There are many different test-and-treat strategies that can be conceived, and it is not possible to compare all of them with randomized studies. The most practical recommendation for future research is to explore the predictive accuracy (sensitivity and specificity) of the tests of interest or combinations thereof by performing cohort studies, case control studies, or by reanalyzing baseline data from existing RCTs (if available). To be clinically useful and informative, such studies should enroll patients representative of those seen in clinical practice, and limit verification and other biases. Such studies should be relatively feasible given the high prevalence of the conditions of interest, the widespread availability of the considered diagnostic technologies, and the fact that most patients treated with PCI or CABG already receive extensive workup before their revascularization.

Enhancing Patient Participation (Third Thematic Area)

Life expectancy is not the only outcome that matters to patients and those close to them. While the importance of quality of life is generally appreciated by physicians and decisionmakers it is not always clear whether, how, and to which extent it is considered. To facilitate the participation of patients in decisions regarding their treatment, one has to develop and evaluate decision aids.

A suitable study design to elicit patient preferences is a survey of patients, or a qualitative research study that uses focus groups of patients. It is at best unclear whether a study of patient preferences will have a major impact on the remaining research agenda. However, compared to designing and undertaking a new RCT, such studies on patient preferences are quite feasible, and relatively inexpensive.

A suitable study design to develop a decision aid is a qualitative research study that uses focus groups of patients, relatives of patients, and physicians. Based on key informant input, there are no decision aids in routine clinical use. Therefore, the question of evaluating decision aims may be premature.

Assessing Performance (Fourth Thematic Area)

The fourth thematic area for future research is development of evidence-based performance measures to be used as feedback to health care facilities or practitioners towards improving patient-relevant outcomes. Evidence suggests that establishing performance measures and active monitoring of physician performance has positive impact on the quality of delivered health care. In principle, it should be feasible to identify process-based measures of performance in cardiovascular care, given the existence of multiple interventions with strong support from RCTs in the domain of cardiovascular disease. However, for most chronic diseases, including CAD, it is not possible to use simple "all-or-nothing" measures to quantify optimal care; simple measures are limited to quantifying poor care. Therefore, developing effective evidence-based performance measures will require both qualitative and quantitative research to identify appropriate measures and test their impact on clinical outcomes.

A first step could be the establishment of focus groups of scientists with relevant expertise to propose specific measures based on the current evidence base and analyses of administrative data. Health care professionals as well as participants with expertise in bioinformatics, quality control, or operations research would offer complementary expertise. To inform these focus groups it may be important to conduct analyses of administrative data related to the processes of interest.

The actual evaluation of whether the implementation of a performance monitoring and feedback system could be readily performed in observational studies that measure performance before and after the implementing the monitoring and feedback mechanism. A much stronger design would be a cluster RCT, where one would randomize health care facilities to implement versus not implement the system. However, unless there is substantial infrastructure already in place, such a cluster RCT would not be easy to perform.

Table II. Prioritized research designs to address future research needs when studying the comparative effectiveness of PCI and CABG

Thematic area	#	*Population*	*Intervention*	*Comparator*	• ***Primary outcomes*** • ***Secondary outcomes***	*Design*	*Feasible*	*Research cost*
(A) Comparative effectiveness and safety	1	General population of elderly patients >75 years old	On-pump CABG	PCI with DES (or with DES and BMS)	• Total mortality or composite of total mortality or myocardial infarctions • Other objective and subjective outcomes	MIPD of RCT or registry data	Yes	Low
						Prospective comparative observational study, preferably nested in an existing cohort or registry	Yes	Medium
						RCT	Probably	High
	2*	Heart failure	[As above]	[As above]	• Composite of total mortality or myocardial infarctions† • Other objective and subjective outcomes	[same options as in row A1]	[same options as in row A1]	[same options as in row A1]
	3*	Renal disease stage 3 or 4	[As above]	[As above]	[As above]	[same options as in row A1]	[same options as in row A1]	[same options as in row A1]
	4	Diabetes	[As above]	[As above]	[As above]	[same options as in row A1]	[same options as in row A1]	[same options as in row A1]
(B) Testing to predict treatment response‡	1	General population of revascularized patients	Invasive and noninvasive tests	Not applicable	Predictive sensitivity or specificity or related metrics	Retrospective analysis of RCT-based data	Yes	Low
						Prospective cohort	Yes	Medium
(C) Enhancing patient participation‡	1	General population of revascularized patients	Not applicable	Not applicable	Description of patient preferences	Qualitative research	Yes	Low
						Survey	Yes	Low
	2	General population of revascularized patients	Not applicable	Not applicable	Development of decision support tools	Qualitative research	Yes	Low or medium
					Evaluation of decision support tools	RCT	Unclear	Unclear
(D) Assessing performance‡	1	General population of revascularized patients	Not applicable	Not applicable	Development of evidence-based performance measures	Qualitative research; observational studies	Unclear	Unclear
					Evaluation of systems that monitor performance	Before-after observational studies or cluster RCT	Unclear	Unclear

BMS=bare metal stent; CABG=coronary artery bypass graft surgery; DES=drug eluting stent; MIPD=meta-analysis of individual patient data; RCT=randomized controlled trial.

* We considered heart failure and renal disease as equally pressing needs.

† So that power calculations result in a feasible trial (<2500 sample size).

‡ A thorough evaluation of the literature for these research gaps was not available (they were not included in the Stanford CER).

Conclusions

Based on our review of the Stanford CER, input from key informants, newly published and ongoing studies, and our insights from quantitative analyses we identified four thematic areas of future research needs to inform choice between mechanical revascularization procedures in patients with nonacute CAD:[a]

A. Comparative effectiveness and safety of PCI vs. CABG: In the first area, pressing priorities are studies on the comparative effectiveness and safety of DES[b] vs. CABG in elderly patients (older than 75 years); patients with heart failure or patients with stage 3 or 4 chronic kidney disease; and patients with diabetes. These patient subgroups are often underrepresented in RCTs but are frequently encountered in everyday clinical practice. Further, for the elderly and for diabetics there is suggestive evidence of a subpopulation-by-treatment interaction. An efficient way to address this research need is to capitalize on already existing data by performing meta-analyses of individual participant data from existing RCTs or from large observational studies. An alternative approach is to perform *de novo* RCTs, especially if existing data are not applicable to the population of interest, or suggest a subpopulation-by-treatment interaction that is clinically important.[c]

B. The role of testing to inform choice of revascularization procedure: In the second area the priority is to perform studies to assess the ability of invasive (arteriography) and noninvasive tests (MR or CT angiography, resting or exercise SPECT, exercise treadmill testing with or without echocardiography) to predict differential response to PCI or CABG. There are many different test-and-treat strategies that can be conceived, and it is not possible to compare all of them with randomized studies. A practical recommendation for future research is to explore the predictive accuracy of single tests or combinations thereof by performing cohort studies, case control studies, or by reanalyzing baseline data from existing RCTs (if available). To be clinically useful and informative, such studies should enroll patients representative of those seen in clinical practice, and take steps to limit verification bias and other biases.

C. Enhancing patient participation: In the third area the identified research needs pertain to eliciting and measuring patient preferences and facilitating shared decisionmaking by developing and evaluating decision aids. A suitable study design to elicit patient preferences is a survey of patients, or a qualitative research study that uses focus groups of patients. Decision aids can be developed in qualitative research studies that utilize focus groups of patients, relatives of patients, and physicians. Based on key informant input, there are no decision aids in routine clinical use. Therefore, the question of evaluating decision aims may be premature.

D. Assessing performance: Development of evidence-based performance measures, monitoring of practice based on these measures and providing feedback to health care facilities and practitioners can improve revascularization outcomes and may reduce inequalities in provided health care at the national level. The optimal research design to develop reliable process-based measures of performance is not clear. One option is to

[a] Because we relied only on key informant input without having the benefit of a thorough evaluation of the literature for the last three thematic areas, and it is possible that the identified evidence gaps have been addressed.

[b] DES in particular. BMS still represents an important comparator.

[c] Alternatively, one could perform efficient *de novo* nonrandomized comparative studies and proceed to a *de novo* RCT only if there is suggestive evidence of an important treatment-by-subgroup interaction.

perform qualitative research using focus groups of health care professionals as well as participants with expertise in quality control or operations research. Evaluating process-based performance measures would be typically performed in large-scope studies that analyze administrative data before and after the implementation of performance monitoring and feedback systems.

Background

Percutaneous coronary interventions (PCI) with or without stents and coronary artery bypass graft surgery (CABG) are the two broad categories of interventions for mechanical revascularization of atherosclerotic coronary arteries in patients with coronary artery disease (CAD). The choice of revascularization procedure is often determined by coronary anatomy and the technical feasibility of either PCI or CABG. Therefore the clinical dilemma pertains to patients in whom both procedures are technically feasible and whose coronary disease is neither too limited nor too extensive. Generally, both approaches would be clinically relevant for patients with single-vessel disease of the proximal left anterior descending artery, most types of two-vessel disease, as well as patients with three-vessel disease that is not particularly extensive. Because PCI and CABG differ in their procedural risk and their initial and downstream costs, assessing their comparative effectiveness and safety is of great interest.

PCI and CABG have already been compared in several randomized controlled trials (RCTs)[1-7] and analyses of large clinical registries.[8-11] The aim of this report is to identify needs for future research in the comparison between PCI and CABG. We use as a basis the 2007 comparative effectiveness report (CER) by the Stanford–University of California at San Francisco Evidence-based Practice Center (Stanford–UCSF EPC) that assessed PCI vs. CABG for coronary artery disease (for brevity, we hereafter refer to the report as "Stanford CER").[12] The report summarized evidence published through 2006 to address the following key questions:

Key Question 1a: In patients with ischemic heart disease and angiographically proven single- or multiple-vessel disease, what is the comparative effectiveness of PCI compared to CABG in reducing the occurrence of adverse objective outcomes[a] and improving subjective outcomes?[b]

Key Question 1b: Over what period are the comparative benefits of PCI and CABG sustained?

Key Question 2: Is there evidence that the comparative effectiveness of PCI and CABG varies based on (predefined) patient and procedure related factors?[c]

---

[a] Long-term and short-term objective outcomes referred to outcomes that impacted patients' health, including, but not limited to, periprocedural death or complications, non-fatal myocardial infarctions, congestive heart failure, stroke, nosocomial infections, respiratory failure or other pulmonary complications, acute or chronic renal failure, cardiac arrhythmias, and long-term survival and event-free survival (major adverse cardiac events).

[b] Subjective outcomes referred to outcomes that impacted patients' perceived quality of life, functional health status, or general health status. Subjective outcomes included freedom from angina and quality of life.

[c] Demographic factors (age, sex, race, or other demographic risk factors), comorbidities (coronary disease risk factors, diabetes, obesity or other comorbid disease), angiographic factors (extent of disease—single and multi-vessel disease), left ventricular function, PCI/CABG-specific factors (bare metal stents, balloon angioplasty, and drug-eluting stents, use of minimally invasive techniques, use of internal mammary arteries), clinical presentation (stable angina or unstable angina, based on NYHA functional class I-IV).

Findings of the Stanford CER

This is a distillation of the findings of the Stanford CER on Key Questions 1a and b:

1. Procedural (short-term) outcomes:
    a. Procedural survival was high for both interventions and was not significantly different between PCI and CABG.
    b. Freedom from procedural strokes significantly favored PCI compared with CABG.
    c. Freedom from procedural myocardial infarction was heterogeneously defined and was not significantly different between the two treatments.
2. Long-term objective outcomes:
    a. Long-term survival was significantly better for CABG compared to PCI in the older trials (pre 1994, where PCI procedures mostly were balloon angioplasties). However, long term survival was not significantly different between CABG and PCI in more recent trials (stent era).
    b. Similarly, long term freedom from myocardial infarction was not statistically significant between the compared treatments.
3. Long-term subjective outcomes
    a. Freedom from angina was greater for CABG rather than PCI interventions.
    b. Quality of life outcomes favored CABG over PCI between 6 months and 3 years of follow-up but equalized thereafter. The degree of improvement in quality of life was correlated with relief of angina.

Regarding Key Question 2, the Stanford CER evaluated the evidence for comparative effectiveness of PCI and CABG based on (predefined) patient-, provider-, disease- and intervention-related factors. However, the report noted that in contrast to the fairly robust evidence concerning overall clinical outcomes, there was much less evidence from randomized trials to gauge whether and how comparative effectiveness varied across these factors. Most clinical trials reported only on survival and did not report outcomes in key patient subgroups.

Currency of the Stanford Report

The Stanford CER summarized evidence up to 2006 and includes most large RCTs that are available at this point. Therefore it is, as of this writing, a relevant report. However, some developments that were outside the scope of the original CER or occurred after the report was completed, need to be considered.

First, medical treatment was outside the scope of the Stanford CER. Therefore, the CER is applicable only to patients in whom revascularization is warranted. However, no revascularization (optimal medical therapy) is now considered a valid clinical option in patients with stable CAD, particularly after the publication of the COURAGE RCT.[13] COURAGE (n=2287) did not find significant differences in mortality, myocardial infarction and other outcomes between PCI (mainly with bare metal stents, BMS) and optimal medical therapy (including aggressive cholesterol lowering therapy, beta blockers, angiotensin converting enzyme inhibitors, and antiplatelets when applicable). A subsequent network meta-analysis of COURAGE and several other RCTs did not find significant differences between medical therapy and PCI with BMS, or medical therapy and PCI with drug-eluting stents (DES).[14] This makes it difficult to clearly specify the pool of CAD patients to whom the Stanford CER pertains, as there is randomized evidence that a large number of them can be treated without revascularization.

Further, additional evidence has been published since the completion of the Stanford CER, and this could affect the conclusions of the report on the comparative effectiveness and safety of PCI vs. CABG, as well as its recommendations for future research:

- Several RCTs that had been included in the Stanford CER have presented long term outcome data based on longer periods of followup.
- Several new RCTs have published results after the completion of the Stanford CER. Among them the SYNTAX trial,[15] a large RCT comparing DES with CABG in patients with three-vessel or complex left main CAD, added a considerable amount of evidence for this comparison. In addition, the CARDia trial,[16] a randomized trial in diabetic patients with multivessel CAD or complex single vessel disease informs the comparison of PCI with stents (68% DES) with CABG in this subgroup.
- Several ongoing RCTs are expected to report results in the next 2 or 3 years.

Research Gaps Identified in the Stanford CER

Table 1 summarizes research gaps and proposals for future studies identified by the Stanford CER. The Stanford CER did not describe a specific methodology for identifying the research gaps, for suggesting what the future research needs are or how they should be addressed. We have mapped the research gaps to the key questions in the CER's analytic framework.

Table 1. Research gaps and proposals for future research from the Stanford CER

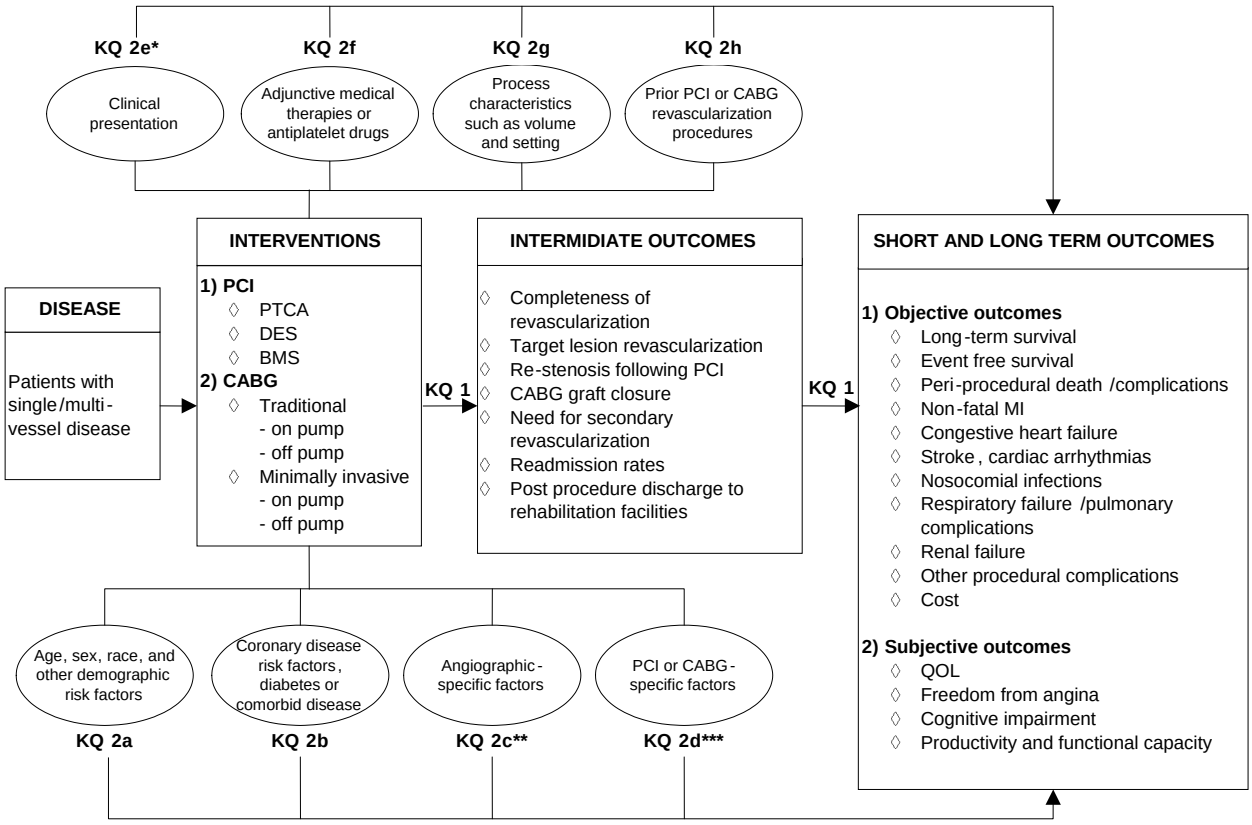
Research gap	Suggestions for future research	Corresponding Key Question
Lack of analyses of PCI and CABG outcomes according to patient characteristics. Specifically: • Sex • Age • Chronic kidney disease • Left ventricular dysfunction	Collaborative pooling of individual patient-level data from randomized trials to (a) enhance statistical power and (b) reduce publication bias A more extensive collaborative study to pool individual patient data from both balloon-era and stent-era trials to increase the number of patients and outcome events improving statistical power even further in patient subpopulations. More direct assessments of the impact of stents on the comparative effectiveness of PCI and CABG and assessment of whether relative efficacy changes over extended followup.	KQ 1, KQ 2a and 2b
Paucity of data on the comparative effectiveness and safety of DES in particular, and especially in the long term.	Further clinical trials with extended followup and large enough to detect clinically meaningful differences in outcome to assess whether the availability of DES has affected the comparative efficacy of PCI and CABG Because the procedural risk of CABG in large registries has declined progressively over time, several trials comparing contemporary CABG with PCI using DES were anticipated (including the FREEDOM (NCT00086540) and SYNTAX trials (NCT 00114972)]	KQ 1, KQ 2d
Paucity of data on the relationship between procedural volume and outcomes of minimally invasive approaches to CABG.	Further research on the association of procedure volume with outcome to examine additional outcome measures, both short term (e.g., nonfatal myocardial infarction, completeness of revascularization) and long term (e.g., survival, angina relief, freedom from repeat procedures), preferably in large patient cohorts using contemporaneous CABG and PCI and applying the same analytic methods.	KQ 2d, KQ 2g

Paucity of data on metrics for quality of care for PCI and CABG procedures.	Development of evidence-based process measures for PCI and CABG to facilitate efforts to improve quality of care and provide better performance measures than procedure volume.

CABG, coronary artery bypass graft surgery; DES, drug eluting stents; PCI, percutaneous coronary intervention.

The analytic framework in Figure 1 illustrates the key questions of the Stanford CER, and maps the evidence gaps that were identified.[a]

Figure 1. Reconstructed analytic framework showing the evidence gaps identified by the Stanford CER



*** KQ 2e**: Stable or unstable angina, based on NYHA functional class I-IV, acute coronary syndrome, cardiogenic shock, acute myocardial infarction with or without ST elevation, or silent ischemia.
**** KQ 2c**: The number of diseased vessels amenable to bypass or stenting, vessel territory of stenosis (e.g., left main or anterior coronary arteries, right coronary artery, circumflex coronary artery), diffuse versus. focal stenoses, left ventricular function, or prior revascularization procedures.
***** KQ 2d**: For CABG: cardiopulmonary bypass mode (normothermic vs. hypothermic), type of cardioplegia used (blood vs. crystalloid), or use of saphenous vein grafts, single or bilateral internal mammory artery grafts, or other types of bypass grafts. For PCI: use of balloon angioplasty, bare metal of drug eluting stents.

[a] This analytic framework is not included in the Stanford CER. It was constructed by Tufts EPC to facilitate the description of the evidence gaps that were identified in the Stanford CER. These evidence gaps guided the future research recommendations section of the report.

Methods

This Future Research Needs document is intended to inform and support researchers and those who fund research to ultimately enhance the body of comparative effectiveness evidence so that it is useful for decisionmakers. The current document is a pilot project undertaken by the Tufts EPC that emulates the envisioned process for developing Future Research Needs documents: after the completion of a CER, EPCs would prioritize the identified *evidence gaps* into *research needs*, and suggest potential research designs. The premise is that not all evidence gaps are of equal importance, and because resources are finite, future research should address the most important among the answerable evidence gaps using the most efficient research methods.

By their very nature Future Research Needs documents refer to the same clinical context as the CER they are based on. This means that they pertain to similar patient populations, interventions and comparators. For the current pilot we refer to patients with nonacute CAD in whom revascularization is warranted; the intervention is PCI variants and the comparator is CABG variants.

Management of coronary artery disease without revascularization (e.g., only with optimal medical therapy) is a valid clinical option in patients with stable CAD, but it was not included in the Stanford CER and is not considered in this document. We comment on the implications of this decision in the Discussion Section.

Overview of the Stepwise Approach

We sought a practical approach to identifying evidence gaps and prioritizing them into research needs. We used a continuum of qualitative and quantitative methodologies, as outlined below. In the following paragraphs we provide details on our methods, and our rationale.

1. *Initial list of evidence gaps (List 1).* We generated an initial list of evidence gaps based on the Stanford CER and initial feedback from a group of key informants.
2. *Interim expanded list of evidence gaps (List 2).* We searched the literature to identify trials that have been published after the Stanford CER. We also searched ClinicalTrials.gov to identify ongoing studies in the field. We then proceeded to further discuss the initial list of evidence gaps (List 1) in one-to-one interviews with key informants taking into account information from the literature searches. This resulted in an interim expanded list of evidence gaps, which included more potential target topics for future research compared to the initial list.
3. *Pruned (final) list of important evidence gaps (List 3).* Here we decided to prune the length of the interim list of evidence gaps (List 2) to contain its scope. This step was not anticipated in the beginning of the pilot project but it was considered necessary in view of the expansion of the evidence gaps list at the prior steps. We pruned the list of the previous step internally and according to *post hoc* criteria, but without direct and explicit feedback from the key informants. We took into consideration data from the literature and the inputs of key informants in the previous steps.
4. *Recommendations for future research*. We recommended specific research designs for future research. To this end we used *ad hoc* criteria to assess the importance of each evidence gap in List 3 to the US setting. Further we estimated the feasibility of different research designs, asked for additional key informant feedback, and performed focused

modeling analyses to pinpoint for which parameters it is most important to obtain additional information through future research.

Step 1. Generating the Initial List of Evidence Gaps

We formed a group of key informants to help us better understand the evidence gaps identified in the Stanford CER (Table 1) and how they relate to clinical practice and the current trends in the field.

Identification of Key Informants and First Round of Key Informant Feedback

We identified six key informants, namely a medical officer of a funding agency for cardiovascular research, a medical director in a major payor, an interventional cardiologist, a CABG surgeon, a general cardiologist, and a clinical researcher and methodologist with contributions to the topic who was also an author of the Stanford CER. We did not use a particular method to identify key informants, such as random sampling from a large pool of candidates. A patient representative was not included because we deemed that the patient perspective is peripheral to this exercise. We elaborate on our decision not to include a patient representative in the Discussion Section.

We did not use a formal process such as a Delphi process to elicit and compile input from key informants. Instead, key informants provided initial feedback via a teleconference. In the teleconference we started discussing which subpopulations, interventions, comparators and outcomes represent evidence gaps based on the Stanford CER. The teleconference did not exhaust the discussion. Based on the feedback from the teleconference we generated the initial list of evidence gaps (List 1, described in the Results Section and in Appendix B).

Step 2. Generating the Expanded List of Evidence Gaps

We then proceeded to further refine the list of evidence gaps. To inform additional discussions with the key informants, we searched for relevant, recently published or ongoing studies that were not included in the Stanford CER.

Trials Published After the Stanford CER and Additional Ongoing Studies

The Stanford CER was published in 2007, and reviewed evidence through 2006. To assess its currency and whether any of the identified evidence gaps have been addressed in the meanwhile, we searched PubMed for randomized controlled trial reports between 2006 and 2010 (last search August 4, 2010). We also searched ClinicalTrials.gov to identify registered ongoing interventional (randomized or nonrandomized) or observational studies (e.g., prospective cohorts) that included patients with nonacute CAD and could inform the comparison of PCI with CABG. Our search strategies are listed in Appendix A.

Second Round of Key Informant Feedback

We then invited key informants to participate in one-to-one teleconferences to continue the discussion of the evidence gaps also considering recent evidence and ongoing studies. Four of the six key informants participated in one-to-one interviews. The interviews aimed to identify the key informants' perception of the major questions in the field, and if applicable, their rationale on why specific questions may be more important than others. Key informants suggested additional areas for future research that were related to the comparison of PCI vs.

CABG, but were not in the scope of the Stanford CER. This resulted in an interim expanded list of evidence gaps (List 2 in Appendix B), which included additional potential target topics for future research. Finally, key informants provided feedback on criteria to prioritize specific research designs (to be used in Step 4).

Step 3. Generating the Final List of Important Evidence Gaps

The expanded list of evidence gaps was too long to effectively develop research designs for, so we eliminated less important gaps based on data from the literature and inputs from the key informants provided during the previous steps.

First, we classified the gaps in the interim expanded list into four thematic areas of future research. The four thematic areas are not overlapping, could be pursued "independently," and are amenable to different types of study designs (and therefore do not all require extensive resources to pursue). The four thematic areas were:

A. Comparative effectiveness and safety of PCI vs. CABG
B. The role of testing to inform choice of revascularization procedure
C. Enhancing patient participation
D. Assessing performance

Only the first thematic area was directly related to the scope of the Stanford CER. The remaining were added based on key informant input in the previous steps.

We developed the *final list of important evidence gaps* (List 3 in Appendix B) by pruning the interim expanded list (List 2) within each of the four thematic areas using *post hoc* criteria, as described below.

