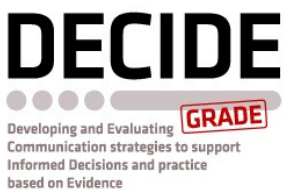

Project no. 258583
Project acronym: DECIDE
Project title: Developing and Evaluating Communication strategies to support Informed Decisions and practice based on Evidence

SP1 - Cooperation

Collaborative Project
Small or medium scale research project

FP7 – HEALTH – 2010 – two-stage

Deliverable 4.1 – Strategies for effectively and efficiently communicating evidence-based diagnostics recommendations

Due date of deliverable: 31 December 2013
Actual submission date: 21 December 2013

Start date of project: 01 January 2011 Duration: 60 Months

Organisation name of lead beneficiary for this deliverable: Academisch Medisch Centrum bij de Universiteit van Amsterdam

Coordinator: Professor Shaun Treweek

Signed:  Date: 21/12/2013

Deliverable 4.1 – Strategies for effectively and efficiently communicating evidence-based diagnostics recommendations

Nature: Report

Dissemination Level: Public (PU)

Owner
Name: AMC
Lead Beneficiary: AMC
E-mail: p.m.bossuyt@amc.uva.nl

Context
Author(s): AMC
Work Package: WP4

Document Status
Version: 1.0
Last modified: Saturday 21 December 2013
Status: Released
Approved by: Shaun Treweek
Date Approved: 21 December 2013

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of contents

Executive summary

The DECIDE project, which started on the 1st of January 2011, aims to build on the work of the GRADE working group (www.gradeworkinggroup.org) by developing and evaluating ways of effectively communicating and supporting the uptake of evidence-based recommendations about prevention, treatment and rehabilitation for different target groups. The project also develops strategies for recommendations about diagnostic tests and health system policies. In this report we present the progress of the development and evaluation of strategies for communicating evidence-based recommendations about diagnostic tests (Work Package 4).

The generally favoured strategy for developing recommendations is to use a comparative approach (i.e., comparing alternative strategies) and to focus on outcomes that are relevant for patients. Diagnostic tests pose a challenge, because of the limited availability of direct evidence linking tests to patient outcomes, and the generally lower quality and higher risk of bias in most evaluations of diagnostic tests. Because of the limited availability of direct evidence guideline developers often will have to base their recommendations on evidence from test accuracy studies, making inferences about the downstream consequences on patient outcomes based on assumptions.

The GRADE working group is developing a comprehensive framework for rating the confidence in the effect estimates of a body of evidence obtained from diagnostic tests studies and linking this evidence to patient outcomes. The work of WP4 is based on these developments.

To achieve the objectives of DECIDE, each of the project work packages is structured in three phases: Strategy development and user testing (Phase 1), evaluation of the strategies in randomised clinical trials (Phase 2), and testing the strategies with real guidelines (Phase 3). In WP4 we use methods similar to the ones developed in the work packages addressing the different target groups (WP 1-3) and health system policies (WP5), which is and will continue to run in parallel. For example, we have developed and user tested an Evidence to Decision (EtD) framework for

recommendations about diagnostic tests, consistent with the presentation format used across all work packages and focussing on patient outcomes. Strategies that are being developed in the other work packages, such as the iSoF tables, the interactive glossary and the EtD for coverage decisions will be applied to diagnostic tests, and tested as part of the different phases.

Because the development of evidence-based recommendations of diagnostic tests is less well understood, compared to recommendations for therapeutic interventions, we are also fine-tuning and elaborating the methods for arriving at diagnostic recommendations. We believe that this step is essential for achieving effective dissemination strategies.