Most Important Evidence Gaps in Comparative Treatment Effectiveness and Safety (First Thematic Area)

We considered as more pressing research gaps that pertained to subpopulations representing a high "disease burden" among patients with nonacute CAD in the US. We estimated the frequency of subpopulations representing evidence gaps in 4 analyses of large clinical registries[a] (Appendix Table B3)[8-11] as a crude proxy of the corresponding "disease burden." We selected the 6 subpopulations with the largest percentage in the publication by Hannan et al.,[8] which is based on two comprehensive registries from the state of New York, and had the largest sample size. We favored interventions that are in routine use in current clinical practice, and that are likely to be routinely used in the midterm future (i.e., for the following 5 years). We deemed that any future research should collect as much information on outcomes as practically feasible.

Evidence Gaps in the Role of Testing To Inform Choice of Revascularization Procedure (Second Thematic Area)

Based on key informant input from previous steps we distinguished evidence gaps related to the ability of testing to predict patient response to treatment with PCI or CABG. Testing could therefore have an effect on patient outcomes, in that it would affect treatment decisions, which in

[a] These were selected among those identified in the Stanford CER using the following criteria: The analysis included patients revascularized in the 1990's in North America or Europe, described at least 1000 patients treated with PCI and 1000 treated with CABG, reported patient characteristics and performed multivariate statistical analyses.

turn affect patient outcomes.[a] The Stanford CER did not review testing to predict revascularization outcomes, and therefore we did not have a systematic overview the state of the evidence available. Instead of explicitly selecting individual tests or combinations of tests for further study, we considered an index list of 4 types of pretreatment testing (invasive and noninvasive) in the following step.[b]

Evidence Gaps in Enhancing Patient Participation (Third Thematic Area)

The third thematic area of evidence gaps pertained to understanding patient preferences and facilitating shared decisionmaking between patients and their physicians. This area of research was not in the scope of the Stanford CER. We selected two particular gaps that were amenable to observational or experimental study, were indicated as important by the key informants, and covered the range of key informant comments on enhancing patients' voice.

Evidence Gaps in Assessing Performance (Fourth Thematic Area)

Finally, based on prior key informant input, there are no validated process-based performance measures that could quantify optimal care. The potential payoff of validated performance measures at the level of a health system is substantial. Therefore, we considered this as an important evidence gap to be addressed through future research.

Step 4. Making Recommendations for Future Research

We proposed specific research designs to address the evidence gaps in the pruned (final) list of the previous step. We considered the four thematic areas of future research separately; in principle, the evidence gaps in the four thematic areas are nonoverlapping and can be pursued independently.

We prioritized research needs based on the evidence gaps in the pruned final list (List 3) and considering predefined prioritization criteria. We solicited explicit input from key informants on the relative importance of the evidence gaps in the final list, and obtained insights from modeling analyses. We considered the following *ad hoc* criteria to prioritize research designs: feasibility in terms of research costs, feasibility in terms of projected study duration, likelihood that the study will have nontrivial findings,[c] likelihood that the study will provide unbiased results (to inform clinical practice), and likelihood that ongoing research will address the evidence gap.

Third Round of Key Informant Input

All key informants were asked to rank the evidence gaps in the pruned list and provide any additional feedback via an e-mailed questionnaire. The key informants were asked to consider the qualitative criteria of Appendix Table B4 in their ranking. We did not automatically

[a] Here we are interested in testing to guide treatment choice (via predicting differential response to treatment). We do not refer to other settings such as screening of asymptomatic individuals, diagnosis, or treatment monitoring. In particular screening of asymptomatic individuals was identified as a very interesting research area, but it was deemed to be outside the scope of our exercise.

[b] Invasive testing with coronary arteriography, non-invasive computerized tomography angiography (CTA) or magnetic resonance angiography (MRA), resting or exercise single photon emission computerized tomography (SPECT), and noninvasive exercise treadmill testing with or without echocardiography.

[c] If there is prior information from large scale analyses of randomized data suggesting, e.g. a treatment effect modification.

accept the suggestions of the key informants, but considered it in our prioritization of research designs together with the aforementioned criteria.

Focused Modeling To Identify Important Parameters To Address in Future Research

One would perform further research in topics where decisions have to be made, but there is substantial remaining uncertainty on parameters that can affect these decisions. For example, consider the choice between PCI and CABG among elderly patients.[a] For which parameters is it most important to obtain additional information through future research? Assuming a reasonably simple specification of this decisional problem, such parameters may be the prevalence of procedural deaths or procedural strokes; the relative effects for procedural deaths or strokes across the compared interventions; the frequency of long term deaths, myocardial infarctions, strokes, or repeat revascularizations; the corresponding relative effects across revascularization options; patient preferences (utilities); and immediate and downstream health care costs. This has practical implications, because different parameters are naturally amenable to different research designs.[b]

To gain insights on important parameters we developed simple mathematical models to analyze the choice between PCI and CABG. Details on our quantitative approaches (modeling strategy, assumptions, data sources, statistical analyses) are presented in the Methods Section of Appendix D. Briefly, we followed an operational process to develop simple decision models that compare different treatment options (PCI with BMS, PCI with DES, CABG). We used Markov models with the following health states: (1) asymptomatic (no prior stroke), (2) recent myocardial infarction (no prior stroke), (3) asymptomatic (post stroke), (4) recent myocardial infarction (post stroke), (5) repeat stroke in patients who had had a prior stroke; (6) dead (Appendix Figure D1). Estimates for model parameters were derived from published sources, including the Stanford CER,[12] a subsequent meta-analysis of individual patient data from RCTs;[17] a network meta-analysis of DES and BMS,[14] three large recently published randomized trials (CARDia,[16] COURAGE,[13] and SYNTAX[15]) and three cost-effectiveness analyses.[18-20] The previous studies were identified from key informant input or from focused literature searches. We used network meta-analysis to obtain consistent estimates for all treatment effects.[21] We analyzed the models using a time horizon of 10 years, as this is our time horizon for making recommendations for future research.

Specification of Three Index Models

It was not feasible to perform quantitative analyses for all identified subpopulations in the pruned final list of research gaps, because we would have to develop separate models for each one.[c] Instead, we decided to develop models for three index subpopulations and extrapolate any

[a] Actually, this is one of the important evidence gaps in the first thematic area (comparative effectiveness of PCI vs. CABG).

[b] To better define prevalence of procedural or long term events one may opt to analyze registries or perform observational studies. To get more information on treatment effect modification one would have to reanalyze existing comparative data or perform new RCTs. Other research designs would be more suitable to other parameters (see subsequent paragraph on "Candidate research designs").

[c] Initially we envisioned that in our discussions with the Key Informants we would identify only one or two important subpopulations in which we would perform modeling analyses. However, this was not the case.

insights to the rest. We chose the index subpopulations based on ease of modeling and availability of good data to parameterize the decision models:

- "RCT-type participants:" a reference scenario simulating a cohort of 65 year old patients with nonacute CAD and no major comorbidities. The decision is between revascularization with DES, BMS or CABG. Most data to parameterize the models were obtained from RCTs. This model is based on more robust data than the other two models (Appendix D).
- "Elderly participants (older than 75 years):" a cohort of elderly (75 year old) CAD patients with nonacute CAD. The decision is between revascularization with DES, BMS and CABG. Because there is not a lot of information on elderly patients specifically, this model makes extrapolations based on specific assumptions (Appendix D).
- "Diabetics:" a cohort of 65-year old diabetic patients. The decision is between revascularization with DES and CABG. As was the case with the model on the elderly, we make several assumptions to parameterize this model (Appendix D).

Modeling Analyses

We performed three types of quantitative analyses:

1. We analyzed our deterministic decision models using quality-adjusted life years (QALY) as the decision relevant quantity. Our aim was to identify "influential" parameters, i.e., parameters that exert the maximum influence on the decision relevant quantity in one-way sensitivity analyses.[a] Theoretically, those parameters should be considered as research priorities, since reducing the uncertainty around them would have the biggest effect on the decision uncertainty.[b]
2. We then included costs and performed cost-effectiveness analyses. The decision relevant quantities were incremental cost-effectiveness ratios between treatment pairs. Again we identified influential parameters using one-way sensitivity analyses. As in the decision analysis approach, influential parameters are more likely to represent priority research needs compared to less influential ones.
3. Finally, we recast the models as probabilistic models and performed value of information analyses. First we calculated the expected value of perfect information (EVPI). EVPI is expressed in monetary units and represents the opportunity cost incurred by having to make decisions based on imperfect information. EVPI represents an *upper bound* to the expected returns of future research. EVPI can be considered as the value of reducing uncertainty for the overall decision tree, i.e. it can be considered as indicative of the value of future research on a broad field, but it cannot prioritize specific research topics or guide study designs. Then we calculated the expected value of perfect information for groups of parameters (EVPPI), which places an upper bound to the value of research *on specific (groups of) parameters*. EVPPI estimates for the specific (groups of) parameters can be used to select the specific topics that future research should address. In EVPPI analyses we organized groups of parameters of interest so that each group could be addressed by a single future study (Appendix D).

[a] In one-way sensitivity analyses we change the value of each parameter over a prespecified uncertainty range (its corresponding 95% confidence interval) while keeping all other parameters at their baseline values, and record the effect on the decision relevant quantity.

[b] This is a simplification. One-way sensitivity analyses underestimate the uncertainty inherent in the model and cannot handle correlated parameters.

Interpretation of Modeling Results

We perform modeling analyses to only to identify "influential parameters" and to rank them according to their relative "influence." This is an *atypical use of modeling*, in that we stop short of providing insights on treatment choices for different assumptions and circumstances. However this is a conscious choice:[a] Ranking of influential parameters is likely to remain stable even if more elaborate models are used, or if better data are used to populate parameters for which we made simplifying assumptions. In contrast, exact values may change substantially.

Candidate Study Designs

Candidate study designs will differ across types of research needs. Effectiveness or efficacy of treatments can be most definitively addressed in RCTs, and secondarily in well conducted nonrandomized comparative observational studies. In contrast, eliciting patient preferences can be meaningfully performed with nonexperimental designs (e.g., in a survey). We list the candidate study designs for different types of questions of future research (Table 2).

Broadly speaking and without considering other factors such as feasibility, an RCT is the most suitable study design to *obtain unbiased estimates of effectiveness or efficacy* of specific interventions in specific populations (other options are listed in Table 2). While one could suggest using RCTs to compare various test-and-treat strategies for patient management, this is often not possible.[22] The first stage should be to *evaluate the performance of tests in predicting differential response to the treatments of interest*, and decide whether an RCT is necessary in a second step (see Results Section).[22] This can be achieved by modeling the treatment-by-test-results interaction in patients who received the treatments of interest and have been followed up for sufficient time to observe their response. An attractive design is to "nest" the study of predictive performance of testing in an RCT that compares the interventions of interest, by applying the test at baseline. An alternative is to reanalyze data of existing RCTs, provided that test results are available for all participants at baseline (or are missing at random in a minority of RCT participants). Other options are listed in the Table. Finally, RCTs are not appropriate for eliciting patient preferences, or for developing decision support tools or performance measures. Surveys of patients or qualitative research studies are possible study designs for these latter cases, as discussed in the Results Section.

Table 2. Candidate study designs for addressing different types of research needs

Type of question to be addressed	*General design category*	*New data*
Treatment effects of interventions (efficacy, effectiveness or safety)	Randomized controlled trial	Yes
	Nonrandomized comparative trial	Yes
	Observational studies—meta-analysis of individual patient data from RCTs	No
	Observational studies—analysis of large clinical registry	No
Predictive performance of tests (predictive sensitivity, specificity or related quantities)	Observational studies—Reanalysis of RCT data	No
	Observational studies—Prospective nonrandomized comparative trial	Yes
	Observational studies—Reanalysis of existing data from nonrandomized comparative studies	No

[a] This is admittedly a defensive stance. It is very influenced by the fact that our models are not developed with the same rigor as some of the well-known and elaborate models that are used to analyze clinical decisions; they have not been calibrated using external data; and their predictions have not been validated in external data. They are operational models that are constructed to make use of the summary information obtained from an evidence report, that can still offer broad insights.

Type of question to be addressed	*General design category*	*New data*
	Observational studies—Retrospective nonrandomized comparative trial	Yes
	Observational studies—Case control	Yes
Elicitation of patient preferences and development/evaluation of decision support tools	Observational studies—Survey	Yes
	Observational studies—Various[a]	Yes
Development of performance measures	Observational studies—Various[b]	Yes

Ordered within each thematic area according to need for novel data collection. Within each area studies are (broadly) ordered by their internal validity.

Feasibility of Study Designs

Studies that do not require new data collection are in principle feasible, provided that access to existing data can be agreed upon or has already been granted. An analysis of an existing registry can be completed within a year. A meta-analysis of individual patient data can be conducted in a time horizon of two years.[c] The feasibility of such studies, generally, does not depend on the desired sample size.

We considered that a study of primary data collection would be infeasible if it were too expensive or complex to conduct; if it required too long a followup, say beyond 5–7 years; or if it relied on information or data that is not yet available. We acknowledge that deliberations on feasibility are inherently subjective.

Generally, RCTs are among the most expensive research designs. Recently completed or ongoing large efficacy RCTs may be examples of "expensive" research. Some of the largest recent efficacy RCTs in nonacute CAD have sample sizes in the neighborhood of 2500 patients.[d] Using this as a reference we commented on the feasibility of other research designs based on the following assumptions:

- Prospective nonrandomized comparative trials are less expensive to conduct than efficacy RCTs of similar size.
- Cohorts or case control studies of the predictive performance of tests for response to treatments are substantially less expensive to perform than efficacy RCTs of similar size.
- Surveys of, e.g., patients to elicit preferences should be generally economically feasible, as they are expected to be much cheaper than a large RCT of a few thousand people.

[a] A variety of research designs may be pertinent, ranging from qualitative research in focus groups, to randomized trials of using versus not using the decision support aid. We do not expand here, but discuss specifics in the Results and Discussion sections.

[b] A variety of research designs may be pertinent to developing performance measures. We do not expand here, but discuss specifics in the Results and Discussion sections.

[c] A meta-analysis of individual patient data can take longer to complete than an analysis of an existing and available database. There are logistical complications including but not limited to identification of data sources, convincing investigators to participate, standardizing definitions of interventions and outcomes, complying with HIPAA, and harmonizing datasets.

[d] For example COURAGE[13] (NCT00007657) compared optimal medical therapy and PCI in 2,287 patients, FREEDOM[22] (NCT00086540) compares PCI and CABG in approximately 2,400 patients, and STICH[23,24] (NCT00023595) compared CABG and medical treatment in 2,136 patients.

Sample Size Calculations for RCTs

We performed sample size calculations using standard formulae for a two-sided chi-squared test at the 0.05 level of significance. We assumed a true relative effect of 0.80 favoring the intervention arm, an allocation ratio of 1:1, no loss to followup, no crossover between treatments, and no sequential monitoring. We made power calculations for 3 and 5 years of followup assuming a range of constant annual event rates in the comparator intervention corresponding to 5-year cumulative proportion of primary events at 5, 10, 15, 20, 30 or 40 percent. To estimate the duration of a trial so that the mean followup is 5 years we assumed a minimum follow up of 2.5 years, an accrual period of 5 years and a constant accrual rate throughout the accrual period. Because of our simplifying assumptions, we probably overestimate the power attained at various total sample sizes.

Handling Conflicts of Interest

In order to minimize conflicts of interest we suggested proposed research designs internally, using predetermined criteria and incorporating insights from modeling analyses. Key informants, all of whom were screened for potential conflicts of interest were consulted to ensure that important evidence gaps were considered and to identify criteria for prioritization, but they were not directly involved in the final prioritization.

Results

We first describe the results of our process to generate the final list of research gaps (Steps 1 through 3). Subsequently, we describe the results of our prioritization of future research needs (Step 4), and suggest specific research designs.

Step 1. Generating the Initial List of Evidence Gaps

The evidence gaps identified by the Stanford CER are listed in Table 1. Based on key informant input via a teleconference, we generated the initial list of evidence gaps (List 1 in Appendix B). Briefly, key informants underlined the importance of the subpopulations of elderly patients (older than 71 or even older than 75 years), women (who tend to have smaller body size), and subpopulations defined by racial descent (particularly Asian, who tend to have smaller body size), and diabetes (major comorbidity that needs further research).

They also added to the list of evidence gaps the subpopulations of patients with high body mass index (above 30kg/m^2), heart failure (as ejection fraction less than 35 percent), and chronic renal disease prior to dialysis (stages 3 or 4). The rationale was that obese patients often have other comorbidities and can pose technical challenges to both revascularization options, and that both heart failure and chronic kidney disease may interact with treatment effects based on CAD pathophysiology.

In addition, key informants suggested that that there is little evidence on patients' preferences regarding the revascularization procedures themselves, or the downstream outcomes with each procedure. For example, some patients may prefer the risk of a major surgery if they have a better chance to be relieved of anginal symptoms, even if they know that there is no evidence of difference in long term survival between CABG and PCI. The need for developing decision aids to facilitate shared decisionmaking by patients, those near them, and their physicians was also mentioned.

Step 2. Generating the Expanded List of Evidence Gaps

The teleconference in Step 1 did not exhaust the discussion, and further input was sought in one-to-one interviews.[a] To inform the discussion in the interviews we searched the literature to identify trials that have been published after the Stanford CER. We also searched ClinicalTrials.gov to identify ongoing studies in the field.

Recently Published and Ongoing Studies

The literature search revealed new or ongoing trials relevant to two of the identified evidence gaps: the efficacy and safety of PCI vs. CABG for diabetics, and the relative efficacy and safety of DES vs. CABG.[b] Five RCTs published after the completion of the Stanford CER generally used DES in their PCI arms, either exclusively, or in combination with BMS. None included balloon angioplasty. Two of the 5 RCTs included well over 1,500 patients;[15,26] one included approximately 500 patients[16] and the remaining two less than 200.[27,28] Two of the largest RCTs enrolled only diabetic patients: CARDia compared DES or BMS with CABG[16] while BARI 2D[26]

[a] Four out of six Key Informants participated in one-to-one interviews.

[b] Appendix Figure B1 presents the search flow and Appendix Table B1 summarizes their findings. Appendix Table B2 shows ongoing RCTs identified in ClinicalTrials.gov.

randomized patients to either prompt revascularization with intensive medical therapy vs. intensive medical therapy alone but stratified randomization by the choice of revascularization method (PCI vs. CABG). Two trials identified in ClinicalTrials.gov, FREEDOM[23] (n=2400, NCT00086540) and VACARDS (n=790, NCT00326196)[a] included only diabetic patients, suggesting that more data on diabetics will be available soon. The majority of ongoing trials used DES stents, again suggesting that additional evidence on DES vs. CABG will be available soon.

Expanded List of Research Gaps (List 2 in Appendix B)

Briefly, in one-to-one interviews, key informants suggested that the recently published and ongoing studies on diabetics are important, but do not necessarily attenuate the need for further research on the interaction of diabetes and treatment.

The previous list of evidence gaps (List 1) was further expanded to include the subpopulation of patients what have received prior revascularization procedures (PCI or CABG). The rationale was revascularization in these patients may pose technical challenges, and that these patients may be at different "baseline" risk than others. Regarding variants of revascularization procedures, key informants suggested focusing on PCI with DES and secondarily BMS; and on-pump traditional CABG with arterial grafts. The rationale was that these are the most commonly used procedures in current clinical practice and therefore should be the priority from a health system perspective. At the same time, it was appreciated that the comparative effectiveness of newer technologies such as PCI with bioabsorbable stents, or hybrid PCI and CABG approaches may be of immediate importance to payors, because they have to make coverage decisions even when evidence is incomplete.

Further, key informants suggested adding to the list several evidence gaps that were outside the scope of the Stanford CER. The first pertained to testing to guide choice of revascularization procedure (among patients in whom revascularization is warranted). The rationale was that testing may be able to predict differential response to PCI or CABG and therefore may result in better long term outcomes. A series of invasive and noninvasive tests were suggested, including invasive testing with coronary arteriography, non-invasive computerized tomography angiography (CTA) or magnetic resonance angiography (MRA), resting or exercise single photon emission computerized tomography (SPECT), noninvasive exercise treadmill testing with or without echocardiography. Second, key informants reiterated the importance of assessing patient preferences regarding revascularization procedures, and the importance of developing and evaluating decision aids to support shared decisionmaking. Finally, they suggested that an important evidence gap pertains to developing evidence-based performance measures.

The interim expanded list of evidence gaps (List 2) is provided in Appendix B.

Step 3. Generating the Final List of Important Evidence Gaps

We pruned the interim list of evidence gaps (List 2) to contain its scope.[b] We organized the final list of evidence gaps (Table 3, List 3 also shown in Appendix B) in to four thematic areas of future research. In the following paragraphs we justify our selections.

[a] ClinicalTrials.gov lists VACARDS as "terminated", but offers no additional information. This may mean that the study has been terminated for efficacy, for futility, or because patient accrual was too slow.
[b] This step was not anticipated in the beginning of the pilot project.

Table 3. Pruned (final) list of important evidence gaps

Thematic area	*List of evidence gaps*

Thematic area 1: Comparative effectiveness of PCI vs. CABG
 	
(Evidence gaps are formed by choosing a subpopulation and a comparison and evaluating all outcomes in the list)
	Subpopulations:
	• Age >75 years
	• Prior PCI
	• Diabetes
	• Women
	• Congestive heart failure
	• Stage 3 or 4 of chronic kidney disease
	Comparisons:
	• BMS vs. on pump traditional CABG with arterial grafts
	• DES vs. on pump traditional CABG with arterial grafts
	Outcomes:
	• 30-day outcomes -objective:
	  ○ Periprocedural death
	  ○ Nonfatal MI
	  ○ Nonfatal stroke
	  ○ Unplanned urgent revascularization with CABG during the PCI
	  ○ Renal failure requiring renal replacement therapy
	  ○ Post procedural discharge to rehabilitation facilities
	  ○ Health care costs
	  ○ Readmission rates for cardiac reasons such as heart failure and unstable angina
	  ○ Others like nosocomial infections are most relevant to CABG, and pulmonary complications affect the length of stay
	• Intermediate outcomes (objective) at 1 year. The relative ranking would be similar to the applicable 30 day outcomes.
	• Longer term outcomes (objective) would be 5 years. The relative ranking would be similar to the applicable 30 day outcomes.
	• Subjective outcomes:
	  ○ QoL measured by generic or disease-specific instruments

Thematic area 2: Testing to inform choice of revascularization procedure	The evidence gap pertains to whether testing before revascularization can guide the choice of revascularization procedure, e.g., by predicting clinical response in the long term. Examples of invasive and non-invasive tests: • Invasive coronary arteriography • Non-invasive CTA or MRA • Resting or exercise SPECT • Exercise treadmill test with or without echocardiography
Thematic area 3: Enhancing patient participation	• Eliciting and measuring patient preferences • Facilitating shared decisionmaking between patients and their physicians by developing and then evaluating decision support tools.
Thematic area 4: Assessing performance	• Based on prior key informant input, there are no validated process-based performance measures that could quantify optimal care

BMS=bare metal stents; CABG=coronary artery bypass grafting; CTA=computerized tomography-based angiography; DES=drug eluting stents; MI=myocardial infarction; MRA=magnetic resonance angiography; PCI=percutaneous coronary intervention; QoL=quality of life; SPECT=single photon emission computerized tomography

Comparative Effectiveness of PCI vs. CABG (First Thematic Area)

We prioritized the patient subpopulations in the interim expanded list of the previous step (List 2) according to their relative frequency among patients who receive revascularization in the US (based on Hanan et al.,[8] see also Appendix Table B3).[a] The five most common patient groups are those listed in the first five bullets under the Subpopulations heading in Table 3. Especially for renal disease, the registries report only percentage with creatinine >2 or >2.5 mg/dL. In all

[a] The prevalence of a subpopulation was used as a crude proxy of "burden of disease."

likelihood this corresponds to stage 4 or stage 5 (renal replacement therapy). Stage 3 will be quite more frequent. Relying on prior key informant input, we included renal disease (stage 3 or 4) as a sixth most important subpopulation.

For PCI we considered angioplasty with BMS and angioplasty with DES. Although currently DES are much more commonly used compared to BMS, prior input from key informants suggested that BMS remain a PCI option that needs further consideration, especially in light of the higher DES costs. We excluded from further consideration balloon angioplasty, or other treatments that are not currently favored, such as brachytherapy. Based on prior input from key informants, we excluded bioabsorbable stents, because they are not yet routinely used and because there are a lot of remaining evidence gaps with DES and BMS already.

Based on prior input by key informants, for CABG, we considered the typical on-pump CABG with arterial grafts, which is the most common procedure. We did not prioritize minimally invasive coronary artery bypass (MIDCAB) or hybrid PCI and MIDCAB approaches because they are not used routinely, or they are used in very specific situations, respectively.

Regarding outcomes of interest, we prioritized clinical outcomes (subjective or objective, procedural, short term or long term) over surrogate measurements of revascularization success such as measurements of flow restoration or luminal diameter of revascularized vessels. It is desirable that future studies record as many of the outcomes in Table 3 as practically feasible.

Testing To Inform Choice of Revascularization Procedure (Second Thematic Area)

The identified evidence gap pertains to the ability of several types of commonly available testing procedures to predict response with treatment (see Table 3). We do not refer to other settings such as screening of asymptomatic individuals, diagnosis, or treatment monitoring. In particular screening of asymptomatic individuals is an important research area, but it was deemed to be well outside the scope of the pilot project.

Enhancing Patient Participation and Assessing Performance (Third and Fourth Thematic Areas)

The pertinent evidence gaps are listed in Table 3.

Step 4. Recommendations for Future Research

We made recommendations for future research within the aforementioned thematic areas. The most elaborate prioritization was performed for the comparative effectiveness of PCI vs. CABG (first thematic area), as this is most directly related to the Stanford CER. The other three thematic areas were not in the scope of the Stanford CER, but represent important evidence gaps according to the assessments of the key informants and our team. In the following sections, we present our rationale for what future research to perform by thematic area of future research. The actual list of proposed future research is summarized in Table 4.