A key conclusion from the brainstorming sessions in the initial phases of the DECIDE project was that the WP4 group did not consider it appropriate to move to user-testing without a thorough understanding of what is currently being used by guideline producers to present and grade evidence about diagnostic tests and to support their recommendations. This is similar to WP3, which considered it appropriate to do some qualitative work with patients to gain an understanding of what the public already knew about, and wanted from, guidelines before moving to user-testing. This observation has led to two substantial literature reviews and to two series of in depth interviews with guideline developers. In these we have identified current approaches, perceptions and preferences, as well as problematic areas in the development and dissemination of evidence-based recommendations about diagnostic tests.

In the next phase of WP4, we used the results from the brainstorming sessions, from the literature review and reviews, and from user testing of the diagnostic EtD framework, to work on topics that warrant further investigation. These include development of a clinical pathway as part of the guideline development process, projecting outcomes in the EtD framework (linking different types of evidence) and identifying necessary and sufficient conditions to make strong recommendations based on technical or clinical performance of medical tests (instead of comparison of patient outcomes).

1. Introduction

The DECIDE project, which started on the 1st of January 2011, aims to build on the work of the GRADE working group (www.gradeworkinggroup.org) by developing and evaluating ways of effectively communicating and supporting the uptake of evidence-based recommendations about prevention, treatment and rehabilitation for different target groups. The project also develops strategies for recommendations about diagnostic tests and health system policies.

The DECIDE project is structured into five main investigational work packages, each aimed at a different target (stakeholder) group: Healthcare professionals (WP1), policymakers and managers (WP2), public, patients and carers (WP3), users of evidence on diagnostic tests (WP4), and users of evidence on health system policies (WP5). To achieve the objectives, each of these work packages is structured in three phases: Strategy development (Phase 1), Evaluating of the strategies in randomised clinical trials (Phase 2), and testing the strategies with real guidelines (Phase 3). In WP4 we use methods similar to the ones used in the work packages addressing the different target groups (WP 1-3) and health system policies (WP5), which will run in parallel.

DECIDE's assessment of the effectiveness of communication strategies will provide an empirical, theoretically-informed basis for better understanding of the factors that influence the effectiveness of communication strategies on the various actors in healthcare.

In this deliverable we present the progress of the development and evaluation of strategies for communicating evidence-based recommendations about diagnostic tests (WP4).

2. Methods

Because the development of evidence-based recommendations of diagnostic tests is less well understood compared to recommendations for therapeutic interventions, we will – in addition to the general methods – fine-tune and elaborate the methods for arriving at diagnostic recommendations to achieve effective dissemination strategies. In collaboration with other work packages we will follow the general methods by applying the strategies that are being developed for the different target groups, such as the iSoF tables, the interactive glossary and the EtD for coverage decisions, to diagnostic tests and test strategies.

While the DECIDE approach originally consists of strategy development and user testing, evaluating the strategies and testing the strategies with real guidelines (DoW), we realized very early in the process that there was uncertainty in development of recommendations about diagnostic tests. Before venturing to the communication side, we first focused on the groundwork: identifying strategies and challenges.

Phase 1 Identifying strategies and challenges; groundwork for strategy development
We used brainstorm sessions, systematic reviews and interviews with guideline developers to identify key issues and challenges in communicating evidence-based recommendations about diagnostic tests (AMC & UHF). Furthermore, to identify challenges in assessing the quality of evidence of diagnostic test accuracy (DTA) evidence we applied the GRADE approach for diagnostic test and test strategies to Cochrane DTA reviews (AMC & UHF). DECIDE partner UHF developed and user tested versions of a diagnostic Evidence to Decision (EtD) framework.

Phase 2 Developing and evaluating strategies for communicating recommendations
Methods are currently under development.

In this report we will describe the results of phase 1 (2011 - 2013) in detail, and present an outline of phase 2 (2013 - 2015).

3. Identifying strategies and challenges in developing recommendations

3.1 Brainstorming

The key elements in the first phase, as specified in the grant application, were refined in brainstorming sessions at the Geneva Kick-Off meeting and at the Barcelona DECIDE meeting. All DECIDE partners participated in the brainstorming sessions.