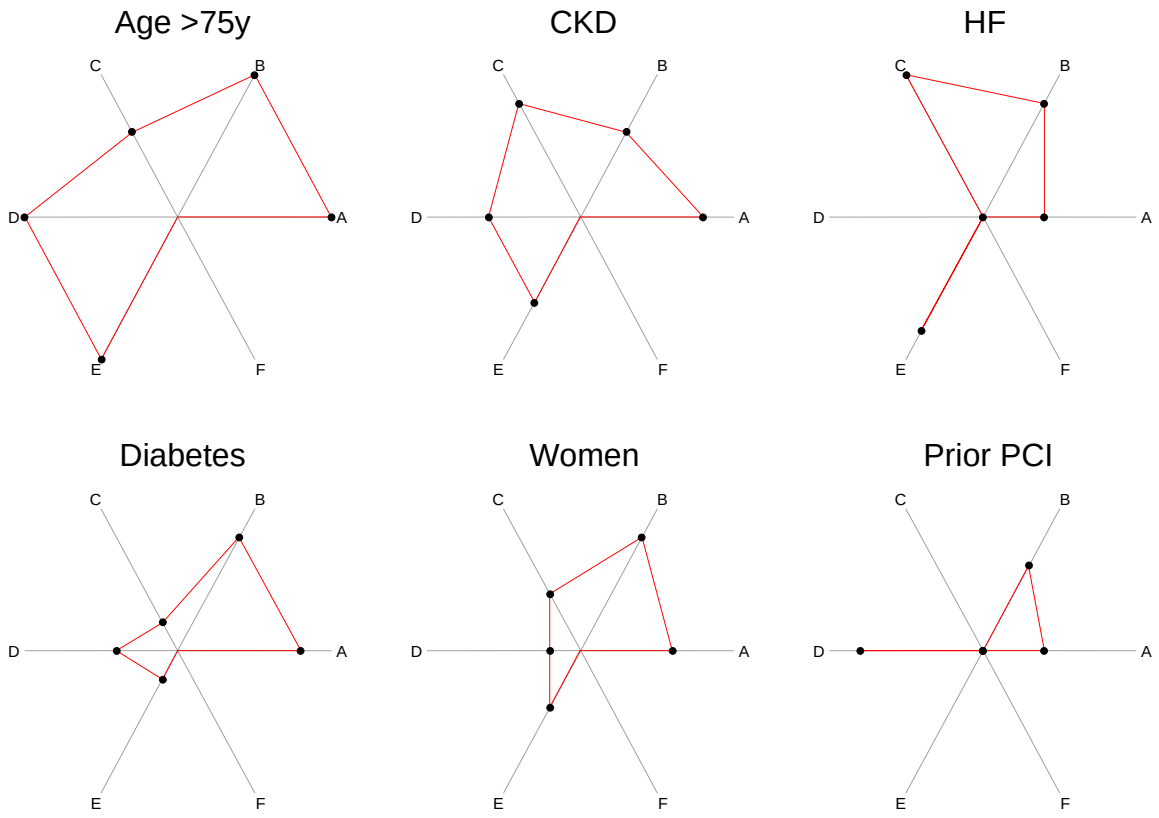
Comparative Effectiveness of PCI vs. CABG (First Thematic Area)

Summary of Key Informant Input

Figure 2 shows how key informants ranked the six priority subpopulations according to their perceived overall importance. Overall, the provided ranks suggest that the most important

subpopulation is the elderly; followed by chronic kidney disease, heart failure, diabetes, and women; and finally those who had received prior revascularization with PCI. Some key informants noted that explicit ranking of subpopulations may be problematic, for several reasons. For example, one may suggest that almost all six subpopulations are equally important. Another explanation was that the populations are not mutually exclusive and therefore ranking them is not straightforward. On the other extreme, a key informant suggested that we probably do not have a good understanding of the real determinants of differential response to PCI or CABG. If for example the major determinant is the extent of the disease, more elaborate study of subpopulations defined by clinical characteristics, personal history or comorbidity will not be the most efficient way to advance knowledge.

Figure 2. Star graph of explicit ranking of six subpopulations by key informants



The figure consists of six panels, each corresponding to a subpopulation of interest. Each key informant is represented by an axis. All axes start from the center of the graph. Each key informant is denoted with a letter from "A" through "F" (the ordering of the letters does not correspond to the order they are mentioned in the text). The ranks they provided are listed on the corresponding axis as black dots. The furthest away a black dot is from the origin, the higher the rank. Some key informants provided tied ranks for some subpopulations. Red thick lines connect the black dots in adjacent axes forming polygons. Broadly speaking, for each subpopulation larger polygon areas tend to correspond to higher overall ranks. Key informant F opted not to provide the requested feedback. The ranks of key informant C were reconstructed based on the key informant's rating in importance criteria per subpopulation (see Methods Section and Appendix B).

CKD=chronic kidney disease; HF=heart failure; PCI=percutaneous interventions; y=years.

Focused Modeling To Identify Important Parameters in Future Studies in the First Thematic Area

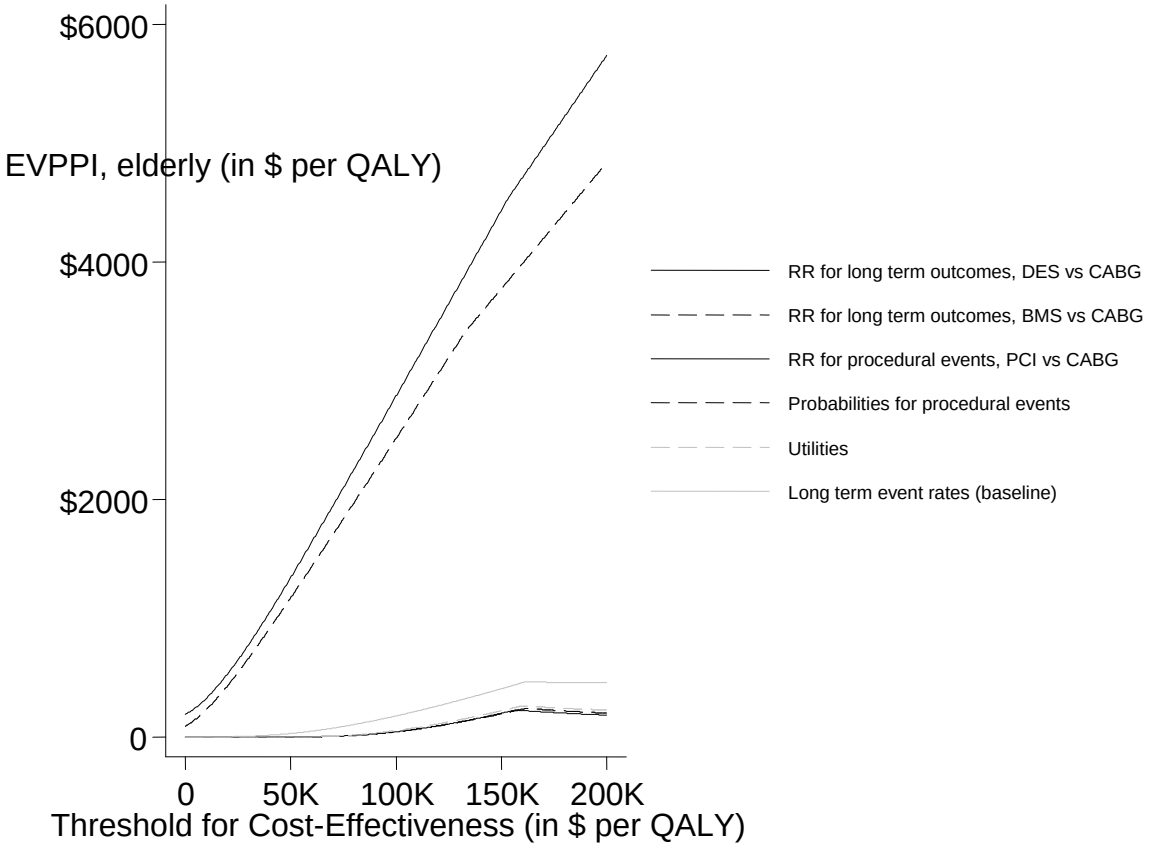
Overall, in all three quantitative approaches (decision analysis, cost-effectiveness analysis and value of information analysis) the decision relevant quantity (quality adjusted life years, incremental cost effectiveness ratio and EVPPI, respectively) was mostly influenced by the uncertainty in the treatment effects between stenting with BMS, stenting with DES and CABG, particularly for the long term outcomes of death, myocardial infarction and stroke. The uncertainty around treatment effects for risk of repeat revascularization did not exert a large influence on the decision relevant quantities, apparently because these are already known with relative precision. Other quantities in the models (such as frequency of procedural events, treatment effects for procedural events, or utilities associated with different health states) exerted much smaller influence on the decision relevant quantities. These observations were consistent *within each of the three modeled subpopulations* we examined: the RCT-type participants, the population of elderly (over 75 years of age), and diabetics. Details of the quantitative analyses are presented in Appendix D.

EVPPI analyses that explicitly consider intervention costs, identified this as a set of parameters for which more information is needed. However, good information on average costs is straightforward to obtain in a specific setting, and therefore we do not consider costs as a target for future research.

For example, Figure 3 illustrates results of analyses on the expected value of perfect information for parameters for the model on elderly patients. (Analyses are qualitatively similar for the models on RCT-type patients and diabetics—see Appendix D.) The figure shows that the value of having perfect information is highest for 2 groups of parameters that stand for the relative effects between treatments for long term outcomes (deaths, myocardial infarctions, strokes and revascularizations). This suggests that future research should inform *on the relative treatment effects between BMS, DES and CABG*. This includes not only comparisons of BMS vs. CABG and DES vs. CABG, but comparisons of *DES vs. BMS* as well![a] However, the comparison between BMS and DES was outside the scope of the Stanford CER, and therefore we do not consider it further. We touch more on this issue in the Discussion Section.

[a] This is an important point. Because our models use CABG as the reference strategy, the relative effects between BMS and DES in our models are implicitly (indirectly) defined based on the relative effects of DES vs. CABG and of BMS vs. CABG. A different parameterization of the model (e.g., having DES as a reference) would show a high expected value of perfect information for parameters for DES vs. BMS.

Figure 3. Expected value of perfect information for groups of parameters over a range of cost-effectiveness thresholds



Shown are calculations of the expected value of perfect information for 6 groups of model parameters (EVPPI). Each group of parameters is represented by a line. The horizontal axis is the "willingness to pay" (cost-effectiveness threshold), i.e., the monetary equivalent of a quality-adjusted life year (QALY). The higher the expected value of perfect information for parameters, the more valuable future research that informs on these parameters is. Refer to Appendix D for more details on methods and complete results.

BMS=bare metal stents; DES=drug eluting stents; EVPPI=expected value of perfect information for parameters; PCI=percutaneous interventions; QALY=quality-adjusted life years; RR=relative risks.

Candidate Study Designs

Modeling analyses suggest that the most influential parameters are the relative treatment effects, particularly for the long term outcomes of death, myocardial infarction and stroke, that is comparative evidence. As mentioned in Table 2 in the Methods Section, information on treatment effects (or on subpopulation-by-treatment interactions) is most efficiently obtained through a reanalysis of already compiled evidence. Another option is to perform *de novo* nonrandomized comparative studies, or preferably, to perform *de novo* RCTs (a more robust design).

Reanalyses of existing data should include all aforementioned outcomes, not only overall survival. Because of the eligibility criteria of existing efficacy RCTs, a meta-analysis of individual participant data from RCTs will not adequately represent older adults or patients with several comorbid conditions. Therefore, it is imperative to adequately study existing non-randomized data. In the field of coronary revascularization there is already a long experience

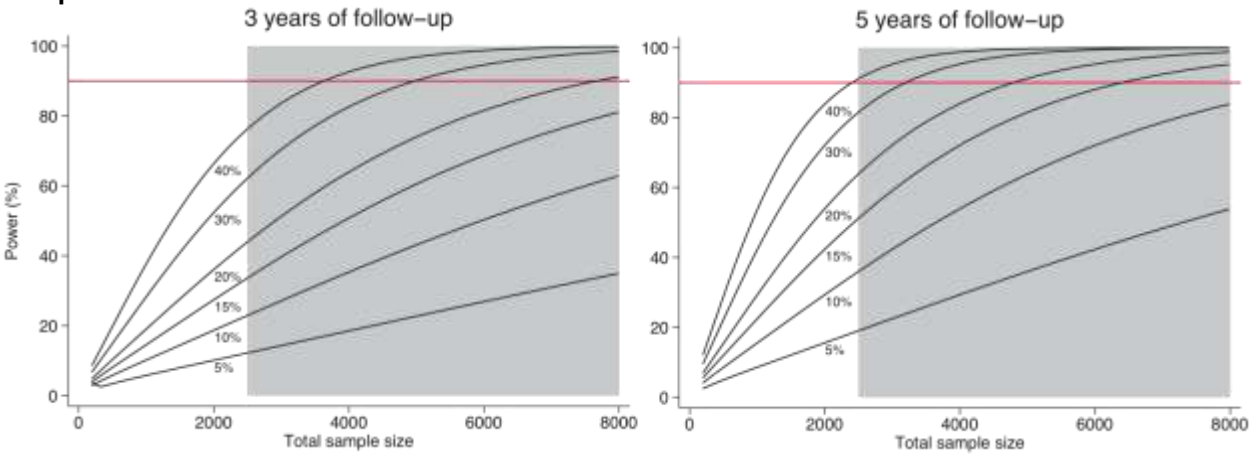
with the use of data from large clinical registries. Capitalizing on the existing registries is probably one of the most efficient ways to provide more information for the identified research gaps. It is likely that to study the interaction of chronic kidney disease or heart failure with treatment, one may have to record more detailed information that that routinely deposited in the registries for a subset (perhaps a random sample) of registry patients. The associated expense for such research should be substantially lower than that of a *de novo* RCT.

Performing a meta-analysis (or reanalysis) of existing individual patient data from randomized trials or observational studies may be challenging, as several parties have to agree and collaborate. Another challenge is that studies may not have gathered data necessary for the analyses. If this is not feasible, one would have to obtain additional information by collecting new comparative data, namely through *de novo* nonrandomized comparative trials or through *de novo* RCTs. While randomized data are preferable, one may consider favoring a nonrandomized comparison as a first step, if it can be performed in a very efficient way. An example would be to "nest" the study in an ongoing registry. This way one would take advantage of the existing infrastructure of the ongoing registry to select and followup patients, and would collect additional clinical and other information on a subset of patients, as applicable.

De novo RCTs would ideally assess all outcomes of Table 3, or as many as practically feasible. However, is it realistic to consider a new RCT to compare PCI vs. CABG? Figure 4 helps put into perspective the feasibility of RCTs to address the identified research needs (Table 4). Based on Figure 4, if the 5-year rate of the primary event of interest in the comparator arms (controls) is 30 percent, a study of approximately 2,500 patients would attain 80 to 90 percent power to find a relative effect of 0.80 over 5 years of followup. The same study would attain a power of 60 percent over a 3 year followup. To get average followup duration of approximately 5 years a trial would have to go on for 6 to 8 years. Therefore, *de novo* RCTs are feasible, but would likely require resources comparable to recent large RCTs.

In all likelihood, the primary outcome of a *de novo* RCT would have to be a composite outcome of death or myocardial infarction (to attain a high event rate). The exception is the subpopulation of elderly patients (older than 75 years), where event rates can be high enough for the outcome of death alone.

Figure 4. Power calculations for superiority RCTs for various 5-year primary event rates in the comparator arm



Plotted are power calculations for 6 different 5-year primary event rates in the comparator arm (5, 10, 15, 20, 30 and 40 percent, as shown next to each line in each panel). The calculations are for a two-sided chi-squared test at the 0.05 level of significance, and assuming a constant annual event rate, a true relative effect of 0.80 favoring the intervention arm, an allocation ratio of 1:1, no loss to follow up, no crossover between treatments, and no sequential monitoring. The gray area denotes sample sizes that may be too large (>2,500 patients total, see Methods Section). The red horizontal line stands for 90% power. Note that to get average followup duration of approximately 5 years a trial would have to go on for 6 to 8 years. This assumes a minimum followup of 2.5 years, a patient recruitment period of 5 years and a constant recruitment rate. In reality, the total sample size would have to be even larger than what is shown in the horizontal axis, as there will be loss to followup and there will also be adjustments for sequential monitoring.

Recommended Future Research (Table 4)

We distinguish three tiers of pressing research needs in the first thematic area (Table 4). It is probably prudent to first perform reanalyses of existing data, and consider performing new RCTs only if existing data are not applicable to the population of interest, or suggest a subpopulation-by-treatment interaction that is clinically important. (Alternatively, one could perform efficient *de novo* nonrandomized comparative studies, as described above, and proceed to a *de novo* RCT only if there is suggestive evidence of an important treatment-by-subgroup interaction.) This is because, in all likelihood, a *de novo* RCT would require resources comparable to recent large multimillion dollar RCTs. Reanalyses of existing data could be collaborative meta-analyses of individual patient data from prior RCTs or from large observational studies, if the populations of interest are underrepresented in prior RCTs.

Currently, DES represent a more pressing research need compared to BMS, because they are used more often than BMS in everyday practice, and this trend will most likely continue in the midterm. The above does not mean that future studies should not use BMS. If numbers allow, an option would be to further randomize patients in the PCI arm to BMS vs. DES.

1. The first tier of future research studies are studies assessing the comparative effectiveness of PCI vs. CABG among the elderly, e.g., patients older than 75 years. Up to 40 percent of people who received revascularization in large registries were 75 years or older, but elderly patients are grossly underrepresented in the existing RCTs. In addition, there is evidence of strong age-by-treatment interaction. In a recent meta-analysis of individual patient data,[a] survival was better with GABG compared with PCI for older patients (p-value for trend was 0.002 across the age categories of younger than 55, between 55 and

[a] Data from this meta-analysis were used in our modeling.

65, and older than 65 years).[17] The magnitude of the treatment effect modification was clinically significant, but it is unclear if it generalizes to older patients.

2. The second tier of future research pertains to further study of patients with heart failure and patients with stage 3 or stage 4 renal disease. Although a recent meta-analysis of individual patient data found no significant interaction between heart failure and treatment effect (p=0.46),[17] patients with heart failure were excluded from many randomized trials that formed the basis of the meta-analysis. Thus we deem that the interaction of congestive heart failure and treatment choice remains unclear. We identified no analysis of randomized data on patients with chronic kidney disease.
3. The third tier pertains to studying the interaction of diabetic patients and treatment effects. There is already suggestive evidence that survival is better with CABG compared with PCI in diabetics (p=0.014 for treatment by diabetes interaction); such an interaction may be of substantial clinical significance if confirmed by further studies.[17] However, at least two large ongoing trials, FREEDOM[23] (n=2400, NCT00086540) and VACARDS (n=790, NCT00326196)[a] included only diabetic patients. Their results will likely be presented within the next 2-3 years. In the light of this upcoming evidence we deemed that the need for additional data in diabetics is less pressing than that in the elderly.

Finally, we deemed that the subpopulation of women and people who have received PCI in the past represent less pressing priorities than the above, and do not recommend them as priority areas for future research. The previously mentioned meta-analysis of individual patient data found no interaction between sex and treatment effect (p=0.25) for overall survival.[17] The meta-analysis did not report analyses with respect to having a prior revascularization, but key informants suggested that there is little if any evidence for an interaction of treatment with history of revascularization with PCI.

Testing To Inform Treatment Choice (Second Thematic Area)

Summary of Key Informant Input

Key informants suggested that it is difficult to prioritize which baseline test or combination of tests is the best in predicting better response with CABG vs. PCI. This is not surprising, given that the Stanford CER did not review the role of testing in guiding treatment. However, key informants noted that if baseline (pre-revascularization) testing were able to predict long term differential response to PCI or CABG, its effects on patient health can be substantial. Overall, the key informants suggested that further elaboration of the ability of testing to affect clinical outcomes is highly desirable.

Candidate Study Designs

Ultimately, the clinical utility of medical tests is measured by whether the information they provide affects patient-relevant outcomes. To a large extent, effects of medical tests are indirect in nature.[29] In principle, a test result affects patient outcomes mainly by influencing treatment choices. This indirectness in the link between testing and its downstream effects poses practical challenges to comparing alternate test-and-treat strategies in clinical trials.[30]

[a] ClinicalTrials.gov lists VACARDS as "terminated," but offers no additional information. This may mean that the study has been terminated for efficacy, for futility, for safety, or because patient accrual was too slow.

In theory, a large number of potential test-and-treat strategies could be considered, even if one is restricted to the four tests in Table 3. For example, one can explore using a single test, using several tests simultaneously, or using sequential testing (subsequent tests can be used to confirm the results of previous tests[a]; or earlier tests are used to triage who gets subsequent invasive testing[b]). In all these cases, one should also consider the threshold for test positivity. Each permutation of tests, test combinations and positivity thresholds represents a possible test-and-treat strategy. It is not possible to compare all test-and-treat strategies using RCTs.[31] Instead, one should first quantify test performance in the clinical context of interest, and then assess the necessity of explicitly comparing test-and-treat strategies in RCTs. Based on the predictive accuracy of tests, one can deduce whether further study with an RCT of testing vs. no testing is mandatory, or whether a decision can be reached without further studies, following Lord et al.[22]

Recommended Future Research (Table 4)

Therefore, the most practical recommendation for future research is to explore the predictive accuracy (sensitivity and specificity) of the tests of interest[c] or combinations thereof by performing cohort studies, case control studies, or by reanalyzing baseline data from existing RCTs (if available). To be clinically useful and informative, such studies should enroll patients representative of those seen in clinical practice, and limit verification and other biases.[32,33] Such studies should be relatively feasible given the high prevalence of the conditions of interest, the widespread availability of most diagnostic technologies considered by the key informants, and the fact that most patients treated with PCI or CABG already receive extensive workup before their revascularization.

Enhancing Patient Participation (Third Thematic Area)

Summary of Key Informant Input

Input from key informants suggested that life expectancy is not the only outcome that matters to patients and those close to them. While the importance of quality of life is generally appreciated by physicians and decisionmakers it is not always clear whether, how, and to which extent it is considered. For example, a key informant suggested that current clinical practice guidelines seem to focus on life expectancy, without giving substantial weight to e.g., the likelihood of symptoms after revascularization (i.e., to quality-adjusted life expectancy).

Key informants pointed out the need for development and evaluation of decision aids, to facilitate shared decisionmaking. At the same time, it was appreciated that shared decisionmaking is probably not the norm in current clinical practice in PCI vs. CABG. In practice, a physician could sway a patient's choice by focusing on the pros or cons of either treatment. A desire of patients for ultimate treatment decisions regarding CAD treatment to be made by doctors has been observed in empirical studies.[34] Therefore, a major issue is not only the development, but also the evaluation of decision aids in real-life clinical settings.

[a] This way the whole battery of tests enhances specificity.
[b] Especially if subsequent testing is invasive.
[c] Here, arteriography, MRA or CTA, resting or exercise SPECT, exercise treadmill testing with or without echocardiography.

Candidate Study Designs

Although the importance of considering patient preferences in treatment decisions is well recognized,[35,36] it is unclear what the best approach for developing decision support tools is. A primary reason for this is the presence of multiple gaps in research for such aids: little is known about their actual effects across different decisional contexts, their acceptability to patients and physicians, or their potential effects on health-related quality of life. In addition, the development, dissemination and implementation of decision support tools are fraught with substantial methodological challenges. These challenges include the *de facto* inability to evaluate decision aids in double-blind trials, potential selection biases due to physicians' preconceptions (either positive or negative) and difficulties in reliably measuring the impact of decision support tools on clinical outcomes.

O'Connor et al.[37] discussed the research and policy implications of developing decision support tools with a focus on the challenges of designing future studies. If strong candidate decision aids for patient decisionmaking were available, then randomized trials, preferably randomizing physicians (rather than the individual patients they care for) to either employing or not employing decision aids (i.e., with cluster RCTs), would provide an optimal but possibly costly method of evaluating their effects on clinical outcomes.[38]

Recommended Future Research (Table 4)

A suitable study design to elicit patient preferences is a survey of patients, or a qualitative research study that uses focus groups of patients. It is at best unclear whether a study of patient preferences will have a major impact on the remaining research agenda. However, compared to designing and undertaking a new RCT, such studies on patient preferences are quite feasible, and relatively inexpensive.

A suitable study design to develop a decision aid is a qualitative research study that uses focus groups of patients, relatives of patients, and physicians. Based on key informant input, there are no decision aids in routine clinical use. Therefore, the question of evaluating decision aims may be premature.

Assessing Performance (Fourth Thematic Area)

Summary of Key Informant Input

All key informants agreed that development of evidence-based performance measures, monitoring of practice based on these measures and providing feedback to health care facilities and practitioners can improve revascularization outcomes and may reduce inequalities in provided health care at the national level.

Candidate Study Designs

In principle, it should be feasible to identify *process-based measures of performance* in cardiovascular care, as there are a lot of interventions with strong support from RCTs in the domain of cardiovascular disease. Evidence suggests that establishing performance measures and active monitoring of physician performance has positive impact on the quality of delivered health care.[39] However, for most chronic diseases, including CAD, it is not possible to use simple "all-or-nothing" measures to quantify optimal care; simple measures are limited to quantifying poor care.[40] Therefore, developing evidence-based performance measures for major

diseases can be considered as a priority for a health system, as it carries the promise of reducing inequalities in health care access and delivery.

It is not clear what the optimal study design is for developing process-based performance measures.[41] One would have to define the exact details of the clinical area to be measured, select key aspects of care for measurement, design specifications for the measures, develop a data collection strategy, test the scientific strength (validity, reliability, feasibility) of the measures, and then monitor practice and provide feedback.[42-44]

It appears that to develop performance measures one option is to perform qualitative research where focus groups of scientists with relevant expertise would propose specific measures based on the current evidence base and analyses of administrative data. Health care professionals as well as investigators with expertise in bioinformatics, quality control, or operations research would offer complementary expertise. To inform these focus groups it may be important to conduct analyses of administrative data related to the processes of interest.

The actual evaluation of whether the implementation of a performance monitoring and feedback system could be readily performed in observational studies that measure performance before and after the implementing the monitoring and feedback mechanism. A much stronger design would be a cluster RCT, where one would randomize health care facilities to implement versus not implement the system. However, unless there is substantial infrastructure already in place, such a cluster RCT would not be easy to perform.

Recommended Future Research (Table 4)

The optimal research design to develop reliable process-based measures of performance is not clear. One option is to perform qualitative research using focus groups of health care professionals as well as participants with expertise in bioinformatics, quality control or operations research. Evaluating process-based performance measures would be typically performed in large-scope studies that analyze administrative data before and after the implementation of performance monitoring and feedback systems, or in large cluster-randomized trials.

Table 4. Prioritized research designs to address future research needs when studying the comparative effectiveness of PCI and CABG

Thematic area	#	*Population*	*Intervention*	*Comparator*	• ***Primary outcomes*** • ***Secondary outcomes***	*Design*	*Feasible*	*Research cost*
(A) Comparative effectiveness and safety	1	General population of elderly patients >75 years old	On-pump CABG	PCI with DES (or with DES and BMS)	• Total mortality or composite of total mortality or myocardial infarctions • Other objective and subjective outcomes	MIPD of RCT or registry data	Yes	Low
						Prospective comparative observational study, preferably nested in an existing cohort or registry	Yes	Medium
						RCT	Probably	High
	2*	Heart failure	[As above]	[As above]	• Composite of total mortality or myocardial infarctions† • Other objective and subjective outcomes	[same options as in row A1]	[same options as in row A1]	[same options as in row A1]
	3*	Renal disease stage 3 or 4	[As above]	[As above]	[As above]	[same options as in row A1]	[same options as in row A1]	[same options as in row A1]
	4	Diabetes	[As above]	[As above]	[As above]	[same options as in row A1]	[same options as in row A1]	[same options as in row A1]
(B) Testing to predict treatment response	1	General population of revascularized patients	Invasive and noninvasive tests	Not applicable	Predictive sensitivity or specificity or related metrics	Retrospective analysis of RCT-based data	Yes	Low
						Prospective cohort	Yes	Medium
(C) Enhancing patient participation	1	General population of revascularized patients	Not applicable	Not applicable	Description of patient preferences	Qualitative research	Yes	Low
						Survey	Yes	Low
	2	General population of revascularized patients	Not applicable	Not applicable	Development of decision support tools	Qualitative research	Yes	Low or medium
					Evaluation of decision support tools	RCT	Unclear	Unclear
(D) Assessing performance	1	General population of revascularized patients	Not applicable	Not applicable	Development of evidence-based performance measures	Qualitative research; observational studies	Unclear	Unclear
					Evaluation of systems that monitor performance	Before-after observational studies or cluster RCT	Unclear	Unclear

*so that power calculations result in a feasible trial (<2500 sample size).

BMS=bare metal stent; CABG=coronary artery bypass graft surgery; DES=drug eluting stent; MIPD=meta-analysis of individual patient data; RCT=randomized controlled trial.

Discussion

Key Informants

In the current project we used input from a limited group of key informants to identify evidence gaps regarding treatments for stable coronary artery disease amenable to revascularization therapy. We subsequently organized them in four thematic areas, namely comparative effectiveness and safety; use of testing to guide treatment choice; enhancing patient participation and assessing performance. We requested feedback from key informants on how they would prioritize future research needs, but we did not automatically accept their suggestions. We considered their feedback as an additional input along with the Stanford CER, the results of our literature searches, and our insights from focused quantitative analyses. This approach offers a degree of protection from potential conflicts of interest on behalf of the key informants. On the other hand it may have resulted in the omission of real priorities from the final list of research needs. For example, we opted not to prioritize an RCT in women high enough in our list, because there is prior evidence that there is no (detectable) sex-by-treatment interaction (Box 1).

By its very nature, a document on Future Research Needs should be as specific as feasible in its assessments. Key informant input was invaluable in *expanding the list* of important evidence gaps, yet expanded lists are more challenging to prioritize. Key informants could not readily identify a specific subpopulation of patients with CAD as the most important priority. While they did provide ranks when they were explicitly asked to do so, their feedback was that, generally, differences between the importance of research needs that received different ranks are not always clear cut. Ideally, a more formal process to engage key informants would be desirable. For example there is evidence that a Delphi process can lead to the development of research recommendations consistent across various stakeholder groups.[45] However, there are logistical challenges to routinely performing Delphi processes to prioritize research needs. We believe that any approach to elicit the perceived research needs from key informants should be primarily practical.

The size and composition of the key informant group is also important to discuss (Box 1). We opted to use a small group of six so that we have the opportunity to perform one-to-one interviews. We selected key informants in a nonsystematic way, i.e. we did not use formal sampling methods to select among a large pool of candidates, but aimed for a diverse and representative group that was likely to generate a diverse list of research needs. We invited a medical officer of a funding agency for cardiovascular research, a medical director of a major payor, an interventional cardiologist, a CABG surgeon, a general cardiologist, and a clinical researcher and methodologist with contributions in the field.

We did not invite a patient representative, because we deemed that the patient perspective is not particularly important in the current phase of prioritizing research needs. Unless they have specialized knowledge, patients cannot contribute to identifying priority subpopulations or interventions which represent pressing research needs. An argument in favor of including patient representatives is that they can identify important outcomes that elude the attention of health care professionals. However, there is a long research experience in cardiovascular medicine, and we can reasonably expect that the most important objective and subjective outcomes are already very well known. It may be more fruitful to include patient input when designing a new study, in an effort to augment the list of secondary or tertiary outcomes that the study records.

There is a growing literature in diverse medical fields comparing research priorities as perceived by different stakeholder groups, including researchers and patients. Oliver et al. attempted to develop an evidence based approach for involving "consumers" of research in setting the research agenda for the UK National Health System.[46] They noted that the existing literature consisted mostly of descriptive reports "by researchers who were key actors in involving consumers" and indicated that there does not exist a rigorous evidence base for consumer involvement in research agenda setting. O'Donnell et al. surveyed UK-based funders for their approach to involving consumers about what health-related research should be funded.[47] They highlighted several issues of concern regarding the difficulty in selecting "appropriate" consumers to provide input, difficulties in understanding the research proposals discussed, possible "distortions" of funding decisions due to consumer biases and uncertainty about how to reconcile discrepancies between professionals' and consumers assessments of research priorities.[47,48]

In addition, consideration has to be given to whether the key informants can effectively convey what the important questions are for decisionmakers (patients, clinicians, funders, policy makers). There is only limited evidence that surveys of experts using "nominal group" techniques can produce recommendations for future research that are representative of an entire research community.[49] Formal survey methods for eliciting research priorities from research groups have been used in a relatively limited number of diverse fields such as occupational medicine,[50,51] hematology,[52] nephrology,[45] dental medicine,[53] and nursing.[54,55] However, formal surveys are not practical to perform for the development of future research needs documents.

Quantitative Analyses Based on Simple Models

We explored whether the use of quantitative analyses offers additional insights. Quantitative analyses require a mathematical model (decision model). It is difficult to develop, debug, and validate a detailed decision model in a short period of time.[a] Instead, we adopted the operational approach of identifying a published model that is simple and has face validity, and reparameterized it to use information from the evidence report and recent large studies. The advantage of using quantitative methods is that they make explicit all assumptions and data sources used. Further, quantitative methods require a well formulated decisional context, much like a systematic review requires a well formulated answerable clinical question.

One insight we obtained during our value of information analysis is that by its very nature, a simple model *cannot be used to prioritize subpopulations*. To do so, one would have to use a model that includes all the subpopulations of interest, and is appropriately parameterized using subpopulation-by-treatment interactions. It follows that this can be easily done only if the subpopulations of interest are mutually exclusive, which is seldom the case. For example, to prioritize between "diabetics" and "elderly" we would have to explicitly model their intersection ("elderly diabetics"), and we do not have enough data to do so. This has been described elsewhere as well.[59]

We used the models as tools to identify where future research is needed the most, albeit indirectly: presumably, one would recommend further research to inform the parameters that are most uncertain in the one-way sensitivity analyses of decision and cost-effectiveness models. This is an *atypical use of modeling*. We only identify and rank "influential" parameters and stop

[a] It is may be feasible to develop a decision model *de novo* if it is sufficiently simple. Even fairly simple representations of clinical dilemmas can provide useful insights.[56-58]

short of providing insights on treatment choices for different assumptions and circumstances, leaving a lot to be desired. However this is a conscious choice:[a] Ranking of influential parameters is likely to remain stable even if more elaborate models are used, or if better data are used to populate parameters for which we made simplifying assumptions. In contrast, exact values may change substantially.

An additional insight from the quantitative exercises was that it is imperative to use estimates from network meta-analysis to parameterize models that compare three or more treatment options. Because treatment effect estimates from network meta-analysis are consistent between them, changing the parameterization of the model will always result in identical conclusions, as is logically expected.[57] Models that are not parameterized based on consistent estimates from network meta-analysis can give different conclusions under different parameterizations, and can be misleading.[60]

There is a growing literature on methods and applications of VOI analysis. The UK's National Institutes for Clinical Excellence routinely includes probabilistic decision and cost effectiveness analyses in their systematic reviews of health technologies and encourage use of VOI analysis to inform research priorities. In the US there is a tendency to avoid of cost-effectiveness analyses,[61,62] and this may extend to VOI analyses. VOI analysis is a theoretically motivated method to prioritize future research. Our application of VOI analysis afforded us two insights: First, in view of the relatively large expected costs of medical interventions in the US, estimates of the EVPI and of EVPPI will tend to be much higher compared to those typically reported in other countries. Second, any uncertainty in cost-related parameters is going to influence EVPI estimates substantially especially at low willingness-to-pay thresholds.[59] We found that focusing on the EVPPI estimates of non-cost-related subgroups of parameters is more informative for prioritizing future research, because uncertainty around costs is relatively straightforward to reduce.[b] In our case we focus on the ranking of different groups of parameters according to their EVPPI, but not on the actual EVPPI values. Although we ignore information by focusing on ranks, we also gain by having results that are less likely to change if we modify our models.

Our choice of a 10 year horizon for modeling is arbitrary. In VOI analyses it is equivalent to assuming a price "shock" ten years from now, e.g., that a new highly cost-effective treatment will appear after ten years that will render all choices moot.[58] Another way to state this is that we simply do not wish to consider a time horizon longer than 10 years as this is already too far into the future, when treatment advances, changes in treatment prices, or changes in the population case mix may drastically change the decisional context. Generally, longer time horizons would tend to increase the actual estimates of the value of future research. Because in our approach we are in relative ranking rather than actual values, considering longer time horizons would most likely not change our insights.

Overall, we believe that modeling is useful for making specific future research recommendations, and not so much for broad-stroke qualitative statements. Ideally, if modeling is contemplated as part of the Future Research Needs documents, provisions should be made

[a] This is admittedly a defensive stance. It is very influenced by the fact that our models are not developed with the same rigor as some of the well-known and elaborate models that are used to analyze clinical decisions; they have not been calibrated using external data; and their predictions have not been validated in external data. They are operational models that are constructed to make use of the summary information obtained from an evidence report, that can still offer broad insights.

[b] By conducting, for example, surveys of hospitals offering specific procedures or utilizing pre-existing databases.

early on, during the actual conduct of the CER. This way one would ensure that appropriate data will be identified, appraised and used in the model in a timely manner.[57,58]

The Implications of an Incomplete "Intervention Space" in the Decisional Context

An additional insight from our quantitative analyses pertained to the importance of having a well defined decisional context. We already commented on the fact that no revascularization (i.e., use of medical therapy) is a valid clinical option in many if not most patients in whom both PCI and CABG would be considered.[13] We excluded medical therapy from our considerations, because evidence on medical therapy was not reviewed in the Stanford CER. Adding medical therapy to the incomplete "intervention space" of our decisional context could dramatically change our suggestions on future research needs. For example, we may have recommended that medical therapy be added to the interventions to be compared in future RCTs. This of course can potentially affect the feasibility of future studies.

Further, in our qualitative approach we did not expand on the comparison between DES and BMS, as our attention was focused on comparisons of PCI methods vs. CABG. After performing quantitative analyses however, it became evident that information is needed on the comparative effectiveness and safety across all three interventions, BMS, DES and CABG. In other words, this means that additional information on BMS vs. DES comparisons may be a valid alternative for future research. It should be noted that different parameterizations of our model could be used to estimate the uncertainty that exists regarding the relative treatment effects of DES and BMS.

Box 1. Potential methodological questions in developing future research needs documents

What is the optimal role for the key informants? Is their main role to flesh out the evidence gaps or should they be the ones who prioritize the future research needs?
Is a handful of key informants enough to ensure good coverage of the major considerations in the topic? What is the best composition for the key informant group? Is it important that they have a strong methodological background?
Explore the use of simplified models to gain insights for future research. Explore whether the ranking of the importance of different parameters would change if more elaborate models were used instead of simplified ones. Validate or refute our assumption that ranking of the importance of different parameters is relatively robust to bad quality data on costs and utility weights or other model inputs apart from treatment effects.
Explore whether and when simple models can be used to *prioritize subpopulations*.
Explore the role of network meta-analysis in modeling. Explore empirically the implications of not using consistent estimates of treatment effects in modeling analyses. Explore the implications on incoherence in a network-meta-analysis on modeling results and conclusions.

Conclusion

Based on our review of the Stanford CER, input from key informants, newly published and ongoing studies and our insights from quantitative analyses we identified three thematic areas of future research needs to inform choice between mechanical revascularization procedures in patients with nonacute CAD: comparative effectiveness and safety; use of testing to guide treatment choice; and elicitation of patient preferences, development of decision aids and evidence-based performance measures.

In the first area, pressing priorities are analyses of individual participant data or new RCTs on the comparative effectiveness and safety of DES vs. CABG in elderly patients (older than 75 years); patients with heart failure or patients with stage 3 or 4 chronic kidney disease; and patients with diabetes.

An additional priority is to perform studies of the ability of invasive (arteriography) and non invasive tests (MRA or CTA, resting or exercise SPECT, exercise treadmill testing with or without echocardiography) to predict differential response to PCI or CABG. Pending such data it is unclear whether RCTs are needed to compare testing to guide choice of revascularization vs. no testing.

It is relatively inexpensive to perform studies to enhance understanding of patient preferences. It is also important to explore the development of decision support aids as well as to develop evidence-based performance measures.

References

1. Coronary artery bypass surgery versus percutaneous coronary intervention with stent implantation in patients with multivessel coronary artery disease (the Stent or Surgery trial): a randomised controlled trial. Lancet 2002 Sept 28;360(9338):965-970.
2. The final 10-year follow-up results from the BARI randomized trial. J Am Coll Cardiol 2007 Apr 17;49(15):1600-1606.
3. Henderson RA, Pocock SJ, Sharp SJ, et al. Long-term results of RITA-1 trial: clinical and cost comparisons of coronary angioplasty and coronary-artery bypass grafting. Randomised Intervention Treatment of Angina. Lancet 1998 Oct 31;352(9138):1419-1425.
4. Hueb W, Lopes NH, Gersh BJ, et al. Five-year follow-up of the Medicine, Angioplasty, or Surgery Study (MASS II): a randomized controlled clinical trial of 3 therapeutic strategies for multivessel coronary artery disease. Circulation 2007 Mar 6;115(9):1082-1089.
5. Hueb WA, Bellotti G, de Oliveira SA, et al. The Medicine, Angioplasty or Surgery Study (MASS): a prospective, randomized trial of medical therapy, balloon angioplasty or bypass surgery for single proximal left anterior descending artery stenoses. J Am Coll Cardiol 1995 Dec;26(7):1600-1605.
6. Sedlis SP, Ramanathan KB, Morrison DA, et al. Outcome of percutaneous coronary intervention versus coronary bypass grafting for patients with low left ventricular ejection fractions, unstable angina pectoris, and risk factors for adverse outcomes with bypass (the AWESOME Randomized Trial and Registry). Am J Cardiol 2004 Jul 1;94(1):118-120.
7. Serruys PW, Ong AT, van Herwerden LA, et al. Five-year outcomes after coronary stenting versus bypass surgery for the treatment of multivessel disease: the final analysis of the Arterial Revascularization Therapies Study (ARTS) randomized trial. J Am Coll Cardiol 2005 Aug 16;46(4):575-581.
8. Hannan EL, Racz MJ, Walford G, et al. Long-term outcomes of coronary-artery bypass grafting versus stent implantation. N Engl J Med 2005 May 26;352(21):2174-2183.
9. Malenka DJ, Leavitt BJ, Hearne MJ, et al. Comparing long-term survival of patients with multivessel coronary disease after CABG or PCI: analysis of BARI-like patients in northern New England. Circulation 2005 Aug 30;112(9 Suppl):I371-I376.
10. Pell JP, Walsh D, Norrie J, et al. Outcomes following coronary artery bypass grafting and percutaneous transluminal coronary angioplasty in the stent era: a prospective study of all 9890 consecutive patients operated on in Scotland over a two year period. Heart 2001 Jun;85(6):662-666.
11. Dzavik V, Ghali WA, Norris C, et al. Long-term survival in 11,661 patients with multivessel coronary artery disease in the era of stenting: a report from the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) Investigators. Am Heart J 2001 Jul;142(1):119-126.
12. Bravata D, McDonald K, Gienger A, et al. Comparative effectiveness of percutaneous coronary interventions and coronary artery bypass grafting for coronary artery disease. Comparative Effectiveness Review No 9 (prepared by Stanford UCSF-EPC) 2007.
13. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 2007 Apr 12;356(15):1503-1516.
14. Trikalinos TA, Alsheikh-Ali AA, Tatsioni A, et al. Percutaneous coronary interventions for non-acute coronary artery disease: a quantitative 20-year synopsis and a network meta-analysis. Lancet 2009 Mar 14;373(9667):911-918.
15. Serruys PW, Morice MC, Kappetein AP, et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. N Engl J Med 2009 Mar 5;360(10):961-972.

16. Kapur A, Hall RJ, Malik IS, et al. Randomized comparison of percutaneous coronary intervention with coronary artery bypass grafting in diabetic patients. 1-year results of the CARDia (Coronary Artery Revascularization in Diabetes) trial. J Am Coll Cardiol 2010 Feb 2;55(5):432-440.
17. Hlatky MA, Boothroyd DB, Bravata DM, et al. Coronary artery bypass surgery compared with percutaneous coronary interventions for multivessel disease: a collaborative analysis of individual patient data from ten randomised trials. Lancet 2009 Apr 4;373(9670):1190-1197.
18. Rao C, Aziz O, Panesar SS, et al. Cost effectiveness analysis of minimally invasive internal thoracic artery bypass versus percutaneous revascularisation for isolated lesions of the left anterior descending artery. BMJ 2007 Mar 24;334(7594):621.
19. Yock CA, Boothroyd DB, Owens DK, et al. Cost-effectiveness of bypass surgery versus stenting in patients with multivessel coronary artery disease. Am J Med 2003 Oct 1;115(5):382-389.
20. Bischof M, Briel M, Bucher HC, et al. Cost-Effectiveness of Drug-Eluting Stents in a US Medicare Setting: A Cost-Utility Analysis with 3-Year Clinical Follow-Up Data. Value Health 2009 Mar 11.
21. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002 Aug 30;21(16):2313-24.
22. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? Ann Intern Med 2006 Jun 6;144(11)850-855.
23. Farkouh ME, Dangas G, Leon MB, et al. Design of the Future REvascularization Evaluation in patients with Diabetes mellitus: Optimal management of Multivessel disease (FREEDOM) Trial. Am Heart J 2008 Feb;155(2):215-223.
24. Jones RH, Velazquez EJ, Michler RE, et al. Coronary bypass surgery with or without surgical ventricular reconstruction. N Engl J Med 2009 Apr 23;360(17):1705-1717.
25. Velazquez EJ, Lee KL, O'Connor CM, et al. The rationale and design of the Surgical Treatment for Ischemic Heart Failure (STICH) trial. J Thorac Cardiovasc Surg 2007 Dec;134(6):1540-1547.
26. Frye RL, August P, Brooks MM, et al. A randomized trial of therapies for type 2 diabetes and coronary artery disease. N Engl J Med 2009 Jun 11;360(24):2503-2515.
27. Buszman P, Wiernek S, Szymanski R, et al. Percutaneous versus surgical revascularization for multivessel coronary artery disease: a single center 10 year follow-up of SOS trial patients. Catheter Cardiovasc Interv 2009 Sept 1;74(3):420-426.
28. Thiele H, Neumann-Schniedewind P, Jacobs S, et al. Randomized comparison of minimally invasive direct coronary artery bypass surgery versus sirolimus-eluting stenting in isolated proximal left anterior descending coronary artery stenosis. J Am Coll Cardiol 2009 Jun 23;53(25):2324-2331.
29. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991 Apr;11(2):88-94.
30. Lijmer JG, Leeflang M, Bossuyt PM. Proposals for a phased evaluation of medical tests. Med Decis Making 2009 Sept;29(5):E13-E21.
31. Trikalinos TA, Siebert U, Lau J. Decision-analytic modeling to evaluate benefits and harms of medical tests: uses and limitations. Med Decis Making 2009 Sept;29(5):E22-E29.
32. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999 Sept 15;282(11):1061-1066.
33. Rutjes AW, Reitsma JB, Di NM, et al. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006 Feb 14;174(4):469-476.
34. Lambert N, Rowe G, Bowling A, et al. Reasons underpinning patients' preferences for various angina treatments. Health Expect 2004 Sept;7(3):246-256.
35. Kassirer JP. Adding insult to injury. Usurping patients' prerogatives. N Engl J Med 1983 Apr 14;308(15):898-901.
36. Kassirer JP. Incorporating patients' preferences into medical decisions. N Engl J Med 1994 Jun 30;330(26):1895-1896.
37. O'Connor AM, Fiset V, DeGrasse C, et al. Decision aids for patients considering options affecting cancer outcomes: evidence of efficacy and policy implications. J Natl Cancer Inst Monogr 1999;(25):67-80.

38. O'Connor AM, Bennett CL, Stacey D, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2009 Jul 8;(3):CD001431.
39. Asch SM, McGlynn EA, Hogan MM, et al. Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. Ann Intern Med 2004 Dec 21;141(12):938-945.
40. Hayward RA. All-or-nothing treatment targets make bad performance measures. Am J Manag Care 2007 Mar;13(3):126-128.
41. Eddy DM. Performance measurement: problems and solutions. Health Aff (Millwood) 1998 Jul;17(4):7-25.
42. Golden WE, Hermann RC, Jewell M, et al. Development of evidence-based performance measures for bipolar disorder: overview of methodology. J Psychiatr Pract 2008 May;14 Suppl 2:18-30.
43. Groce III JB, Translating evidence-based guidelines into performance measures for venous thromboembolism and acute coronary syndrome. Am J Health Syst Pharm 2007 Jun 1;64(11 Suppl 7):S25-S29.
44. van Gaal BG, Schoonhoven L, Hulscher ME, et al. The design of the SAFE or SORRY? study: a cluster randomised trial on the development and testing of an evidence based inpatient safety program for the prevention of adverse events. BMC Health Serv Res 2009;9:58.
45. Kellum JA, Mehta RL, Levin A, et al. Development of a clinical research agenda for acute kidney injury using an international, interdisciplinary, three-step modified Delphi process. Clin J Am Soc Nephrol 2008 May;3(3):887-894.
46. Oliver S, Clarke-Jones L, Rees R, et al. Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. Health Technol Assess 2004 Apr;8(15):1-IV.
47. O'Donnell M, Entwistle V. Consumer involvement in decisions about what health-related research is funded. Health Policy 2004 Dec;70(3):281-290.
48. Herxheimer A. Relationships between the pharmaceutical industry and patients' organisations. BMJ 2003 May 31;326(7400):1208-1210.
49. Vella K, Goldfrad C, Rowan K, et al. Use of consensus development to establish national research priorities in critical care. BMJ 2000 Apr 8;320(7240):976-980.
50. van der Beek AJ, Frings-Dresen MH, van Dijk FJ, et al. Priorities in occupational health research: a Delphi study in The Netherlands. Occup Environ Med 1997 Jul;54(7):504-510.
51. Harrington JM, Calvert IA. Research priorities in occupational medicine: a survey of United Kingdom personnel managers. Occup Environ Med 1996 Sept;53(9):642-644.
52. Brittenham GM, Franks AL, Rickles FR. Research priorities in hereditary hemochromatosis. Ann Intern Med 1998 Dec 1;129(11):993-996.
53. Chestnutt IG, Taylor MM. Prioritisation of research recommendations from a national needs assessment programme. Health Bull (Edinb) 2000 Sept;58(5):396-402.
54. Bond S, Bond J. A Delphi survey of clinical nursing research priorities. J Adv Nurs 1982 November;7(6):565-75.
55. Schmidt K, Montgomery LA, Bruene D, et al. Determining research priorities in pediatric nursing: a Delphi study. J Pediatr Nurs 1997 Aug;12(4):201-207.
56. Dunn VH, Moskowitz AJ, Lau J, et al. Can "hypersimplified" decision trees be used instead of Markov models? [ABSTRACT]. Med Decis Making 1984;4.
57. Claxton K, Ginnelly L, Sculpher M, et al. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. Health Technol Assess 2004 Jul;8(31):1-103, iii.
58. Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. Pharmacoeconomics 2006;24(11):1055-1068.
59. Briggs A, Claxton K, Schulpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.
60. Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty. Med Decis Making 2002 Jul;22(4):359-371.

61. Neumann PJ, Greenberg D. Is the United States ready for QALYs? Health Aff (Millwood) 2009 Sept;28(5):1366-1371.

62. Neumann PJ. Why don't Americans use cost-effectiveness analysis? Am J Manag Care 2004 May;10(5):308-312.

.

Abbreviations

AHRQ	Agency for Healthcare Research and Quality
BMS	Bare metal stent
CABG	Coronary artery bypass graft surgery
CAD	Coronary artery disease
CER	Comparative effectiveness report
CKD	Chronic Kidney Disease
CTA	Computerized tomography angiography
DES	Drug eluting stent
EPC	Evidence-based Practice Center
EVPI	Expected value of perfect information
EVPPI	Expected value of perfect information for parameters
HF	Heart failure
HR	Hazard ratio
hrQoL	Health related quality of life
ICER	Incremental cost-effectiveness ratio
LAD	Left anterior descending coronary artery
LVEF	Left ventricular ejection fraction
MH	Medical subject heading
MI	Myocardial infarction
MIDCAB	Minimally invasive direct coronary artery bypass surgery
MIPD	Meta-analysis of individual patient data
MRA	Magnetic resonance angiography
NA	Not applicable
NCT	Clinical trial registry number
NYHA	New York Heart Association
PCI	Percutaneous coronary intervention
QALY	Quality adjusted life year
RCT	Randomized controlled trial
RR	Relative risk
SCHIP	State Children’s Health Insurance Program
SF-36	Short Form-36 questionnaire
SPECT	Single photon emission computed tomography
UCSF	University of California at San Francisco
VOI	Value of information

Appendixes

Appendix A. Search Strategies
Appendix B. Prioritization Tools
Appendix C. Excluded Studies
Appendix D. Methods and Results of Quantitative Analyses

Appendix A. Search Strategies

PubMed

We searched PubMed for randomized controlled trial reports published after the completion of the Stanford CER (our search covered the period from 2006 to August 4, 2010).

We used the simple search strategy:

"coronary artery bypass[MH] AND (angioplasty[MH] OR stents[MH] OR percutaneous coronary intervention)" and set limits for "Randomized controlled trial" or "Controlled Clinical Trial" or "Clinical Trial, Phase III" or "Clinical Trial, Phase IV" as publication type. We only considered studies published in English.

ClinicalTrials.gov

We searched the ClinicalTrials.gov database to identify ongoing studies comparing any PCI intervention with any type of CABG, including "hybrid" interventions that combine PCI and CABG.

We searched for "coronary artery bypass" and "percutaneous coronary intervention" and selecting the age groups "Adult" and "Adult Senior" and study type "Interventional studies" and separately for "Observational studies."

Appendix B. Prioritization Tools

Appendix B. Contents

Appendix B. Step 1: Initial List of Evidence Gaps (List 1)

The following is the initial list of evidence gaps. It was generated based on the Stanford CER and input from key informants.