3.2 Systematic reviews

Two systematic reviews were performed. In the first systematic review we compared grading systems for medical tests on how they use evidence in guideline development (AMC). Twelve grading systems could be included in the review. All varied in the degree to which methodological and process characteristics were addressed. Five systems for grading evidence about medical tests in guideline development addressed to differing degrees of explicitness the need for and appraisal of different bodies of evidence, the linking of such evidence, and its translation into recommendations. At present, no one system addressed the full complexity of gathering, assessing and linking different bodies of evidence.[1]

The second systematic review was aimed at methods used by organisations developing recommendations about diagnostic tests. We found 44 tools and their modifications to assess the quality of evidence of diagnostic test and strategies. Tools used inconsistent terminology and the criteria for moving from evidence to recommendations were incomplete for most guideline development frameworks that were evaluated (UHF).[2]

3.3 Interviews with guideline developers

To identify hurdles and issues in developing and disseminating evidence-based recommendations about diagnostic tests we performed two qualitative studies involving interviews with guideline developers. For the first study, we conducted 17 in depth interviews with international guideline developers (AMC). DECIDE partner SANTPAU assisted with some of the interviews. A convenience sample of guideline developers was selected with a broad range of experience in guideline development specific for testing from a few years to many years, and the size of organisations they

worked for also ranged from very established international guideline organisations such as NICE to local institutional guideline development groups. All interviews were conducted in English in a semi-structured manner and will be analysed using thematic analysis. A preliminary analysis of the first 10 interviews is currently being conducted.
In the second study we interviewed 23 experts who had a variety of backgrounds, formal training and experience in methods of evaluating evidence and making recommendations about diagnostic tests (UHF). They also represented a variety of national societies and organizations. Diagnostic test accuracy was the factor most commonly considered by organisations when formulating recommendations. However, the majority of experts pointed out that accuracy alone is not sufficient and that recommendations based on accuracy alone may be misleading.[2]

3.4 Applying GRADE for diagnostic tests and test strategies to DTA evidence

The GRADE Working Group developed an approach to assess the quality of evidence of diagnostic tests. The use of this approach in Cochrane reviews of diagnostic test accuracy (DTA) is new. We applied this approach to three Cochrane (DTA) reviews with the aim of better understanding the application of the GRADE criteria to such reviews. For the interpretation of the GRADE criteria, it made a difference whether assessors looked at the evidence from a patient important outcome perspective, or from a test accuracy standpoint. The GRADE criteria inconsistency, imprecision and publication bias were challenging to apply as was the assessment of comparative test accuracy reviews (AMC, UHF and other partners).[Currently under review by the Journal of Clinical Epidemiology].

3.5 Evidence to Decision framework

Using information from prior work on diagnostic test accuracy with guideline developers and authors of systematic reviews DECIDE partner UHF has developed versions of an EtD framework for recommendations about diagnostic tests accuracy information (Appendix 1).

The framework is being developed using an iterative process based on the GRADE approach to clinical practice guidelines, including guidelines about diagnostic tests and strategies, a review of relevant literature, workshops, brainstorming, feedback

from stakeholder, application of the framework's to examples, survey of scientists in the field of diagnostic test and strategies, user testing and planned trials.

The framework is consistent with the presentation format used across all work packages in DECIDE.

4. Developing and evaluating strategies for communicating recommendations

4.1 Pathway - Clinical pathway

Tests in themselves usually do not affect outcome; outcome is affected by downstream clinical management, which is guided by the results of medical testing. Developing recommendations about testing (versus no testing) or about the selection of tests for a specific purpose then requires an identification of the link between test results and downstream clinical actions: the clinical pathway. The clinical pathway is also crucial in explaining, communicating and disseminating guidelines about testing.