List 1. Initial list of evidence gaps

Population	***Intervention/PCI***	***Comparator/CABG***	***Outcomes***	***Perspective; time horizon***	***Other***
CAD in whom revascularization is considered. • 1-vessel (proximal LAD) • Most 2-vessel CAD • Less severe 3-vessel *Modifiers/Subgroups* • Age, sex (women), race (Asian) • Diabetes • High BMI • Heart failure • Chronic kidney disease • Angiographic factors, such as number of vessels; disease location (LM, LAD, RCA, CCA); type of lesions (long lesions, small diameter vessels, altered anatomy) • Presentation (silent ischemia, stable angina, unstable angina, ACS, MI, cardiogenic shock) • Comedications, adjunctive medical therapies?	*Stents* • DES • BMS • Bioabsorbable *Angioplasty* • Laser • Balloon *Other* • Atherectomy	*Traditional (sternotomy)* • On pump • Off pump *Minimally invasive* • On pump • Off pump • Hybrid PCI/CABG *Graft type* • Saphenous vein • Internal mammary artery *Cardioplegia* • Yes • No (beating heart) *Other* with transmyocardial revascularization (TMR)	*Short-term objective* • Completeness of revascularization • Peri-procedural death • Non-fatal MI • Stroke • Unplanned/urgent CABG during same PCI hospitalization • Nosocomial infections • Respiratory failure/Pulmonary complications • Renal failure • Other procedural complications • Post procedure discharge to rehabilitation facilities • Readmission rates • Costs *Intermediate and long term objective* • Long-term overall survival • Long-term cardiovascular survival • Non fatal MI • Event free survival (define events, e.g. MI free or repeat-revascularization-free) • Repeat revascularization procedures • Target lesion revascularization • PCI restenosis • CABG graft closure • Congestive heart failure • Cardiac arrhythmias • Use of anti-angina medications *Short and long term subjective* • QOL • Freedom from angina • Cognitive impairment • Productivity and functional capacity	• Societal • Researcher • Payor • Funder • Patient • 1-2 y • 5 y • 10 y	• Patient preferences • Development and evaluation of decision aids for shared decision-making

ACS=acute coronary syndrome; BMI=Body mass index; BMS=bare metal stents; CABG=coronary artery bypass grafting; CCA=circumflex coronary artery DES=drug-eluting stents; LM=Left main; LAD=left anterior descending artery; MI=myocardial infarction PCI=percutaneous interventions; QOL=quality of life; RCA=right coronary artery

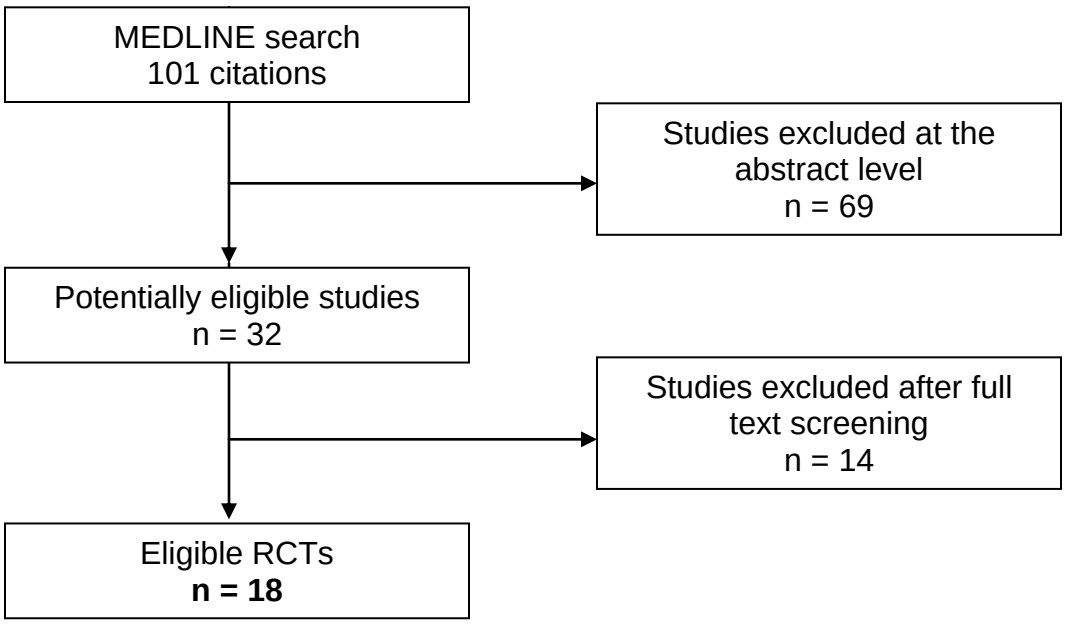
Appendix B. Step 2: Search Results and Interim Expanded List of Evidence Gaps

We then proceeded to further refine the list of evidence gaps. To inform additional discussions with the key informants, we searched for relevant recently published or ongoing studies that were not included in the Stanford CER.

Searches

Appendix Figure B1 shows the flow of the literature. From a total of 101 citations retrieved by our searches we identified 32 potentially relevant studies that appeared after the completion of the Stanford CER and we reviewed them in full text. After full text screening, 18 studies were consider eligible and 14 were excluded (Appendix C lists the excluded studies and reasons for exclusion). Eleven publications were updated reports of RCTs included in the Stanford CER.[1-11] Eight publications[12-19] reported on 5 RCTs that were not included in the Stanford CER. Appendix Table B1 summarizes their findings. Appendix Table B2 shows ongoing RCTs identified in ClinicalTrials.gov.

Appendix Figure B1. Search flow



Reasons for exclusion are presented in Appendix C. RCTs, randomized controlled trials.

Appendix Table B1. Studies published since 2006

Author, year	*Study short name (if available)*	*Sample size[a] PCI/ Surgery*	*Comparison*	*Population*	*Main findings*
Updates of studies included in the Stanford CER report					
Holper, 2007[7]	BARI[b]	1816 / 1317	Balloon vs. CABG	Patients with multivessel coronary artery disease with severe angina or objective evidence of ischemia requiring revascularization; CAD involving 2 or 3 vessels.	Patients with heart failure had increased mortality ten years after initial revascularization. In diabetic patients with heart failure and preserved LVEF there was a significant in cardiac mortality compared to patients without heart failure.
Holmes, 2007[6]	BARI[c]	1371 / 2239	Balloon vs. CABG	Patients with multivessel coronary artery disease with severe angina or objective evidence of ischemia requiring revascularization; CAD involving 2 or 3 vessels.	CABG was associated with significantly lower risk of sudden cardiac death but was not significantly associated with any other causes of long-term mortality at 7.7 years of follow-up.
BARI investigators, 2007[1]	BARI[d]	915 / 914	Balloon vs. CABG	Patients with multivessel coronary artery disease with severe angina or objective evidence of ischemia requiring revascularization; CAD involving 2 or 3 vessels.	PCI and CABG did not differ significantly regarding survival or angina rates at 10 years of follow-up. The PCI group had a substantially higher rate of revascularization. In a subgroup analysis by diabetes status, the CABG group had higher survival than the PCI group. Survival rates were identical among non diabetic subjects.
Martuscelli, 2008[10]	CABRI	120 / 103	Balloon vs. CABG	Patients with LAD, circumflex or right coronary artery chronic occlusion (subgroup of the CABRI trial)	Death or Q-wave MI was significantly lower in the CABG group at a median follow-up of 30 months.
Lopes, 2008[9e]	MASS-I and MASS-II	277 / 277	BMS (>80%) vs. CABG	Single or multivessel CAD (>70% stenosis)	CABG was significantly associated with lower rates of a composite endpoint of death, myocardial infarction or refractory angina compared to PCI at 5 years of follow-up.

[a] Patients for whom outcomes were reported in each study (intention-to-treat populations when data were available). Many of the studies reported subgroup analyses of larger randomized trials.
[b] Included patients both from the randomized (patients randomized to PCI or CABG) and the observational component of the BARI study (patients eligible for randomization who refused to be randomized were entered into a registry and underwent the same baseline investigations).
[c] Included patients both from the randomized (patients randomized to PCI or CABG) and the observational component of the BARI study (patients eligible for randomization who refused to be randomized were entered into a registry and underwent the same baseline investigations).
[d] Final 10-year follow-up data from the BARI study randomized component.
[e] We did not extract data regarding patients assigned to medical therapy (no revascularization).

Author, year	*Study short name (if available)*	*Sample size[a] PCI/ Surgery*	*Comparison*	*Population*	*Main findings*
Hueb, 2007[8]	MASS-II	205 / 2003	PCI (BMS, lasers, atherectomy, or balloon) vs. CABG	Angiographically documented proximal multivessel coronary stenosis (>70% stenosis) and documented ischemia	CABG was statistically significantly associated with lower rates of the composite endpoint of death, Q-wave MI or refractory angina requiring revascularization (primary endpoint). The difference in morality was nonsignificant at 5 years of follow up.
Favarato, 2007[4]	MASS-II	180 / 175	PCI (BMS, lasers, atherectomy, or balloon) vs. CABG	Angiographically documented proximal multivessel coronary stenosis (>70% stenosis) and documented ischemia	Both therapeutic strategies presented significant improvement in all dimensions of the SF-36 hrQoL questionnaire during follow-up. The CABG group had significantly greater improvement in physical and social functioning, vitality and general health when compared to medical treatment and PCI.[a]
Goy, 2008[5]	SIMA	62 / 59	BMS vs. CABG	Isolated proximal LAD stenosis and LVEF>45%	The incidence of death and MI were the same in the two groups at 10 years of follow-up. Patients in the PCI group had a significantly higher rate of the composite outcome of death, myocardial infarction or additional revascularization.
Booth, 2008[2]	SoS	488 / 500	BMS vs. CABG	Symptomatic multivessel CAD and at least one lesion suitable for stent implantation.	Patients in the CABG group had significantly improved survival at a median follow-up of 6 years. There was no significant interaction of the treatment effect with baseline angina grade, severity of coronary artery disease or diabetic status.
Buszman, 2009[3]	SoS	50 / 50[b]	BMS vs. CABG	Symptomatic multivessel CAD and at least one lesion suitable for stent implantation.	No significant difference was observed between the two groups at 10 years of follow-up. There was a significantly higher rate of repeat revascularizations and the composite outcome of death, MI, stroke or repeat revascularization in the PCI arm.
New studies (not included in the Stanford CER report)					
BARI 2D Study Group, 2009[15]	BARI 2D	1605 / 763[c]	Prompt coronary revascularization vs. medical therapy, randomization was stratified according to the method of revascularization (PCI vs. CABG) [a]	Patients with type 2 diabetes and CAD (≥50% stenosis of a major epicardial coronary artery associated with a positive stress test or ≥70% stenosis of a major epicardial coronary artery and classic angina)	In the PCI stratum there was no difference in the composite outcome of death, myocardial infarction or stroke between the revascularization and the medical therapy groups. In the CABG stratum, the rate of the composite outcome was significantly lower in the revascularization group than in the medical therapy group. This interaction between stratum and study group was statistically significant.

[a] The MASS-II trial had a three arm design, comparing PCI, CABG and medical therapy.
[b] Subgroup analysis of 100 patients enrolled in the SoS trial in centers in Poland. The original SoS trial included 988 patients and was conducted in 53 European and Canadian centers.
[c] Patients in the PCI / CABG randomization strata.

Author, year	*Study short name (if available)*	*Sample size[a] PCI/ Surgery*	*Comparison*	*Population*	*Main findings*
Chaitman[14]	BARI 2D	1605 / 763[b]	Prompt coronary revascularization vs. medical therapy, randomization was stratified according to the method of revascularization (PCI vs. CABG)	Patients with type 2 diabetes and CAD (≥50% stenosis of a major epicardial coronary artery associated with a positive stress test or ≥70% stenosis of a major epicardial coronary artery and classic angina)	In the CABG stratum, MI events and the composite outcomes of death or MI and cardiac death or MI were significantly less frequent in the revascularization plus intensive medical therapy group compared to the group that received intensive medical therapy alone at an average of 5.3 years of follow-up.
Kapur, 2010[16]	CARDia	256 / 254	BMS (32%) or DES (68%) vs. CABG	Patients with diabetes and either multivessel CAD or complex single vessel disease[c]	There was no statistically significant difference in mortality or a composite outcome of death, non-fatal MI, or non-fatal stroke at one year (primary outcome). There was a significant difference in favor of CABG in the composite outcome of death, non-fatal MI, non-fatal stroke, and repeat revascularization (secondary outcome). All comparisons were done at 1 year of follow-up.
Buszman, 2008[13]	LE MANS	52 / 53	BMS (65%) or DES (35%) vs. CABG	Unprotected left main CAD (>50% stenosis)	Patients in the PCI group had a significantly higher improvement in LVEF compared to patients in the CABG group at 1 year of follow-up (primary outcome). Patients in both groups performed equally well on stress tests and had similar improvements in angina status. At 1 year of followup, the rates of mortality and the composite outcome of cardiac death, MI, stroke, repeat intervention or in-stent thrombosis were comparable between the two arms. The difference in overall survival was non-significant at an average follow-up of 28 months.
Thiele, 2009[19]	NA	65 / 65	DES vs. MIDCAB	Isolated proximal LAD stenosis (>50% stenosis)	DES was non-inferior to MIDCAB at 12-month follow-up with respect to the composite outcome of cardiovascular death, MI or the need for repeated target vessel revascularization.
Serruys, 2009[18]	SYNTAX	903 / 897	DES vs. CABG	Three-vessel or complex left main CAD	CABG resulted in lower rates of the composite outcome of death from any cause, stroke, MI or repeat revascularization at 1 year compared to PCI.

[a] The BARI 2D trial had a 2x2 factorial design: in the first strategy patients were randomized to undergo either prompt coronary revascularization or medical therapy; in the second strategy patients were assigned to either insulin-sensitization therapy or insulin provision therapy to achieve a target glycated hemoglobin of less than 7.0%. We only abstracted data regarding the revascularization strategies.

[b] Patients in the PCI / CABG randomization strata.

[c] Complex disease was defined as ostial or proximal left anterior descending coronary artery disease.

Author, year	*Study short name (if available)*	*Sample size[a] PCI/ Surgery*	*Comparison*	*Population*	*Main findings*
Morice, 2010[17]	SYNTAX	357 / 348	DES vs CABG	Left main CAD (subgroup analysis of Serruys, 2009[18])	PCI and CABG had comparable outcomes in regards to the composite outcome of all-cause death, cerebrovascular accident/stroke, MI or repeat revascularizations at one year of follow-up.
Banning, 2010[12]	SYNTAX	903 / 897	DES vs CABG	Three-vessel or complex left main CAD	Among diabetic patients, patients on the CABG group had lower rates of the composite outcome of death from any cause, stroke, MI or repeat revascularization compared to patients on the PCI group at 1 year of follow-up. The difference was not significant among non-diabetic patients. The repeat revascularization rate was higher on the PCI arm compared to the CABG arm, regardless of diabetes status.

BMS=bare metal stent; CABG=coronary artery bypass grafting surgery; CAD=coronary artery disease; DES=drug-eluting stent; EPC=Evidence-based Practice Center; hrQoL=health-related quality of life; LAD=left anterior descending; LVEF=left ventricular ejection fraction; MI=myocardial infarction; MIDCAB=minimally invasive direct coronary artery bypass; NA=not available; PCI=percutaneous coronary intervention. Studies are separated by whether they were updates of studies in included in the Stanford CER or presented their first results subsequent to the report. Studies are listed alphabetically; when multiple publications are available from the same study they are listed by year of publication and then by author name.

The five RCTs published after the completion of the Stanford CER generally used DES in their PCI arms, either exclusively, or in combination with BMS. None included balloon angioplasty. Two of the 5 RCTs included well over 1,500 patients,[15,18] one included approximately 500 patients[16] and the remaining two less than 200.[3,19] Two of the largest RCTs enrolled only diabetic patients: CARDia compared DES or BMS with CABG[16] while BARI 2D[15] randomized patients to either prompt revascularization with intensive medical therapy vs. intensive medical therapy alone but stratified randomization by the choice of revascularization method (PCI vs. CABG).

Appendix Table B2. Ongoing studies comparing PCI vs. CABG (from clinicaltrials.gov)

Study	*NCT number*	*Size*	*Comparison*	*Population*	*Completion*
FREEDOM	NCT00086450	2400	PCI vs CABG	Diabetes (T1 or 2)	[soon]
Leipzig	NCT00176397	200	DES vs CABG	Left main	[soon]
VACARDS	NCT00326196	790	DES vs CABG	Diabetes (severe ischemic disease)	Terminated
REHEAT2	NCT00388245	150	PCI vs CABG	Ischemic cardiomyopathy; low left ventricular ejection fraction	[soon]
China	NCT01035034	400	DES vs Hybrid	2-3 vessel disease with left main involvement and denovo lesions of left anterior descending artery	12/2012

DES=drug eluting stents; Hybrid=Hybrid PCI/MIDCAB intervention; MIDCAB=minimally invasive CABG.

Evidence Gaps After One-to-One Interviews with the Key Informants (List 2)

The list is presented as stem questions and options to complete the questions. In the comment columns, we list ongoing studies identified in ClinicalTrials.gov.

List 2: Interim expanded list of evidence gaps

Populations

Imagine that we can order a well designed and conducted study that provides a conclusive answer to the following question.

Stem question		***Subgroup***	***Comment***
For patients with coronary artery disease who may be eligible for both PCI or CABG: What is the comparative effectiveness or safety of PCI vs. CABG in ***THIS SUBGROUP***?	A.	Diabetes	• Lancet MIPD; FREEDOM (n=2400); VACARDS (n=790)
	B.	Older adults (>71 or >75)	•
	C.	High body mass index	•
	D.	LV ejection fraction<35% or ischemic heart failure	• Lancet MIPD; REHEAT2 (n=150)
	E.	Renal disease (stage 3 or 4), prior to replacement therapy	•
	F.	~~Left Main disease~~	• SYNTAX (n=1800); Leipzig_1 (n=200); China_2 (n=400)
	G.	Women	•
	H.	Race (Asian? Smaller body size?)	•
	I.	Prior CABG	•
	J.	Prior PCI	•

Crossed out font: suggested as a nonmajor gap in key informant interviews (CABG consistently better).

Interventions

Imagine that we can order a well designed and conducted study that provides a conclusive answer to the following question.

Stem question		***Treatment variant***	***Comment***
For patients with coronary artery disease who may be eligible for both PCI or CABG: What is the comparative effectiveness or safety of ***THIS_PCI_VARIANT*** vs. ***THIS_CABG_VARIANT***?	A.	PCI: Bare metal stents	•
	B.	PCI: Drug eluting stents	• SYNTAX (n=1800); Leipzig_1 (n=200); VACARDS (n=790) China_2 (n=400); Leipzig_2 (n=130);
	C.	~~CABG: Minimally invasive surgery [MIDCAB]~~	• Leipzig_2 (n=130);
	D.	~~CABG: Hybrid minimally invasive surgery with PCI~~	• China_2 (n=400); • *For Hybrid vs CABG:* China_1 (n=400); POL-MIDES (n=200)
	E.	Medical treatment	• *[If the topic were broader, an important set of questions has to do with Medical vs invasive treatment…]*

Crossed out font: suggested as a nonmajor gap in key informant interviews (not routinely used).

Testing To Inform Treatment Decisions

Imagine that we can order a well designed and conducted study that provides a conclusive answer to the following question.

Stem question		***Test to guide treatment choice***	***Comment***
For patients with coronary artery disease who may be eligible for both PCI or CABG: Does ***THIS_TESTING*** have the ability to predict differential response to PCI or CABG? ~~What is the impact of testing with ***THIS_TEST*** for guiding treatment choice (PCI vs. CABG) on patient-relevant outcomes?~~	A.	Stress test	• [STICH-DECIPHER uses dobutamine echocardiography but measured also several other tests – but CABG vs medical therapy]
	B.	SPECT	• WOMEN study
	C.	MR angiography	•
	D.	CT angiography	• [faCTor64 screening of asymptomatic people]; ~~CTPRIME (terminated)~~
	E.	~~Calcification score~~	•
	F.	~~SYNTAX score~~	•

Crossed out font: based on key informant input it was deemed that the second stem question (leftmost column) depends on the question above it, which should be addressed first. In the second column, key informants suggested dropping these tests; in the last column, the CTPRIME study was terminated.

[Comment: The question refers to testing to inform treatment choice. We are not asking the important question of testing to screen asymptomatic individuals for CAD, to choose between e.g., no treatment, medical treatment, or some kind of revascularization. The screening question is deemed to be too peripheral to our topic.]

Preferences of Patients or Referring Physicians, and Decision Aids

Imagine that we can order a well designed and conducted study that provides a conclusive answer to the following question.

	Question	***Comment***
A.	For *patients* with coronary artery disease who may be eligible for both PCI or CABG: What are the preferences of the patients regarding treatment choice, and which factors influence them?	•
B.	For *(primary care) physicians whose patients have* coronary artery disease and may be eligible for both PCI or CABG: What are the preferences of the referring physician regarding treatment choice, and which factors influence them?	•
C.	For *patients* with coronary artery disease who may be eligible for both PCI or CABG, and their referring physicians: Develop a decision support tool to help physicians and patients in their shared decisionmaking regarding treatment choice.	•

Study of Inequalities and Performance Measures

Imagine that we can order a well designed and conducted study that provides a conclusive answer to the following question.

	Question	***Comment***
A.	How can we develop suitable performance measures for institutions and for individual physicians (CABG surgeons or PCI interventionists)	•
B.	For *patients* with coronary artery disease who may be eligible for both PCI or CABG: What are the major factors that contribute to inequalities in accessing health care?	•

Appendix B. Step 3: Final List of Evidence Gaps After Internal Prioritization (List 3)

See the "Background information provided to key informants" (pages 14–15 of this Appendix) in the next section for a justification of our selection of the items in the final list of evidence gaps (List 3).

List 3. Pruned (final) list of important evidence gaps

Thematic area	*List of evidence gaps*

Thematic area 1: Comparative effectiveness of PCI vs. CABG	*Populations:*
	• Age >75 years
	• Prior PCI
	• Diabetes
	• Women
	• Congestive heart failure
	• Stage 3 or 4 of chronic kidney disease
	Interventions:
	• PCI: Bare metal stents
	• PCI: Drug eluting stents
	• CABG: On pump traditional CABG with arterial grafts
	Outcomes:
	• 30-day outcomes -objective:
	  ○ Periprocedural death
	  ○ Nonfatal MI
	  ○ Nonfatal stroke
	  ○ Unplanned urgent revascularization with CABG during the PCI
	  ○ Renal failure requiring renal replacement therapy
	  ○ Post procedural discharge to rehabilitation facilities
	  ○ Health care costs
	  ○ Readmission rates for cardiac reasons such as heart failure and unstable angina
	  ○ Others like nosocomial infections are most relevant to CABG, and pulmonary complications affect the length of stay
	• Intermediate outcomes (objective) at 1 year. The relative ranking would be similar to the applicable 30 day outcomes.
	• Longer term outcomes (objective) would be 5 years. The relative ranking would be similar to the applicable 30 day outcomes.
	• Subjective outcomes:
	  ○ QoL is very important and can be measured by generic or disease-specific instruments

Thematic area 2: Testing to inform choice of revascularization procedure	The evidence gap pertains to whether testing before revascularization can guide the choice of revascularization procedure, e.g., by predicting clinical response in the long term. Examples of invasive and non-invasive tests: • Invasive coronary arteriography • Non-invasive CT angiography or MR angiography • Resting or exercise SPECT • Exercise treadmill test with or without echocardiography
Thematic area 3: Enhancing patient participation	• Eliciting and measuring patient preferences • Facilitating shared decisionmaking between patients and their physicians by developing and then evaluating decision support tools.
Thematic area 4: Assessing performance	• Based on prior key informant input, there are no validated process-based performance measures that could quantify optimal care

BMS=bare metal stents; CABG=coronary artery bypass grafting; CTA=computerized tomography-based angiography; DES=drug eluting stents; MI=myocardial infarction; MRA=magnetic resonance angiography; PCI=percutaneous coronary intervention; QoL=quality of life; SPECT=single photon emission computerized tomography

Appendix B. Step 4: Background Information Provided to Key Informants

Relative Disease Burden of CAD Subpopulations

In the teleconference, and in one-to one calls with several of the key informants we identified a list of factors that define subpopulations of interest for the comparative effectiveness of PCI vs. CABG. Appendix Table B3 summarizes data on the prevalence of these factors from 4 analyses of large clinical registries of patients with coronary artery disease who received revascularization (identified from the Stanford CER[20]). These included patients treated in the 1990's in North America or Europe, describe at least 1,000 patients treated with PCI and 1,000 treated with CABG, report patient characteristics and performed multivariate statistical analyses.

Broadly, percentages in Appendix Table B3 fall into similar ranges for most factors. Obviously there are differences that can be ascribed to difference in settings, time period—type of PCI or CABG, eligibility criteria for the analyses, or health policy differences across countries. The rows of the Table are ordered by decreasing percentages in Hannan et al.[21] Hannan et al. is the largest of the four, and is based on two comprehensive registries from the state of New York. The percentages in this study are a crude index of the relative "disease burden" in CAD subpopulations defined by the pertinent factors. The five top factors are old age (>75 years), prior PCI revascularization, diabetes, female sex, and congestive heart failure.

Especially for renal disease, the registries report only percentage with creatinine >2 or >2.5 mg/dL. In all likelihood this corresponds to stage 4 or stage 5 (renal replacement therapy). Stage 3 will be quite more frequent, and therefore we include renal disease (stage 3 or 4) as a top 6th subpopulation.

Data on Treatment Effect Modification in CAD Subpopulations

The meta-analysis of individual patient data by Hlatky 2009[22] found that age (3 categories using cutoffs at 55 and 65 years) and diabetes modify the effects of treatment on survival (p=0.002 and 0.014, respectively). The magnitude of the treatment effect modification was clinically important. They found no significant effect modification for sex (p=0.25), heart failure (p=0.46), or abnormal LV function (p=0.87). Racial descent, previous PCI or CABG, body mass index, renal function and left main disease were not examined.

Ongoing Trials of PCI vs. CABG in CAD Subpopulations

There are two ongoing trials (FREEDOM, n=2400 and VACARDS, n=790) that enrolled diabetics and their results will likely be presented within the next 2-3 years. Other ongoing trials are listed in Table B2.

Criteria to Prioritize Research Questions

Appendix Table B4 lists the criteria that we use to rate the future research needs for PCI vs. CABG. Briefly, we use 6 criteria:

Disease burden; impact on practice; impact on patient health; impact on patient satisfaction; address health inequalities; address ethical, legal or social issues (ELSI); costs

Appendix Table B3. Characteristics of patients who underwent PCI or CABG in selected large clinical registries

	Hannan et al. 2005[21]	Malenka et al. 2005[23]	Pell et al. 2001[24]	Dzavik et al. 2001[25]
Description	No prior revascularization, no LM (>50%)	Multivessel disease, <80y, no prior revascularization, no LM (>50%)	Patients who received the interventions	CAD patients who were revascularized
Sample size	59314	14493	9890	22690
Years	1997-2000	1994-2001	1997-1999	1995-1998
Location	NY Registries, US	Northern New England, US	Scotland	Alberta, Canada
Percentages				
Age >75 years	41	33[a]	5	<50
2nd revascularization (1st was PCI)[b]	35	ND	23	13
Diabetes	30	32	12	26
Women	30	28	28	22
Congestive heart failure	16	14	ND	17
Non-European descent (non-white)	12	ND	ND	ND
Left ventricular ejection fraction <30%	8	12	42	ND
Renal disease*	3[c]	[3][d]	ND	3[e]
2nd revascularization (1st was CABG)[f]	5	ND	2	13
Body mass index >30 kg/m^2	ND	ND	27	ND
Left main disease (e.g. >50% stenosis)	NE	NE	ND	ND

ND=no data; NE=not estimable

*Unfortunately, the percentage of patients who had chronic kidney disease stage 3 or 4 was not explicitly mentioned (what is shown corresponds to creatinine >2 or >2.5 mg/dL, or roughly, stages 4 or 5). Stage 3 is more prevalent—and therefore we believe that "renal disease stages 3 or 4" will be among the 6 most prevalent factors in Hannan et al.

[a] >70years but note that there is an upper cutoff of 80 years of age; however in other registries >80y is a small percentage (<5%).

[b] Calculated as proportion revascularized (during the followup) among those who received PCI.

[c] Renal failure (dialysis or creatinine >2.5 mg/dL).

[d] Renal failure or creatinine >2 mg/dL.

[e] Creatinine >200 mmol, approximately >200/88= 2.3 mg/dL but not on dialysis.

[f] Calculated as proportion revascularized (during the followup) among those who received CABG.

Table B4. Criteria to qualitatively rank future research needs

Criterion short name	Description	Comment
Disease burden	*The research question pertains to/corresponds to/addresses an important **disease burden**, or a **high-priority population**?*	• High prevalence of a condition suggests a high disease burden • An important subpopulation can be relatively rare
Impact on practice	*Answering the research question **has the potential to change practice***	• Physicians are likely to adopt the intervention in practice
Impact on patient health	*Answering the research question **has the potential to improve patients' health***	• If the study finds an effective intervention, can it affect patient health directly? • e.g., testing affects health indirectly e.g., through treatment choices
Impact on patient satisfaction	*Answering the research question **has the potential to improve patients' satisfaction***	• Relates to patient's preferences
Address health inequalities	*Answering the research question **has the potential to reduce health inequalities** without adverse impact on specific subpopulations*	• Variation in practice • Variation in access • Variation in measured performance (for institutions)
Address ethical, legal or social issues	*Answering the research question has the potential to allow assessment of **ethical, legal or social issues (ELSI)** pertaining to the disease*	• Malpractice • Minorities • Other ethical concerns
Costs	*Answering the research question will have **nontrivial impact on costs**/economic considerations*	• High unit cost • High volume • High opportunity costs

We will rate each research need in each criterion as:
1=I strongly disagree; 2=I disagree; 3=I am indifferent; 4=I agree; 5=I strongly agree.

Appendix B. Step 4: Questionnaire to Key Informants

I expect that you will need **less than 60 minutes** to provide an opinion on the research needs listed below. I would appreciate even a partly filled in reply.

Your name here: _____

Populations

Is the following question describing a future research need?
*"For patients with non-acute coronary artery disease who are eligible for both PCI and CABG: What is the comparative effectiveness and safety of PCI vs. CABG in **THIS SUBGROUP**?"*

Subgroup	***Overall, rank the subgroups using letters, A=most important, E=least important***	***Is this research need important in the criteria of Table A3?*** *[NA=I will not answer;1=I strongly agree; 2= I agree; 3= I am indifferent; 4=I disagree; 5=I strongly disagree]*						
		Disease burden	***Impact on practice***	***Impact on patient health***	***Impact on patient satisfaction***	***Address health inequalities***	***Address Ethical, Legal, Social, Issues***	***Costs***
Age >75 years								
Prior PCI								
Diabetes								
Women								
Congestive heart failure								
Stage 3 or 4 of chronic kidney disease								

Please provide any comments here:

Interventions

Is the following question describing a future research need?

*"For patients with non-acute coronary artery disease who are eligible for both PCI and CABG: What is the comparative effectiveness or safety of **THIS_PCI_VARIANT** vs. CABG?"*

PCI Variant	***Overall, rank the treatment variants using letters, A=most important, B=least important***	***Is this research need important in the criteria of Table A3?*** *[NA=I will not answer;1=I strongly agree; 2= I agree; 3= I am indifferent; 4=I disagree; 5=I strongly disagree]*						
		Disease burden	***Impact on practice***	***Impact on patient health***	***Impact on patient satisfaction***	***Address health inequalities***	***Address Ethical, Legal, Social, Issues***	***Costs***
Bare metal stents		NA						
Drug eluting stents		NA						

Please provide any comments here:

Testing To Predict Treatment Response

Is the following question describing an important future research need?

*"For patients with non-acute coronary artery disease who are eligible for both PCI and CABG: What is the ability of **THIS_TEST** (prior to revascularization) to predict differential treatment response to PCI or CABG, in terms of (predictive) sensitivity and specificity?"*

[Comment: We are not asking the important question of testing to screen asymptomatic individuals for CAD. The screening question is deemed to be too peripheral to our topic.]

Test	***Overall, rank the tests using letters, A=most important, D=least important***	***Is this research need important in the criteria of Table A3?*** *[NA=I will not answer;1=I strongly agree; 2= I agree; 3= I am indifferent; 4=I disagree; 5=I strongly disagree]*						
		Disease burden	***Impact on practice***	***Impact on patient health***	***Impact on patient satisfaction***	***Address health inequalities***	***Address Ethical, Legal, Social, Issues***	***Costs***
Invasive coronary arteriography		NA						
Non-invasive CT angiography or MR angiography		NA						
Resting or exercise SPECT		NA						
Exercise treadmill test with or without echocardiography		NA						

Please provide any comments here:

Patient Preferences and Decision Aids

Is the following an important future research need?

"For patients with non-acute coronary artery disease who are eligible for both PCI or CABG: Develop a decision support tool to help physicians and patients in their shared decisionmaking regarding treatment choice."

This would presuppose elicitation of patient preferences

Is this research need important in the criteria of Table A3? *[NA=I will not answer;1=I strongly agree; 2= I agree; 3= I am indifferent; 4=I disagree; 5=I strongly disagree]*						
Disease burden	***Impact on practice***	***Impact on patient health***	***Impact on patient satisfaction***	***Address health inequalities***	***Address Ethical, Legal, Social, Issues***	***Costs***
NA	NA					

Please provide any comments here:

Performance Measures

Is the following an important future research need?

"For patients with non-acute coronary artery disease who are eligible for both PCI or CABG: Develop suitable performance measures for institutions and for individual physicians (CABG surgeons or PCI interventionists)."

Is this research need important in the criteria of Table A3? *[NA=I will not answer;1=I strongly agree; 2= I agree; 3= I am indifferent; 4=I disagree; 5=I strongly disagree]*						
Disease burden	***Impact on practice***	***Impact on patient health***	***Impact on patient satisfaction***	***Address health inequalities***	***Address Ethical, Legal, Social, Issues***	***Costs***
NA						

Please provide any comments here:

Out of Scope Question: Adding Optimal Medical Therapy as an Alternative Treatment Option Alongside PCI and CABG

Optimal medical therapy was outside the scope of the Stanford CER, and therefore it was not included in this exercise. However, optimal medical therapy came up in our discussions. Please indicate how strongly you agree or disagree with adding optimal medical therapy in all aforementioned research needs, e.g., the "questions" would start:

"*For patients with non-acute coronary artery disease in who are eligible for PCI, CABG or optimal medical therapy: … "*

Answer (1 through 5)	

[NA=I will not answer;1=I strongly agree; 2= I agree; 3= I am indifferent; 4=I disagree; 5=I strongly disagree

Please provide any comments here:

Appendix B References

1. The final 10-year follow-up results from the BARI randomized trial. J Am Coll Cardiol 2007 Apr 17;49(15):1600-1606.
2. Booth J, Clayton T, Pepper J, et al. Randomized, controlled trial of coronary artery bypass surgery versus percutaneous coronary intervention in patients with multivessel coronary artery disease: six-year follow-up from the Stent or Surgery Trial (SoS). Circulation 2008 Jul 22;118(4):381-388.
3. Buszman P, Wiernek S, Szymanski R, et al. Percutaneous versus surgical revascularization for multivessel coronary artery disease: a single center 10 year follow-up of SOS trial patients. Catheter Cardiovasc Interv 2009 Sept 1;74(3):420-426.
4. Favarato ME, Hueb W, Boden WE, et al. Quality of life in patients with symptomatic multivessel coronary artery disease: a comparative post hoc analyses of medical, angioplasty or surgical strategies-MASS II trial. Int J Cardiol 2007 Apr 4;116(3):364-370.
5. Goy JJ, Kaufmann U, Hurni M, et al. 10-year follow-up of a prospective randomized trial comparing bare-metal stenting with internal mammary artery grafting for proximal, isolated de novo left anterior coronary artery stenosis the SIMA (Stenting versus Internal Mammary Artery grafting) trial. J Am Coll Cardiol 2008 Sept 2;52(10):815-817.
6. Holmes Jr DR, Kim LJ, Brooks MM, et al. The effect of coronary artery bypass grafting on specific causes of long-term mortality in the Bypass Angioplasty Revascularization Investigation. J Thorac Cardiovasc Surg 2007 Jul;134(1):38-46.
7. Holper EM, Brooks MM, Kim LJ, et al. Effects of heart failure and diabetes mellitus on long-term mortality after coronary revascularization (from the BARI Trial). Am J Cardiol 2007 Jul 15;100(2):196-202.
8. Hueb W, Lopes NH, Gersh BJ, et al. Five-year follow-up of the Medicine, Angioplasty, or Surgery Study (MASS II): a randomized controlled clinical trial of 3 therapeutic strategies for multivessel coronary artery disease. Circulation 2007 Mar 6;115(9):1082-1089.
9. Lopes NH, Paulitsch FS, Gois AF, et al. Impact of number of vessels disease on outcome of patients with stable coronary artery disease: 5-year follow-up of the Medical, Angioplasty, and bypass Surgery study (MASS). Eur J Cardiothorac Surg 2008 Mar;33(3):349-354.
10. Martuscelli E, Clementi F, Gallagher MM, et al. Revascularization strategy in patients with multivessel disease and a major vessel chronically occluded; data from the CABRI trial. Eur J Cardiothorac Surg 2008 Jan;33(1):4-8.
11. Vaina S, Voudris V, Morice MC, et al. Effect of gender differences on early and mid-term clinical outcome after percutaneous or surgical coronary revascularisation in patients with multivessel coronary artery disease: insights from ARTS I and ARTS II. EuroIntervention 2009 Jan;4(4):492-501.
12. Banning AP, Westaby S, Morice MC, et al. Diabetic and nondiabetic patients with left main and/or 3-vessel coronary artery disease: comparison of outcomes with cardiac surgery and paclitaxel-eluting stents. J Am Coll Cardiol 2010 Mar 16;55(11):1067-1075.
13. Buszman PE, Kiesz SR, Bochenek A, et al. Acute and late outcomes of unprotected left main stenting in comparison with surgical revascularization. J Am Coll Cardiol 2008 Feb 5;51(5):538-545.
14. Chaitman BR, Hardison RM, Adler D, et al. The Bypass Angioplasty Revascularization Investigation 2 Diabetes randomized trial of different treatment strategies in type 2 diabetes mellitus with stable ischemic heart disease: impact of treatment strategy on cardiac mortality and myocardial infarction. Circulation 2009 Dec 22;120(25):2529-2540.

15. Frye RL, August P, Brooks MM, et al. A randomized trial of therapies for type 2 diabetes and coronary artery disease. N Engl J Med 2009 Jun 11;360(24):2503-2515.
16. Kapur A, Hall RJ, Malik IS, et al. Randomized comparison of percutaneous coronary intervention with coronary artery bypass grafting in diabetic patients. 1-year results of the CARDia (Coronary Artery Revascularization in Diabetes) trial. J Am Coll Cardiol 2010 Feb 2;55(5):432-440.
17. Morice MC, Serruys PW, Kappetein AP, et al. Outcomes in patients with de novo left main disease treated with either percutaneous coronary intervention using paclitaxel-eluting stents or coronary artery bypass graft treatment in the Synergy Between Percutaneous Coronary Intervention with TAXUS and Cardiac Surgery (SYNTAX) trial. Circulation 2010 Jun 22;121(24):2645-2653.
18. Serruys PW, Morice MC, Kappetein AP, et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. N Engl J Med 2009 Mar 5;360(10):961-972.
19. Thiele H, Neumann-Schniedewind P, Jacobs S, et al. Randomized comparison of minimally invasive direct coronary artery bypass surgery versus sirolimus-eluting stenting in isolated proximal left anterior descending coronary artery stenosis. J Am Coll Cardiol 2009 Jun 23;53(25):2324-2331.
20. Bravata D, McDonald K, Gienger A, et al. Comparative effectiveness of percutaneous coronary interventions and coronary artery bypass grafting for coronary artery disease. Comparative Effectiveness Review No 9 (prepared by Stanford UCSF-EPC) 2007.
21. Hannan EL, Racz MJ, Walford G, et al. Long-term outcomes of coronary-artery bypass grafting versus stent implantation. N Engl J Med 2005 May 26;352(21):2174-2183.
22. Hlatky MA, Boothroyd DB, Bravata DM, et al. Coronary artery bypass surgery compared with percutaneous coronary interventions for multivessel disease: a collaborative analysis of individual patient data from ten randomised trials. Lancet 2009 Apr 4;373(9670):1190-1197.
23. Malenka DJ, Leavitt BJ, Hearne MJ, et al. Comparing long-term survival of patients with multivessel coronary disease after CABG or PCI: analysis of BARI-like patients in northern New England. Circulation 2005 Aug 30;112(9 Suppl):I371-I376.
24. Pell JP, Walsh D, Norrie J, et al. Outcomes following coronary artery bypass grafting and percutaneous transluminal coronary angioplasty in the stent era: a prospective study of all 9890 consecutive patients operated on in Scotland over a two year period. Heart 2001 Jun;85(6):662-666.
25. Dzavik V, Ghali WA, Norris C, et al. Long-term survival in 11,661 patients with multivessel coronary artery disease in the era of stenting: a report from the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) Investigators. Am Heart J 2001 Jul;142(1):119-126.

Appendix C. Excluded Studies

Appendix Table C1. List of excluded studies and reasons for exclusion

First author	Year of publication	Title	Journal	PMID	Reason for exclusion
Bypass Angioplasty Revascularization Investigation 2 Diabetes Study Group	2008	Baseline characteristics of patients with diabetes and coronary artery disease enrolled in the Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) trial	Am Heart J	18760137	No follow-up information
M. Y. Chan	2008	Prevalence, predictors, and impact of conservative medical management for patients with non-ST-segment elevation acute coronary syndromes who have angiographically documented significant coronary disease	JACC Cardiovasc Interv	19463332	Not relevant interventions
E. L. Eisenstein	2009	Long-term clinical and economic analysis of the Endeavor zotarolimus-eluting stent vs. the cypher sirolimus-eluting stent: 3-year results from the ENDEAVOR III trial (Randomized Controlled Trial of the Medtronic Endeavor Drug [ABT-578] Eluting Coronary Stent System Versus the Cypher Sirolimus-Eluting Coronary Stent System in De Novo Native Coronary Artery Lesions).	JACC Cardiovasc Interv	20129546	Not relevant interventions
M. E. Farkouh	2008	Design of the Future REvascularization Evaluation in patients with Diabetes mellitus: Optimal management of Multivessel disease (FREEDOM) Trial	Am Heart J	18215589	Trial protocol
Y. Y. Liu	2009	[Comparison between drug eluting stent and coronary artery bypass grafting surgery for the treatment of unprotected left main coronary artery disease in elderly patients]	Zhonghua Xin Xue Guan Bing Za Zhi	20128370	Not English
J. L. Martin	2009	Frequency of coronary artery bypass grafting following implantation of a paclitaxel-eluting or a bare-metal stent into a single coronary artery	Am J Cardiol	19101222	Not relevant interventions
A. C. Pereira	2006	Clinical judgment and treatment options in stable multivessel coronary artery disease: results from the one-year follow-up of the MASS II (Medicine, Angioplasty, or Surgery Study II).	J Am Coll Cardiol	16949484	Results from analyses comparing randomization arms not presented.
D. Poldermans	2007	A clinical randomized trial to evaluate the safety of a noninvasive approach in high-risk patients undergoing major vascular surgery: the DECREASE-V Pilot Study	J Am Coll Cardiol	17466225	Not relevant interventions and population
L. Schwartz	2009	Baseline coronary angiographic findings in the Bypass Angioplasty Revascularization Investigation 2 Diabetes trial (BARI 2D)	Am J Cardiol	19231325	No follow-up information
K. T. Stroupe	2006	Cost-effectiveness of coronary artery bypass grafts vs. percutaneous coronary intervention for revascularization of high-risk patients	Circulation	16966588	Cost-effectiveness analysis based on RCT
P. J. Vlaar	2008	Cardiac death and reinfarction after 1 year in the Thrombus Aspiration during percutaneous coronary intervention in Pcute myocardial infarction Study (TAPAS): a 1-year follow-up study	Lancet	18539223	Not relevant population

First author	Year of publication	Title	Journal	PMID	Reason for exclusion
H. B. Ward	2006	Coronary artery bypass grafting is superior to percutaneous coronary intervention in prevention of perioperative myocardial infarctions during subsequent vascular surgery.	Ann Thorac Surg	16928485	CABG – PCI comparison was not randomized
R. C. Welsh	2010	Prior coronary artery bypass graft patients with ST-segment elevation myocardial infarction treated with primary percutaneous coronary intervention	JACC Cardiovasc Interv	20298996	Not relevant population
S. Vaina	2009	Effect of gender differences on early and mid-term clinical outcome after percutaneous or surgical coronary revascularisation in patients with multivessel coronary artery disease: insights from ARTS I and ARTS II	Euro-Intervention	19284072	Comparison of outcomes between the arms of a randomized trial (ARTS I) and an observational study (ARTS II)

CABG=coronary artery bypass graft surgery; PCI=percutaneous coronary intervention; PMID=PubMed identification number.

Appendix D. Methods and Results of Quantitative Analyses

Appendix D Contents

Rationale for Our Quantitative Approach

The aim of the quantitative approach was to construct a simple mathematical model, perform 3 types of analyses (decision analysis, cost-effectiveness analysis and value of information analysis) and quantify the most influential parameters in the model based on the analyses. For decision and cost-effectiveness analyses we prioritize based on one way sensitivity analysis. In value of information analyses we prioritize groups of parameters by calculating the expected value of perfect information for parameters.

Regarding the decision and cost effectiveness analyses, we note that *we do not use modeling to identify the optimal treatment choice*, e.g., the choice that maximizes quality adjusted life years in the decision analysis or the choice that optimizes the balance of costs and effectiveness in cost-effectiveness analysis. We *use the models as tools to identify where future research is needed the most, albeit indirectly*: presumably, we would recommend further research to inform on the parameters that are most uncertain in the one-way sensitivity analyses in the decision and cost-effectiveness models.

This is an *atypical use of modeling*—and stops short of providing what would normally be considered a most important piece of information, namely insights on treatment choices for different assumptions and circumstances. By their very nature our models are not developed with the same rigor as some of the well-known and elaborate models that are used to analyze clinical decisions; they have not been calibrated using external data; and their predictions have not been validated in external data. They are operational models that are constructed to make use of the summary information obtained from an evidence report, and can offer only broad insights. However, we believe that even if they have not been vetted enough to analyze treatment choices, they are able to serve the purpose for which they have been developed, that is *rank the model parameters* which are obtained from the current state of science, the evidence report, according to the uncertainty they exert on the quantity of interest.

Our third type of analysis, the value of information analysis, is a theoretically motivated way to perform such "prioritizing" of future research. Again, we use the value of information analyses to rank groups of parameters based on their expected value of perfect information; we do not focus on the actual analyses results. As above, we believe that although by ranking we lose a lot of potential information from the analyses we performed, we gain by having results that are less likely to change if we use better models.

Methods for Decision Modeling, Cost-Effectiveness Analysis, and Value of Information Analysis

During the construction of the model it was our intent to maximize the use of data from the Stanford CER report. When necessary we considered additional information preferably from recently published studies that are applicable to the US setting. This is an operational approach that could be routinely followed as part of the preparation of future research needs products following completion of a CER report.

Briefly, we followed an operational process to develop a simple decision model that compares different treatment options (PCI with BMS, PCI with DES, CABG). We aimed to build a probabilistic decision model. Such models allow for a quantitative expression of the uncertainty around model parameters: appropriate distributions for model parameters are selected using standard methods and Monte Carlo simulation methods are used to sample from those

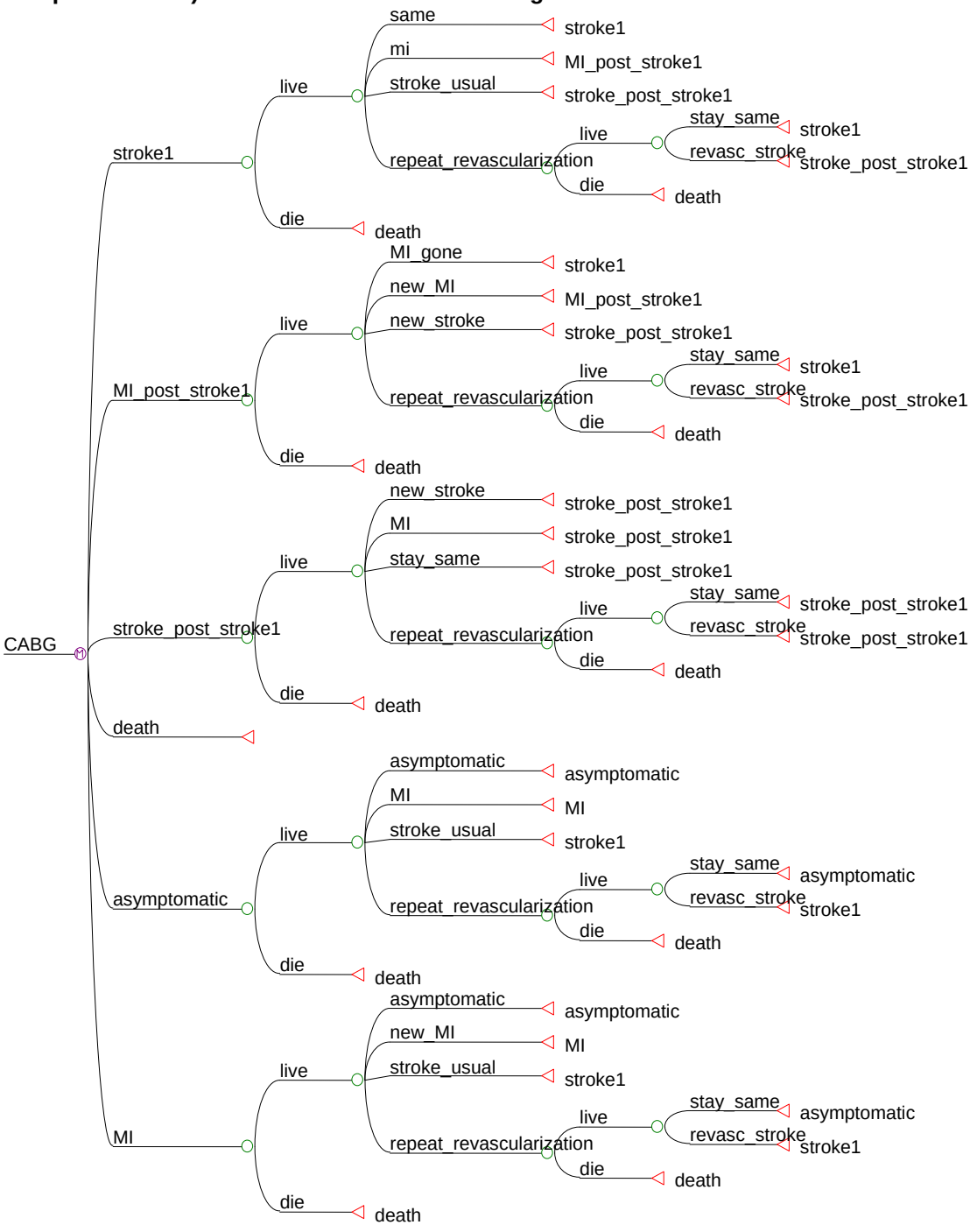
distributions and propagate the uncertainty through the decision model, providing the uncertainty around model estimates of effectiveness and costs.[1,2] Probabilistic sensitivity analysis is currently recommended for all decision and cost-effectiveness analyses.[2,3]

We began the modeling exercise by attempting to identify decision models evaluating a similar decisional context by searching a curated database of cost-effectiveness analyses maintained at the Tufts Medical Center[a] (Cost-effectiveness Analysis Registry, Tufts Medical Center, Boston, MA, last accessed August 25, 2010) for relevant published studies. As expected, several decision models are available related to the choice of coronary artery bypass graft surgery (CABG) vs. percutaneous coronary intervention (PCI). We originally attempted to replicate the model presented by Rao et al,[4] because it was published relatively recently, assumed a societal perspective, conducted a probabilistic sensitivity analyses and reported all data that were necessary to replicate the original results. We were able to perfectly replicate their results regarding treatment effectiveness (to the third decimal, i.e. within the margin of random error) and had moderate success in replicating their results regarding costs (we could not resolve a 10% discrepancy in costs). Because Rao et al.[4] used estimates applicable to the United Kingdom's National Health System (NHS), and we wanted our model to be applicable to a typical USA practice setting, we proceeded by adapting the model to incorporate elements from the analyses reported in Yock et al.[5] and Bischof et al.[6] This required a simplification of the model, specificaly the consideration of different repeat revascularization procedures in aggregate (instead of modeling repeat revascularization with CABG or PCI). Analyses of expected utility using the modified model provided results similar to those in Rao et al.[4]

We considered three treatment alternatives to be of interest: CABG, PCI with bare metal stents (BMS) and with three potential outcomes after the initial revascularization procedure: becoming asymptomatic, experiencing a procedural stroke and experiencing treatment complications resulting in procedural death. Following the initial procedure, for surviving patients, we considered a 6-state Markov model. The states we modeled were (1) asymptomatic (no prior stroke), (2) recent myocardial infarction (no prior stroke), (3) asymptomatic (post stroke), (4) recent myocardial infarction (post stroke), (5) repeat stroke in patients who had a prior stroke; (6) dead. Appendix Figure D1 presents a schematic of the Markov model structure we used for all analyses described herein.

[a] Cost-effectiveness Analysis Registry, maintained at the Center for the Evaluation of Value and Risk in Health, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center. Available at https://research.tufts-nemc.org/cear/default.aspx.

Appendix Figure D1. Markov model structure for the CABG arm. The same structure (with different transition probabilities) was used to model PCI strategies as well



We considered three types of populations in 3 different models for three different types of populations.

- "RCT-type participants:" a base case scenario where a cohort of 65 year olds with typical age, sex and race-related mortality for the USA population. The decision is between

revascularization with DES, BMS and CABG. Most or our data were drawn from randomized clinical trials RCTs, and were used to parameterize this model.