Our systematic review and the interviews show that guideline developers struggle with the methods for defining this clinical pathway. In the second phase of our project we want identify, test and compare different methods for defining the clinical pathway. We will do so by starting with the applications previously identified in the project: cervical cancer screening, thrombosis/pulmonary embolism, optical coherence tomography for diabetic retinopathy and tests for tuberculosis screening.

We will identify existing approaches and key elements for defining the clinical pathway (based on the systematic review, the interviews, and our own experience), develop clinical pathway building blocks as an aid for guideline developers and comparatively apply them to the testing topics.

4.2 Integration - Projecting outcomes in Evidence-to-recommendations framework

In the evidence to recommendation framework, one needs one or more tables with summary of findings on patient-important outcomes. With most medical testing, these outcomes have never been estimated in a single study. The estimates are calculated on the basis of the pretest probability, diagnostic test accuracy, presumed natural

history of disease, anticipated frequency of outcomes related to the disease, treatment efficacy and reported as patient or population outcomes. These are multiple pieces of evidence.

Several issues have to be dealt with in these tables. How should these estimates be calculated? How should the confidence (quality) be expressed when the results of multiple studies are combined? Can we use the same format for communicating these summaries of findings, similar to the formats we use when the evidence is based on a single systematic review, or a single RCT? In staging this part of the project, we will coordinate efforts with an ongoing Cochrane Methods Innovations Fund project about grading the quality of evidence and preparing Summary of Findings Tables for diagnostic tests.

Based on user testing in several settings and workshops (UHF & AMC) we have identified some of the key elements that are required - or likely not required - in the evidence to recommendation framework. Most importantly, while there may be different presentation formats of the framework, for different types of decisions (clinical practice guidelines versus coverage decisions).

4.3 Segmentation - Technical and clinical performance as necessary and sufficient conditions

Based on an analysis of the clinical pathway, it is very well possible that evidence about the technical or clinical performance of medical tests can be used to make strong recommendations for or against testing. This would greatly simplify the development and dissemination of evidence-based recommendations about medical tests.

An example of the former may be a laboratory assay, proposed as a monitoring test. If that test has limited reliability, then it should not be used as a monitoring test. In this case, the technical performance is the basis for a strong recommendation against testing.

An example of the second may be a replacement problem, where one test is replaced by another addressing a similar population, with greater accuracy, relative to a meaningful clinical reference standard, and greater acceptability, similar or lower resource utilization and practical use issues. In that case a strong recommendation for the new test can be made, based on these two elements of the clinical performance of the tests.

In both cases, the recommendations are not based on an explicit comparison of patient outcomes, but on – quite diverse – elements of technical and clinical performance of tests, but after a careful analysis of the clinical pathway and the linked evidence (coming from the reference test).

We will identify and apply a number of strategies for using the GRADE approach to this "segmental" approach for developing recommendations about medical tests.

5. Conclusion

The DECIDE approach is based on strategy development and user testing, on evaluating these strategies, and on testing the strategies with real guidelines. In collaboration with other work packages we will follow this approach by applying the strategies that are being developed for the different target groups, such as the iSoF tables and the EtD for coverage decisions, to diagnostic tests and test strategies.

We realized very early in the process that there was major uncertainty in development of recommendations about diagnostic tests, and that this uncertainty will impact the communication of recommendations. We therefore first aimed to identify strategies and challenges in the development of evidence-based recommendations about diagnostic tests, which have led to substantial publications. In the next phase we will focus on development and evaluation strategies for communicating this specific type of recommendations, based on the results of the first phase. We will address different aspects in linking test results to patient important outcomes, which is the main challenge in developing and communicating recommendations about diagnostic tests. Working in close collaboration with the other DECIDE partners allows us to

investigate when and which communication strategies specifically targeted at diagnostic recommendations are needed.

References

1 Gopalakrishna G, Langendam MW, Scholten RJPM, Bossuyt PMM, Leeflang MMG. Guidelines for guideline developers: a systematic review of grading systems for medical tests. Implementation Science 2013,8:78

2 McMaster University. A systematic evaluation of diagnostic methods, evidence and judgments: trusting the link between diagnostic tests and population outcomes. Final report to the German National Association of Statutory Health Insurance Funds on January 31, 2013

Appendix 1 Diagnostic EtD framework

Should [index test] instead of [comparison] followed by treatment be used to screen, diagnose and manage [health problem]?

Patients: ***Diagnostic intervention:*** ***Comparison:*** ***Implied purpose:*** ***Linked treatment(s):*** ***Anticipated outcomes (prevented and caused by testing and subsequent management if applicable):***	***Background:*** ***Setting:*** ***Perspective:*** *Health system versus individual patient*

	DOMAIN	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS/EXPLANATIONS
PROBLEM	**How common is the problem?** **Is the problem severe?** May skip for individual patient perspective	*No* *Probably No* *Uncertain* *Probably Yes* *Yes* ***Varies*** ☐ ☐ ☐ ☐ ☐ ☐ *No* *Probably No* *Uncertain* *Probably Yes* *Yes* ***Varies***		*Provide information about the importance of the problem for the setting* *Describe if frequency and severity have an impact on considering the overall recommendation.*
DIAGNOSTIC TEST ACCURACY	**What is the diagnostic test accuracy?**	*Very inaccurate* *Inaccurate* *Uncertain* *Accurate* *Very accurate* ***Varies*** ☐ ☐ ☐ ☐ ☐ ☐	**Diagnostic test accuracy** ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	*Describe the diagnostic test accuracy (i.e. sensitivity and specificity) and if it is sufficient to continue developing a recommendation.* **Also see full diagnostic test accuracy evidence profile**
	What is the overall confidence in the diagnostic test accuracy information?	*Very low* *Low* *Moderate* *High* ☐ ☐ ☐ ☐		

BENEFITS AND HARMS

How important are these outcomes?		***The relative importance or values of the main outcomes of interest (pick 5):***
Overall, compared to the alternative, are the anticipated benefits large?		
Overall, compared to the alternative, are the anticipated harms small?		
What is the balance of the benefits and harms/burden?	☐ Benefits outweigh harms/burden ☐ Benefits slightly outweigh harms/burden ☐ Benefits and harms/burden are balanced ☐ Harms/ burden slightly outweigh benefits ☐ Harms/ burden outweigh benefits	**Also see full third layer Summary of Findings** **Describe narratively in the "Details of Judgment column" or rate the benefits and harms by considering:** *Describe narratively in the "Details of Judgment column" or rate the benefits and harms by considering:* *The focus is on patient important outcomes that are calculated on the basis of the pretest probability, diagnostic test accuracy, presumed natural history of disease, anticipated frequency of outcomes related to the disease and to the direct effects of the test, treatment efficacy and complications and reported as patient or population outcomes.* *On average, how important are the outcomes to patients?*

Outcome	**Relative importance**
[Outcome]	-
[Outcome]	-
[Outcome]	-
[Outcome]	-
[Outcome]	-

Critical Outcomes:	*Large benefit*	*Small benefit*	*No effect*	*Small harm/ burden*	*Modest harm/ burden*
1. …	☐	☐	☐	☐	☐
2. …	☐	☐	☐	☐	☐
3. …	☐	☐	☐	☐	☐
4. …	☐	☐	☐	☐	☐
5. …	☐	☐	☐	☐	☐