- "Elderly participants (older than 75 years):" a cohort of elderly (75 year old) CAD patients with nonacute CAD. The decision is between revascularization with DES, BMS and CABG.
- "Diabetics:" a cohort of 65-year old diabetic patients. The decision is between revascularization with DES and CABG.

All analyses were carried out with a 10 year time horizon, with one year Markov cycles. Costs and effects were discounted at 3% per year. Statistical analyses were conducted using Stata version SE 11.1 (Stata Corp. College Station, TX) and decision, cost-effectiveness and value of information analyses were conducted using specialized software (TreeAge Pro Healthcare, Williamstown, MA). For probabilistic decision and cost-effectiveness analyses, as well as expected value of perfect information (EVPI) analyses we based our calculations on 10,000 model iterations. For expected value of partial perfect information (EVPPI) we based our calculations on 1000 iterations of the external sampling loop and 1000 iterations of the internal loop.[7,8]

Model Parameters

Baseline probabilities for the PCI BMS arm: We used the recently published COURAGE trial to obtain probabilities of both procedural and long term outcomes in patients undergoing PCI with BMS.[9] Cumulative event rates for stroke, myocardial infarction and revascularization (during a follow-up of 4.6 years) were converted to annual event rates from which we estimated transition probabilities using standard methods.[1] In addition we calculated the excess mortality due to CAD based on the difference in mortality rates between the average 65 year old USA population and the COURAGE trial.[9]

Treatment effects: To fully utilize the data available from the Stanford CER, we extracted data from the report pertaining to the following outcomes: procedural strokes, procedural mortality, deaths, myocardial infarctions and mortality. We only extracted data from studies that used stents (either BMS or DES) in their PCI arms, since we considered studies of balloon angioplasty not to be reflective of current medical practice. Using these data, we performed meta-analyses of all trials using random effects models and relative risks (RR) as the metric of choice for the four outcomes of interest.[a] Since the number of events for all outcomes was low and the between group differences were small, the relative risk is a good approximation for the hazard ratio (HR), and can thus be applied to event rates. To obtain estimates of treatment effects comparing DES and CABG we used data from the recently published SYNTAX trial.[10] Finally, to obtain estimates of the treatment effect comparing BMS and DES we used a recently published network meta-analysis of treatments for non-acute CAD.[11] To obtain consistent estimates of treatment effects for all three treatments of interest (DES, BMS, CABG) we performed a mixed treatment meta-analysis of the treatment effect estimates obtained from pairwise comparisons of the three strategies.[12]

Utilities: We obtained utility estimates from Rao et al,[4] based on a National Institute for Health and Clinical Excellence (NICE) report on coronary revascularzation.[13] These estimates

[a] In parameterizing the model we treated the relative risks as hazard ratios. This approximation is generally valid, as evident from the Taylor expansions of the complementary log log transformation (which is the link function for hazard ratios) and the log transformation (the link function for relative risks).

were derived from the ARTS clinical trial, using the EQ-5D instrument.[14] To ensure the validity of these estimates, we used the "Utility Weights" function of the CEA Registry (see above) to obtain a comprehensive list of QALY estimates used in cost-effectiveness analyses exploring similar decisional contexts and compared our estimates with those used in published analyses. In all cases the estimates we used were close to the average of estimates used by others, providing reassurance of the representativeness of our analyses. Further the uncertainty of the utility weights in Rao et al. corresponded to the scatter of the utility weight distribution from the CEA Registry.

Costs: Costs were obtained from published sources, relevant to the USA clinical setting. Specifically, we obtained cost estimates from two recent cost-effectiveness analyses by Yock et al.[5] and Bischof et al.[6] To estimate the cost of stroke we used the Diagnosis Related Group (DRG) weights from the Centers for Medicare and Medicaid Services (fiscal year 2007) following the calculation described in Bischof et al.[6] to estimate the costs of stroke (DRG code 559, Acute ischemic stroke with use of thrombolytic agent).

Appendix Table D1 summarizes details of our data sources, Appendix Table D2 demonstrates how we obtained cost estimates for repeat revascularization procedures and Appendix Table D3 presents the parameters estimates we used in the three decision scenarios we explored.

Appendix Table D1. Data sources for model parameters

Model parameter	Estimation method and data sources		
	Typical RCT population	**Elderly**	**Diabetics**
Transition probabilities			
Procedural stroke in PCI	Event probabilities in the PCI arm of the COURAGE trial[9]	As in "Typical RCT population"	As in "Typical RCT population"
Procedural death in PCI			
MI in PCI patients during follow up	Cumulative event rates in the PCI arm of the COURAGE trial (4.6 years of median follow-up)[9] converted to annual transition probabilities[1]	As in "Typical RCT population"	Cumulative event probability in the PCI arm of the CARDia trial[15]
Stroke in PCI patients during follow up			
Revascularization in PCI patients during follow up			
Death among PCI patients during follow-up			
Death after stroke	10 year cumulative event proportions from Rao et al.[4]	As in "Typical RCT population"	As in "Typical RCT population"
Death after MI			
Treatment effects (log transformed RR)			
RR of procedural stroke (BMS/DES vs. CABG)	Meta-analysis of data from the Stanford CER[16]	As in "Typical RCT population"	As in "Typical RCT population"
RR of procedural death (BMS/DES vs. CABG)			
RR of MI (BMS vs. CABG)	Mixed treatment meta-analysis of data from the Stanford CER,[16] Trikalinos et al.,[11] and the SYNTAX trial[10]	Mixed treatment meta-analysis of data from the Stanford CER,[16] Trikalinos et al.,[11] and the SYNTAX trial,[10] with inflated (doubled) standard errors to account for increased uncertainty	Not modeled
RR of stroke (BMS vs. CABG)			
RR of revascularization (BMS vs. CABG)			

Model parameter	Estimation method and data sources		
	Typical RCT population	**Elderly**	**Diabetics**
RR of death (BMS vs. CABG)		Mixed treatment meta-analysis of data from Hlatky et al,[17] Trikalinos et al.,[11] and the SYNTAX trial,[10] with inflated (doubled) standard errors to account for increased uncertainty	
RR of MI (DES vs. CABG)		Mixed treatment meta-analysis of data from the Stanford CER,[16] Trikalinos et al.,[11] and the SYNTAX trial,[10] with inflated (doubled) standard errors to account for increased uncertainty	Based on the CARDia trial[15]
RR of stroke (DES vs. CABG)			
RR of revascularization (DES vs. CABG)			
RR of death (DES vs. CABG)		Mixed treatment meta-analysis of data from Hlatky et al.,[17] Trikalinos et al.,[11] and the SYNTAX trial,[10] with inflated (doubled) standard errors to account for increased uncertainty	From Hlatky et al. meta-analysis of IPD[17]
Utilities			
Asymptomatic	Rao et al.[4]	As in "Typical RCT population"	As in "Typical RCT population"
MI			
Stroke			
Recurrent stroke			
MI post stroke			
Costs			
Cost of PCI with DES	From Bischof et al.[6]	As in "Typical RCT population"	As in "Typical RCT population"
Cost of PCI with BMS			
Cost of CABG			
Cost of repeat revascularization	From Yock et al.[5] taking into account the frequency of PCI/ CABG revascularization in the COURAGE trial,[9] as shown in Appendix D Table 2.	As in "Typical RCT population"	As in "Typical RCT population"
Cost of follow-up (including drug therapy) in PCI	From Yock et al.[5]	As in "Typical RCT population"	As in "Typical RCT population"
Cost of follow-up (including drug therapy) in CABG			
Cost of clopidogrel (for DES revascularization)	Modeled as a fixed increment of PCI costs, based on Filion et al.[18]	As in "Typical RCT population"	As in "Typical RCT population"
Cost of MI	From Bischof et al.[6]	As in "Typical RCT population"	As in "Typical RCT population"
Cost of stroke	Based on MEDICARE DRG weights for ischemic stroke with thrombolysis and using the hospitals selected in Bischof et al.[6]	As in "Typical RCT population"	As in "Typical RCT population"

BMS=bare metal stent; CABG=coronary artery bypass graft surgery; DES=drug eluting stent; DRGs=Diagnosis Related Groups; IPD=individual patient data; MI=myocardial infarction; PCI=percutaneous coronary intervention; RR=relative risk; SD=standard deviation.

Appendix Table D2. Calculation of cost of revascularization

Original procedure	Revascularization procedure	Costs in Yock[ab] Yock[ab] $ [range]	Proportion of such revascularizations in COURAGE	Weighted average cost for revascularization based on initial procedure[c] $ [range]
PCI	PCI	19941 (14956-24926)	73/101 = 0.723	25177 (14956-46614) SE = 7915
PCI	CABG	38845 (25871-46614)	28/101 = 0.277	
CABG	PCI	26093 (15500-33232)	18/21 = 0.857	27916 (15500-46614) SE = 7779
CABG	CABG	38845 (31076-46614)	3/21 = 0.143	

CABG=coronary artery bypass graft surgery; PCI=percutaneous coronary intervention; SE=standard error

Appendix Table D3. Parameter estimates and distributions fit for probabilistic sensitivity analysis. Distribution parameters are presented as means with standard deviations

Parameter	Distribution form	Typical RCT population	Elderly	Diabetics
Probabilities				
Procedural stroke in PCI	Beta	0.0081 (0.0028)		
Procedural death in PCI		0.0068 (0.000552)		
MI in PCI patients during follow up		0.093995 (0.0086053)		0.052239 (0.0135666)
Stroke in PCI patients during follow up		0.019147 (0.0040411)		0.003922 (0.0039062)
Revascularization in PCI patients during follow up		0.198433 (0.0117606)		0.105634 (0.0182069)
Proportion of deaths among patients with stroke during 10 years of follow-up		0.2 (0.05)		
Proportion of deaths among patients with a myocardial infarction during 10 years of follow-up		0.25 (0.075)		
Treatment effects (log-RR)				
RR of procedural stroke (BMS/DES vs. CABG)	Normal	-0.42925 (0.309639)		
RR of procedural death (BMS/DES vs. CABG)		-0.35382 (0.294528)		
RR of MI (BMS vs. CABG)		0.225523 (0.236486)	0.225523 (0.472972)	NA
RR of stroke (BMS vs. CABG)		-0.14734 (0.317057)	-0.14734 (0.634114)	NA
RR of revascularization (BMS vs. CABG)		1.617149 (0.197777)	1.617149 (0.395554)	NA
RR of death (BMS vs. CABG)		-0.09531 (0.11509)	0.195708 (0.160856)	NA
RR of MI (DES vs. CABG)		0.221909 (0.210084)	0.221909 (0.420167)	1.534714 (0.637082)
RR of stroke (DES vs. CABG)		-1.386294 (0.512109)	-1.386294 (1.024218)	-1.96611 (1.031391)
RR of revascularization (DES vs. CABG)		0.807715 (0.159516)	0.807715 (0.319033)	1.821318 (0.483001)

[a] We used the estimates considered as "contemporary costs."
[b] For PCI, we used estimates for primary stenting.
[c] The standard error for the estimates was calculated as 0.25*range, where range was obtained from the maximum and minimum values reported for the procedure specific costs.

Parameter	Distribution form	Typical RCT population	Elderly	Diabetics
RR of death (DES vs. CABG)		0.069973 (0.168394)	0.237482 (0.323178)	0.356675 (0.112387)
Utilities				
Asymptomatic	Beta		0.86 (0.043)	
MI			0.835 (0.417)	
Stroke			0.56 (0.028)	
Recurrent stroke			0.53 (0.027)	
MI post stroke			0.535 (0.027)	
Costs				
Cost of PCI with DES	Gamma		18429 (2910)	
Cost of PCI with BMS			14609 (2601)	
Cost of CABG			37576 (5882)	
Cost of repeat revascularization in PCI			25177 (7915)	
Cost of repeat revascularization in CABG			27916 (7779)	
Cost of follow-up (including drug therapy) in PCI			5116 (425)	
Cost of follow-up (including drug therapy) in CABG			5597 (445)	
Cost of clopidogrel (for DES revascularization)			1215 (NA)	
Cost of MI			11150 (1704)	
Cost of stroke			28253 (5032)	

BMS=bare metal stent; CABG=coronary artery bypass graft surgery; DES=drug eluting stent; MI=myocardial infarction; NA=not applicable; PCI=percutaneous coronary intervention; RR=relative risk.

Model Analyses

We performed the following analyses in a stepwise fashion: a decision analysis focusing on expected benefits (where QALYs was the decision relevant quantity [outcome]), a cost-effectiveness analysis (based on incremental cost-effectiveness ratios, ICERs, for pairs of compared treatments) and a value of information analysis (quantifying the expected value of perfect information, EVPI, for all parameters and the expected value of perfect parameter information, EVPPI, for specific groups of these model parameters). From each analysis we conducted, we present the output that we consider most relevant to making decisions about future research needs.

Decision analysis: Here the interest is in maximizing QALYs regardless of costs. To summarize the influence of parameters such as the treatment effects on the decision, we generate tornado graphs from one-way sensitivity analysis. Specifically, for each parameter of interest we calculated the range of expected utilities for values of the parameter of interest ranging from its lower to its upper confidence interval. The graphs are often called tornado graphs because parameters are graphed in decreasing order of influence thus giving the visual impression of a tornado.

Cost-effectiveness analysis: We carried out our analyses from a societal perspective and assumed that the aim of the decisionmaker was to choose. We generated tornado graphs depicting the range of ICERs in one-way sensitivity analysis of the parameters of interest within their 95% CI.

Value-of-information analysis (VOI): Decisions based on current knowledge may be proven wrong in the future because of uncertainties in the underlying data used, so a VOI analysis assigns a monetary value to reducing uncertainty in a decisional context.[1] VOI analysis

attempts to answer the question: how much is it worth to have information on *all* or *some* parameters of a decision problem (what is the *opportunity cost* of uncertainty)? When the value of simultaneously eliminating all model uncertainty is of interest then the overall EVPI can be calculated. If the value of eliminating uncertainty only regarding one specific parameter, or a group of related parameters, is of interest then EVPPI can be estimated. As both types of analysis convert effectiveness to monetary units to provide the total potential value of conducting future research, they require the specification of willingness-to-pay thresholds. It is widely appreciated that such thresholds are rather arbitrary, and we performed sensitivity analysis from a lower bound of zero to $200,000 per QALY gained.[1] In addition to a willingness-to-pay threshold, typically VOI analysis requires an estimate of the effective population for whom the treatment strategies under study would be applicable. Because our aim was to rank potential research areas within each population of interest, we used the per-person estimates and avoided any comparison across populations. Details of methods for methods of calculating EVPI and EVPPI have been published elsewhere.[1,7,19]

All analyses were repeated for the subpopulation specific analyses (elderly and diabetic patients).

Subpopulation Analyses

For all subpopulation analyses our approach to estimating event rates and probabilities in PCI arms and utilizing relative metrics (hazard ratios) to express the probabilities in the PCI arms was the same as the approach we utilized for the typical RCT population. In addition, the tree and Markov process structure was the same as in the typical RCT population analysis. We briefly discuss here the different sources of data for the two additional subpopulations, the elderly and diabetic patients, particularly regarding treatment effects; further details about the data sources and parameter estimates are presented in Appendix Tables 1 and 3.

Elderly: To perform quantitative analyses in elderly (older than 75 years) patients we modified our base-case model by utilizing the baseline mortality of this older population (based on vital statistics of the USA population), modeling the treatment effect for mortality in this subgroup based on the Hlatky et al.[17] meta-analysis (we used the estimate of the HR for mortality from the oldest subgroup of patients presented in the analysis, i.e.,>65 years of age). Because we did not have an estimate for the age group of interest regarding other long term outcomes (i.e. stroke, MI or revascularization) and because the estimate from the Hlatky et al. meta-analysis pertained to slightly younger patients we "reproduced" the additional uncertainty in these estimates by doubling the standard error of the relative risks for all long term outcomes.[17]

Diabetics: For short term, procedural outcomes we used the data from the PCI arm of the COURAGE trial[9] and calculated the RR for these outcomes comparing PCI with CABG from our meta-analysis of stent trials included in the Stanford CER.[16] We modeled the long term event rates and treatment effects on myocardial infarction, stroke and repeat revascularization based on the CARDia trial.[15] Briefly, this was a randomized comparison of PCI with stent placement vs. CABG where the majority (68%) of the stents used were drug eluting. Outcomes at 1 year of followup where reported in 2010. To represent the cost of clopidogrel prophylaxis, we modeled a fixed cost increment for PCI procedures as we did for the DES arm of the typical RCT population model.[18] For the decision, cost-effectiveness and VOI analysis in diabetic patients we did not model a separate BMS placement strategy as our source of data only presented data in aggregate for BMS and DES.

Results

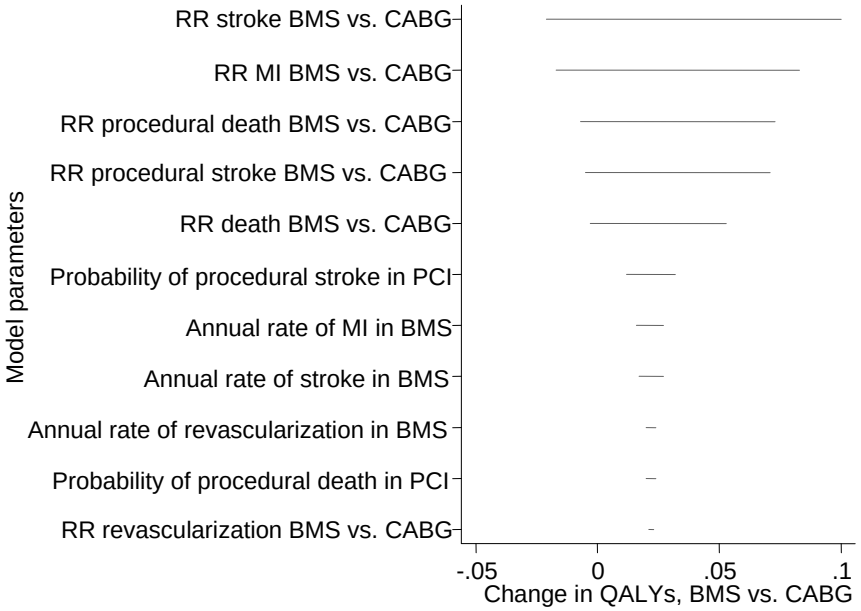
The following paragraphs show results per model (subpopulation) and type of analysis (decision, cost-effectiveness, or value of information analysis).

Typical RCT population

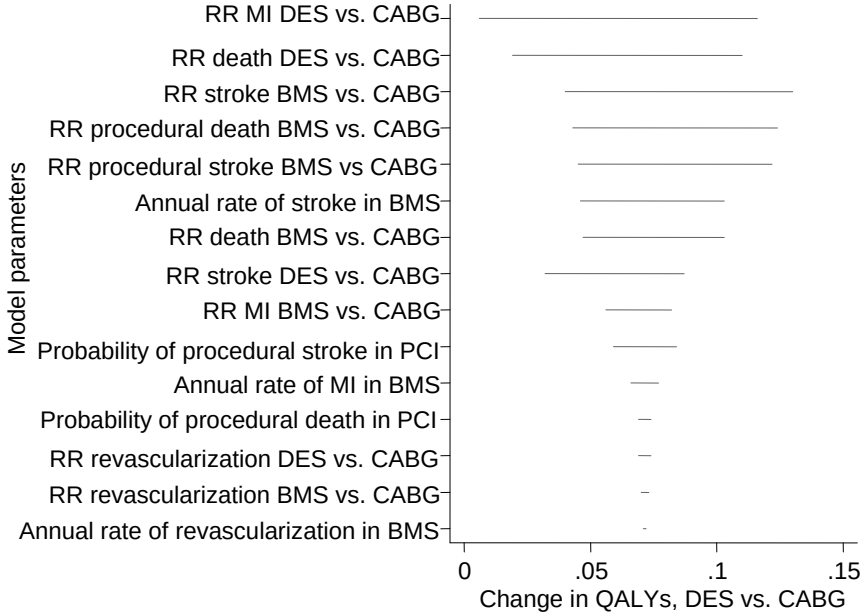
Decision Analysis

To identify parameters with great uncertainty regarding effectiveness, we performed one-way sensitivity analysis for each parameter of interest, separately comparing DES and BMS to CABG. For these analyses each parameter was evaluated for a range equal to its 95% confidence interval and the difference in effectiveness between the largest and smallest value was recorded. The larger this range the more influential the uncertainty of that specific parameter on the relative effectiveness of the two strategies being compared. We then ranked the parameters from the one with the greatest uncertainty to that with the lowest uncertainty. Appendix Figure D2 presents the tornado graph for BMS and Appendix Figure D3 the graph for DES. Parameters relevant to DES do not appear in the BMS tornado graph but BMS-related parameters appear in both graphs due to the specific parameterization of our model (please consult the appendix methods for a detailed explanation of model structure). In general, within each graph, parameters toward the top of the graph are more “influential” in sensitivity analysis along their 95% confidence interval, indicating higher priority for future research.

Appendix Figure D2. Tornado graph of decision analysis, comparing BMS with CABG in the typical RCT population



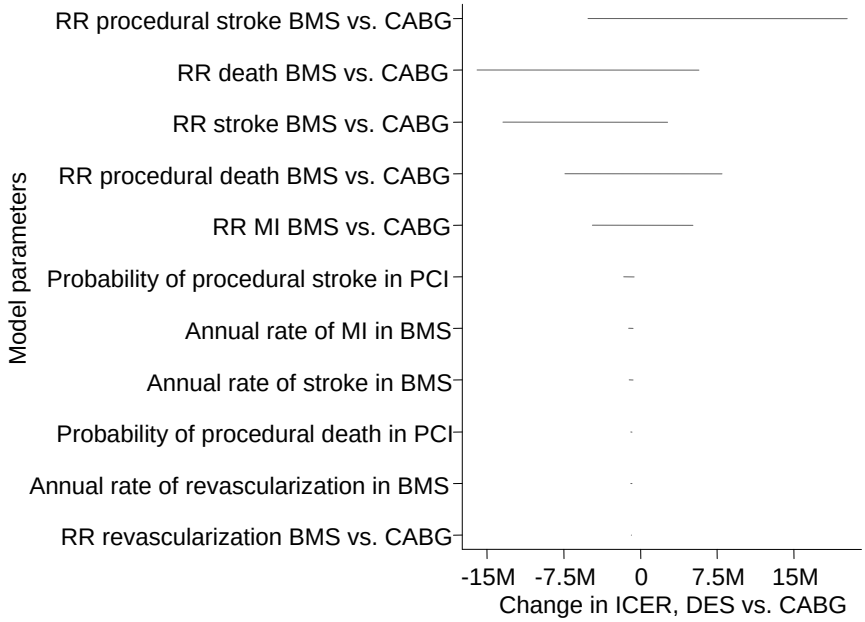
Appendix Figure D3. Tornado graph of decision analysis, comparing DES with CABG in the typical RCT population



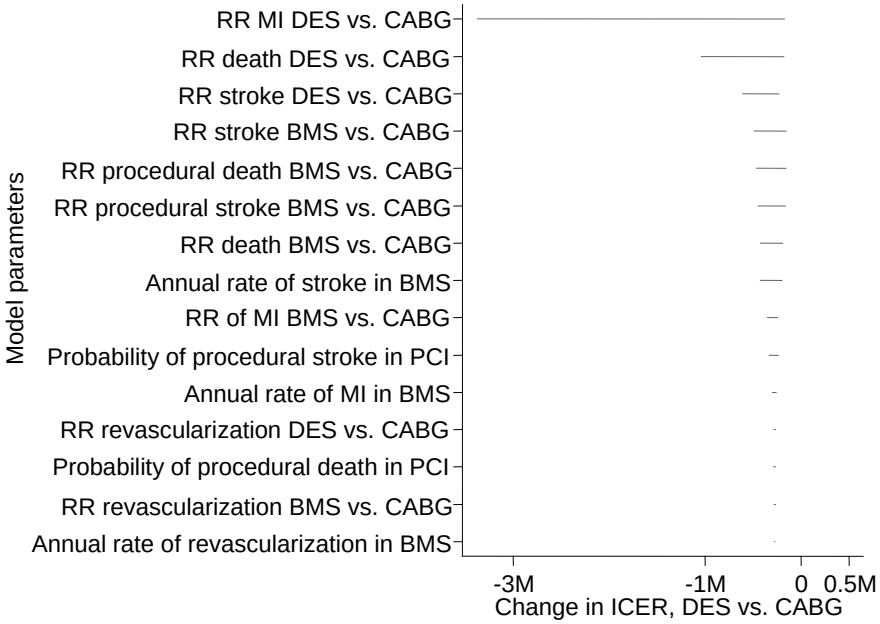
Cost-Effectiveness Analyses

We also performed one-way sensitivity analysis across the range of uncertainty (i.e. the 95% confidence interval) of each parameter under a cost-effectiveness framework. Appendix Figure D4 and D5 present tornado graphs for the same parameters evaluated in decision analysis but here the outcome is the ICER (not QALYs).

Appendix Figure D4. Tornado graph of cost-effectiveness analysis, comparing BMS with CABG in the typical RCT population



Appendix Figure D5. Tornado graph of cost-effectiveness analysis, comparing DES with CABG in the typical RCT population

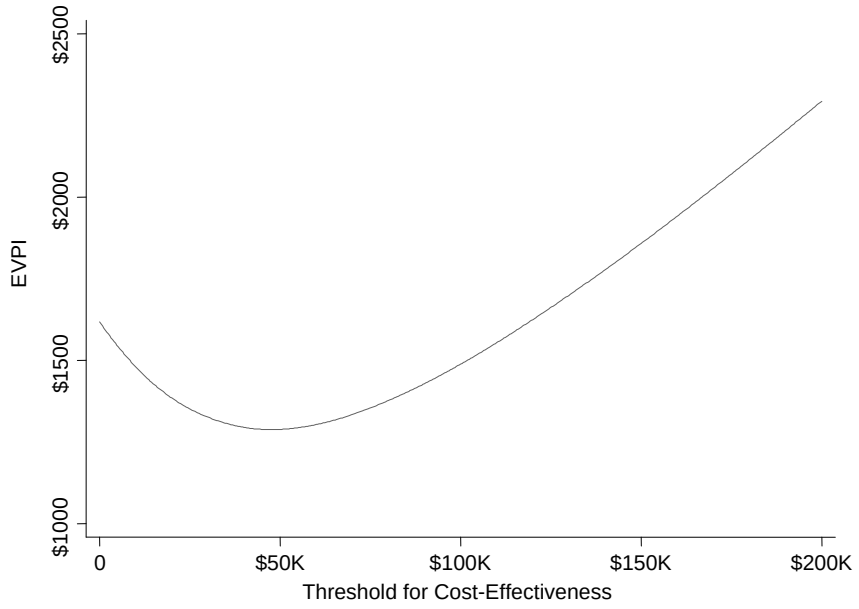


Value of Information Analysis

Expected value of perfect information (EVPI)

For the base case scenario, the overall EVPI (i.e. the value of eliminating any uncertainty around the model parameters) was more that $1250 per patient per year at a willingness-to-pay threshold of $50,000, indicating that there is substantial value in conducting further research to reduce the uncertainty regarding the comparative effectiveness of BMS, DES and CABG. This value places an upper bound on the cost society should be willing to incur to obtain information on this research topic. Appendix Figure D6 presents the estimated EVPI over a wide range of willingness to pay thresholds (0 to $200,000 / QALY).