	Overall, is there certainty about the link between the diagnostic test accuracy information and the linked benefits and harms?	*Very uncertain* ☐ *uncertain* ☐ *Moderately certain* ☐ *Certain* ☐ *Very certain* ☐		**Also see full third layer Summary of Findings**
	What is the overall confidence in the estimates of effect for benefits and harms?	*Very low* ☐ *Low* ☐ *Moderate* ☐ *High* ☐		*This certainty is high if there is moderate or high quality evidence indicating that treatment has clear consequences for patient important outcomes.* *What are the underlying values and preferences for the outcomes associated with the test and the problem is our confidence in these values and preferences?*
VALUES AND PREFERENCES	**Is there similarity about how much people value the main outcomes?**	*Similar* ☐ *Probably similar* ☐ *Uncertain* ☐ *Probably not similar* ☐ *Not similar* ☐		*Source of variability if any:*
RESOURCES	Are the resources required small? **(may skip for individual patient perspective)**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ *Varies* ☐		*What are the costs per resource unit?* *Opportunity cost:* *Is this intervention and its effects worth withdrawing or not allocating resources from other interventions* *Differences across settings:* *Is there lots of variability in resource requirements across settings?*

	Is the incremental cost (or resource use) small relative to the benefits?	*No* *Probably No* *Uncertain* *Probably Yes* *Yes* ***Varies*** ☐ ☐ ☐ ☐ ☐ ☐		*Are the cost (including out of pocket) worth the benefits*
EQUITY	**What happens to health inequities?**	*Increased* *Probably increased* *Uncertain* *Probably reduced* *Reduced* ***Varies*** ☐ ☐ ☐ ☐ ☐ ☐		*Would the implementation of the intervention reduce inequities?*
ACCEPTABILITY	**Is the option acceptable to key stakeholders?**	*No* *Probably No* *Uncertain* *Probably Yes* *Yes* ***Varies*** ☐ ☐ ☐ ☐ ☐ ☐		
FEASIBILITY	**Is the option feasible to implement?***	*No* *Probably No* *Uncertain* *Probably Yes* *Yes* ***Varies*** ☐ ☐ ☐ ☐ ☐ ☐		*Is this intervention generally available? Can it be implemented?*

Recommendation						
Should [index test] instead of [comparison] followed by treatment be used to screen, diagnose and manage [health problem]?						
Overall balance of consequences	*Undesirable consequences* clearly outweigh *desirable consequences*	*Undesirable consequences* probably outweigh *desirable consequences*	The balance between desirable and undesirable consequences is too uncertain*	The balance of desirable and undesirable consequences indicates they are very similar*	*Desirable consequences* probably outweigh *undesirable consequences*	*Desirable consequences* clearly outweigh *undesirable consequences*
	☐	☐	☐	☐	☐	☐
	We recommend against the option or for the alternative	*We suggest not to use the option or to use the alternative*	*No recommendation*		*We suggest using the option*	*We recommend the option*
	☐	☐			☐	☐
Panel decisions	*Describe decision making process if relevant*					
Recommendation (text)	*Formulate clear recommendation*					
Remarks and justification	*Explain the rationale and provide important disclaimers and remarks*					
Implementation considerations	*Describe issues relevant for implementation*					
Research priorities	*Describe research priorities*					

* In this situation no recommendation could be reasonable

Definitions for ratings of the certainty of the evidence (GRADE)**

Ratings	Definitions	Implications
⊕⊕⊕⊕ High	This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different* is low.	This evidence provides a very good basis for making a decision about whether to implement the intervention. Impact evaluation and monitoring of the impact are unlikely to be needed if it is implemented.
⊕⊕⊕◯ Moderate	This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[4] is moderate.	This evidence provides a good basis for making a decision about whether to implement the intervention. Monitoring of the impact is likely to be needed and impact evaluation may be warranted if it is implemented.
⊕⊕◯◯ Low	This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[4] is high.	This evidence provides some basis for making a decision about whether to implement the intervention. Impact evaluation is likely to be warranted if it is implemented.
⊕◯◯◯ Very low	This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[4] is very high.	This evidence does not provide a good basis for making a decision about whether to implement the intervention. Impact evaluation is very likely to be warranted if it is implemented.

*Substantially different: large enough difference that it might have an effect on a decision

**The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care. The working group has developed a common, sensible and transparent approach to grading quality of evidence and strength of recommendations. Many international organizations have provided input into the development of the approach and have started using it.