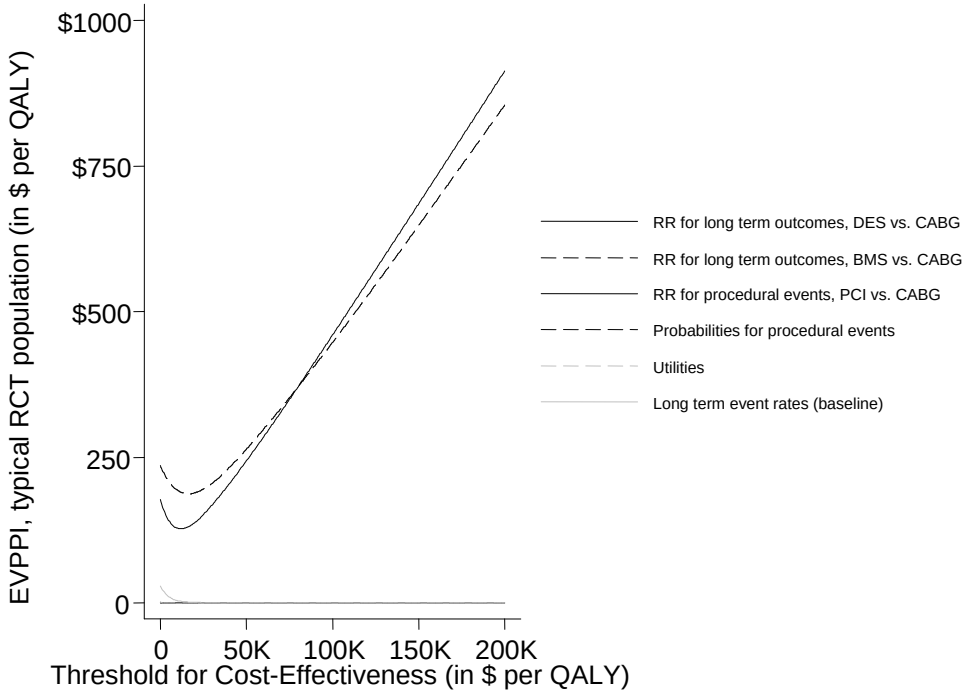
Appendix Figure D6. EVPI over willingness to pay for future research comparing DES, BMS and CABG for the typical RCT population



Expected value of partial perfect information (EVPPI)

Because different study designs can be used to obtain addition information for the different model parameters, analyses of specific sources of uncertainty (i.e. of specific model parameters) are more informative regarding the prioritization of future research. For this reason, we calculated EVPPI for the following groups of model parameters: (1) treatment effects on long-term outcomes (separately for DES vs. CABG and BMS vs. CABG) (2) procedural treatment effects, (3) long term event rates in BMS arms, (4) the probability of procedural outcomes in BMS arms (5) the probability of death after stroke and MI and (6) patient preferences (utilities).[1] Because accurate estimates of costs can be obtained with relatively straightforward research designs (mostly survey-type research or analysis of existing databases) we chose not to display the EVPPI of cost estimates in the following figures. Appendix Figure D7 presents the EVPI for comparing three strategies (DES, BMS and CABG) over a wide range of willingness to pay thresholds (0 to $200,000 / QALY).

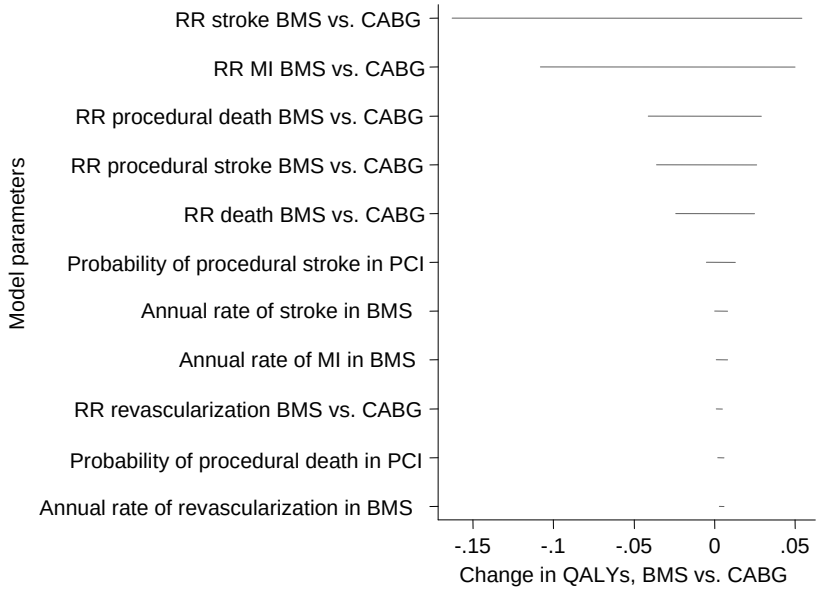
Appendix Figure D7. Graph of EVPPI for future research comparing DES, BMS and CABG for different groups of parameters over a wide range of willingness to pay for the base case scenario in the typical RCT population



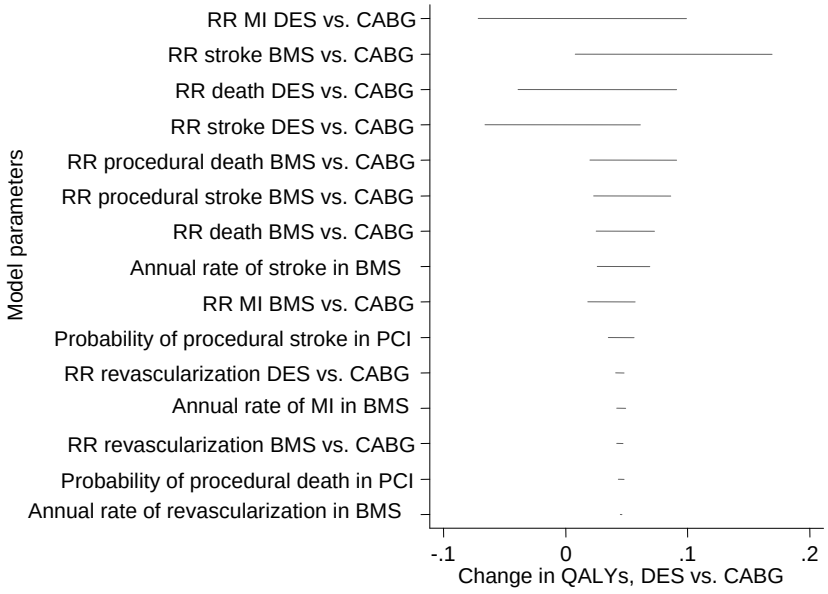
Elderly Patients

Results are listed in the same order as in the first model.

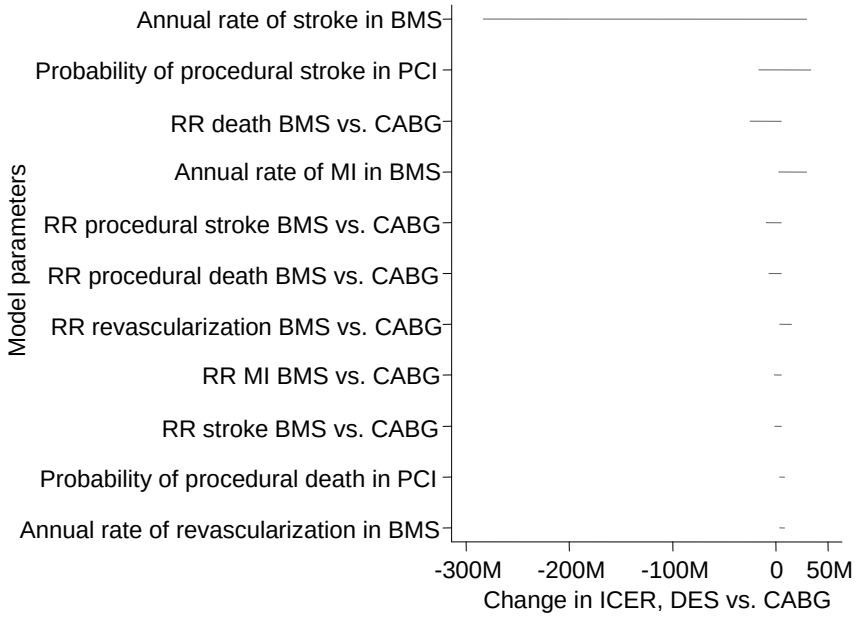
Appendix Figure D8. Tornado graph of decision analysis, comparing BMS with CABG in the model representing elderly patients



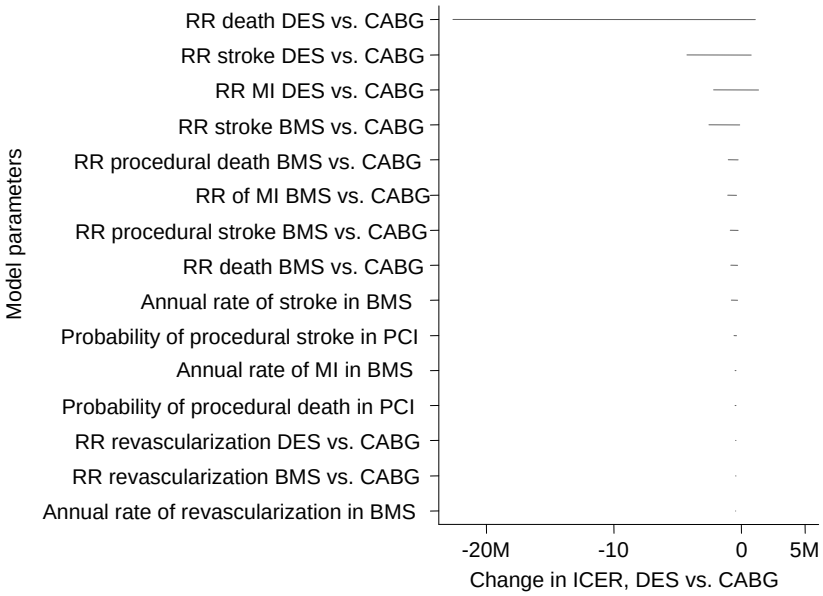
Appendix Figure D9. Tornado graph of decision analysis, comparing DES with CABG in the model representing elderly patients



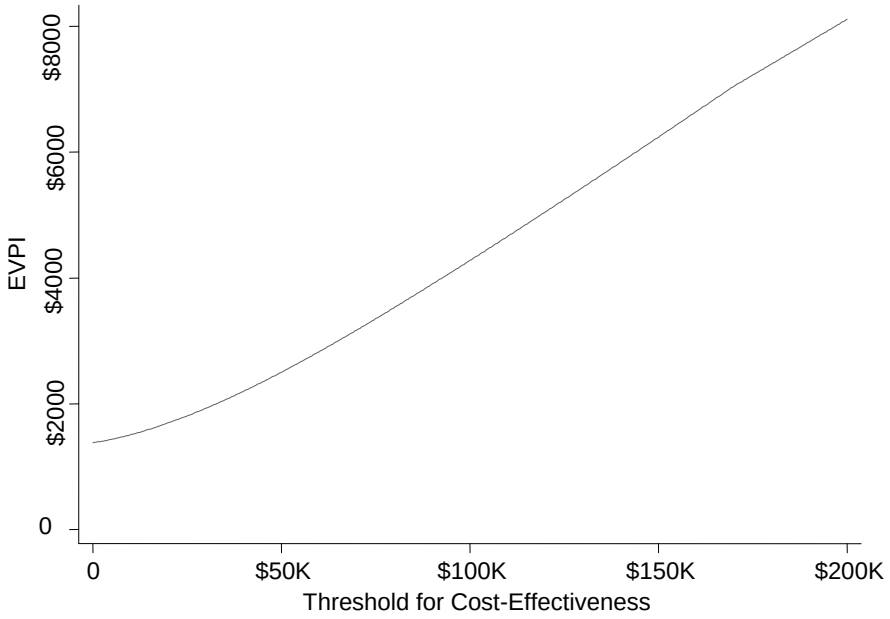
Appendix Figure D10. Tornado graph of cost-effectiveness analysis, comparing BMS with CABG in the model representing elderly patients



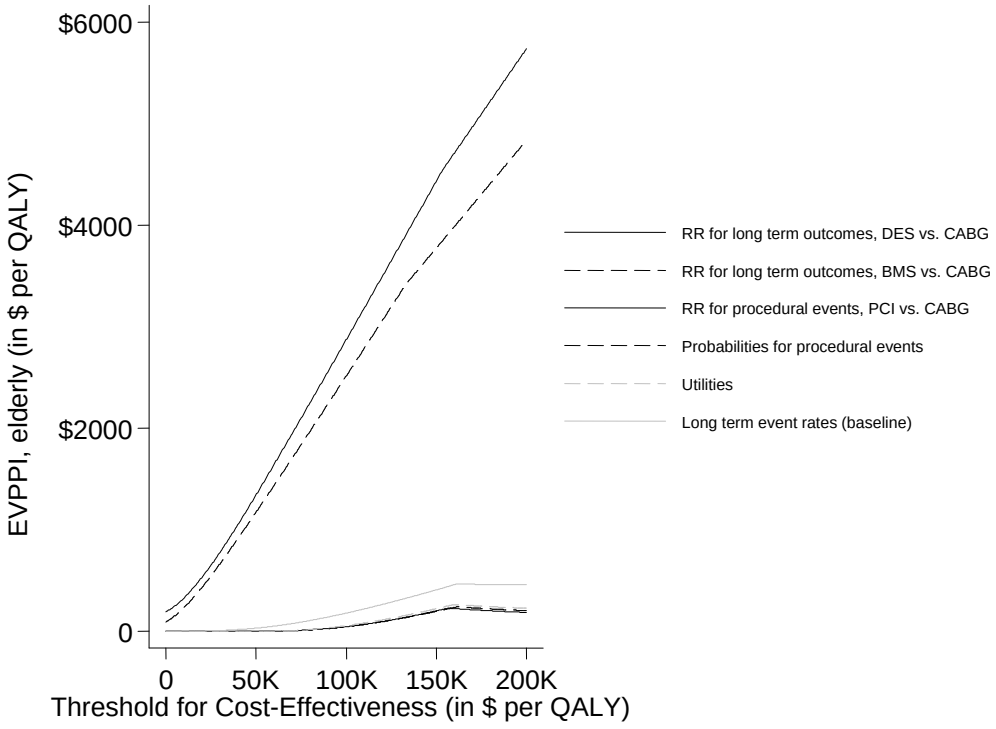
Appendix Figure D11. Tornado graph of cost-effectiveness analysis, comparing DES with CABG in the model representing elderly patients



Appendix Figure D12. EVPI over willingness to pay for future research comparing DES, BMS and CABG for the model representing elderly patients



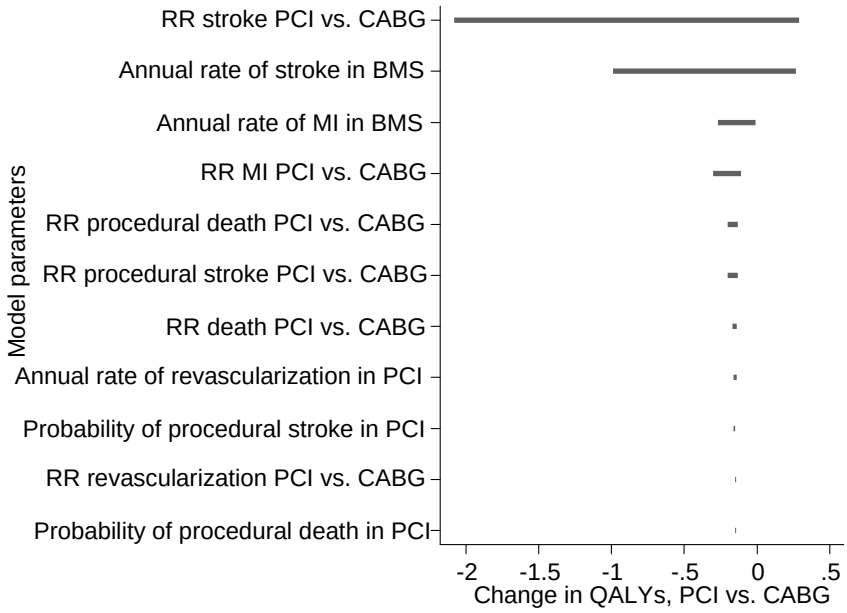
Appendix Figure D13. Graph of EVPPI for future research comparing DES, BMS and CABG for different groups of parameters over a wide range of willingness to pay for the model representing elderly patients



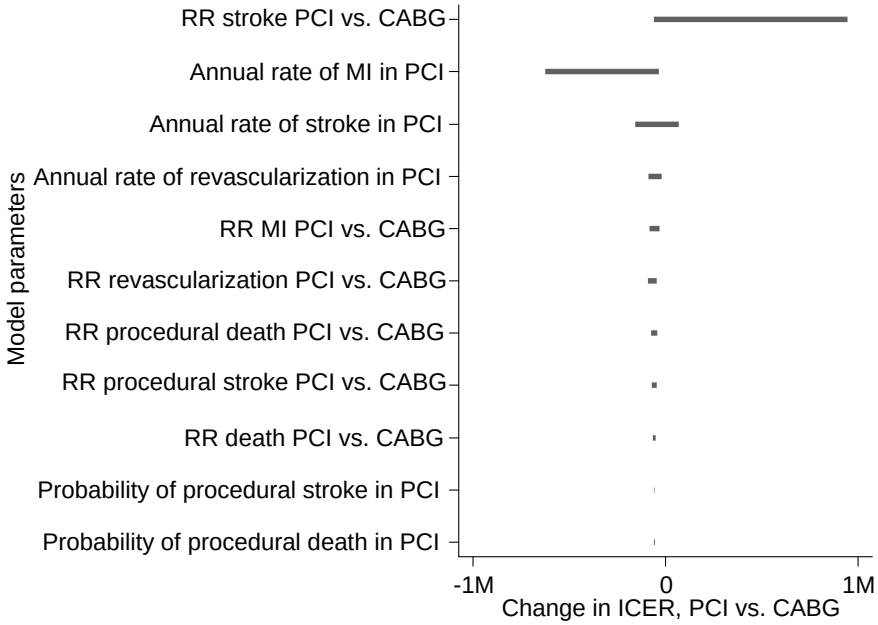
Diabetic Patients

Note that the decision tree (and consequently the decisional context) for the diabetic subpopulation analyses is slightly different from that used for the base case and elderly analyses. This difference was due to modeling considerations regarding the sources of data for each of these analyses, as explained in more detail in the supplementary methods. Results are listed in the same order as in the first model.

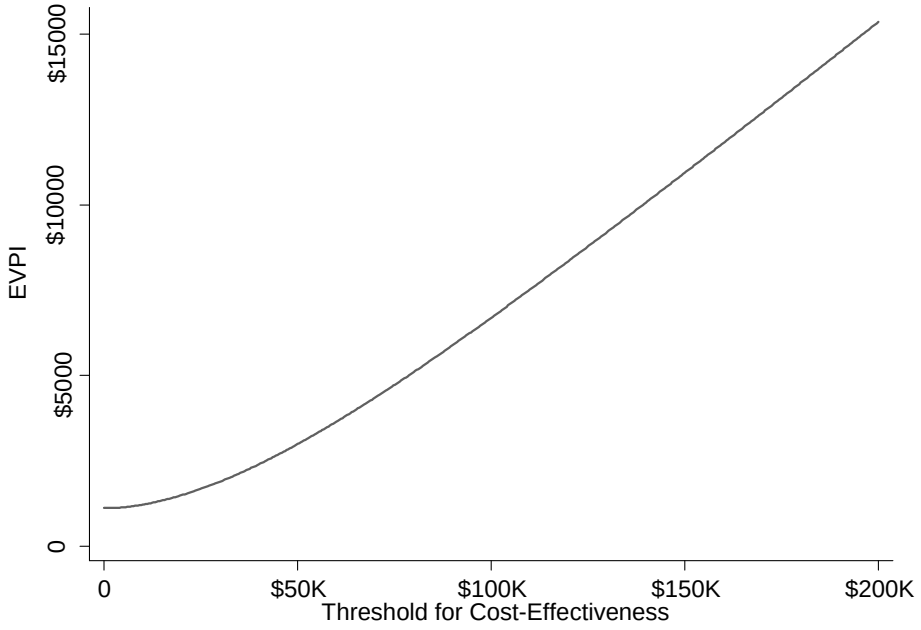
Appendix Figure D14. Tornado graph of decision analysis, comparing PCI with CABG in the model representing diabetic patients



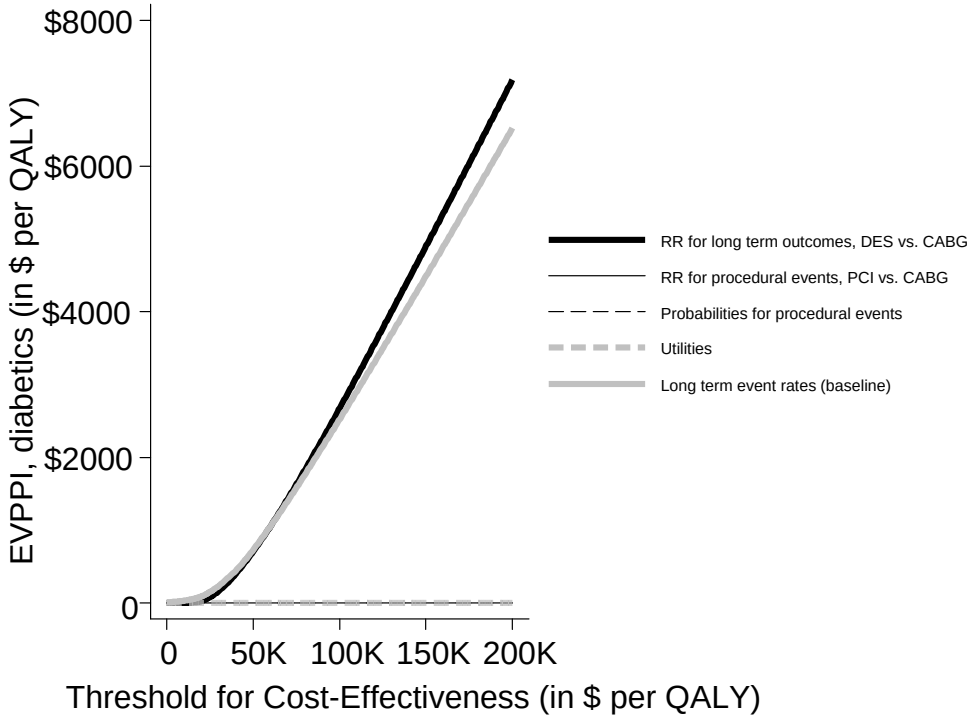
Appendix Figure D15. Tornado graph of cost-effectiveness analysis, comparing PCI with CABG in the model representing diabetic patients



Appendix Figure D16. EVPI over willingness to pay for future research comparing PCI and CABG for the model representing diabetic patients



Appendix Figure D17. Graph of EVPPI for future research comparing PCI and CABG for different groups of parameters over a wide range of willingness to pay for the model representing diabetic patients



Appendix D. References

1. Briggs A, Claxton K, Schulpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.
2. Griffin S, Claxton K, Hawkins N, et al. Probabilistic analysis and computationally expensive models: Necessary and required? Value Health 2006 Jul;9(4):244-252.
3. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005 Apr;14(4):339-347.
4. Rao C, Aziz O, Panesar SS, et al. Cost effectiveness analysis of minimally invasive internal thoracic artery bypass versus percutaneous revascularisation for isolated lesions of the left anterior descending artery. BMJ 2007 Mar 24;334(7594):621.
5. Yock CA, Boothroyd DB, Owens DK, et al. Cost-effectiveness of bypass surgery versus stenting in patients with multivessel coronary artery disease. Am J Med 2003 Oct 1;115(5):382-389.
6. Bischof M, Briel M, Bucher HC, et al. Cost-Effectiveness of Drug-Eluting Stents in a US Medicare Setting: A Cost-Utility Analysis with 3-Year Clinical Follow-Up Data. Value Health 2009 Mar 11.
7. Brennan A, Kharroubi S, O'hagan A, et al. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. Med Decis Making 2007 Jul;27(4):448-470.
8. Oakley JE, Brennan A, Tappenden P, et al. Simulation sample sizes for Monte Carlo partial EVPI calculations. J Health Econ 2010 May;29(3):468-477.
9. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 2007 April 12;356(15):1503-1516.
10. Serruys PW, Morice MC, Kappetein AP, et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. N Engl J Med 2009 Mar 5;360(10):961-972.
11. Trikalinos TA, Alsheikh-Ali AA, Tatsioni A, et al. Percutaneous coronary interventions for non-acute coronary artery disease: a quantitative 20-year synopsis and a network meta-analysis. Lancet 2009 Mar 14;373(9667):911-918.
12. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002 Aug 30;21(16):2313-2324.
13. Hill R, Bagust A, Bakhai A, et al. Coronary artery stents: a rapid systematic review and economic evaluation. Health Technol Assess 2004 Sept 8(35):iii-242.
14. Serruys PW, Unger F, van Hout BA, et al. The ARTS study (Arterial Revascularization Therapies Study). Semin Interv Cardiol 1999 December;4(4):209-219.
15. Kapur A, Hall RJ, Malik IS, et al. Randomized comparison of percutaneous coronary intervention with coronary artery bypass grafting in diabetic patients. 1-year results of the CARDia (Coronary Artery Revascularization in Diabetes) trial. J Am Coll Cardiol 2010 Feb 2;55(5):432-440.
16. Bravata D, McDonald K, Gienger A, et al. Comparative effectiveness of percutaneous coronary interventions and coronary artery bypass grafting for coronary artery disease. Comparative Effectiveness Review No 9 (prepared by Stanford UCSF-EPC) 2007.
17. Hlatky MA, Boothroyd DB, Bravata DM, et al. Coronary artery bypass surgery compared with percutaneous coronary interventions for multivessel disease: a collaborative analysis of individual patient data from ten randomised trials. Lancet 2009 Apr 4;373(9670):1190-1197.

18. Filion KB, Roy AM, Baboushkin T, et al. Cost-effectiveness of drug-eluting stents including the economic impact of late stent thrombosis. Am J Cardiol 2009 Feb 1;103(3):338-344.

19. Tappenden P, Chilcott JB, Eggington S, et al. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. Health Technol Assess 2004 Jun;8(27):iii-78.