Appendix

Appendix 1 – Documentation behind the guideline development

Findings from the review and interview study were grouped row-wise into similar topics or components of an early assessment model and further divided into two groups: analytical components (A0-A5) in Table A1, while Table A2 contains process components (P0-P4).

Table A1 Analytical components for early assessment in health organisations

#	**Findings from interview and review (I = interview, R = review)**
A0	An early decision support should be transparent and easy to communicate (I) Early assessments are likely to be internal decision aids (R)
A1	An early decision support should include a broad set of domains. Five main domains are stressed: 1) strategic fit/misfit, 2) clinical effects or technological aspects, 3) patients' or customers' perspectives, 4) organisational aspects, and 5) cost analysis (I) The majority of existing early assessment models are comprehensive, time-consuming, and not transparent. "Known" cost-effectiveness methods are used. One model judges the strategic fit as step one in an early assessment (R) Both quantitative and qualitative methods are used to measure the domains (I)
A2	Describe uncertainty and risks regarding the IMT (I) + "Various criteria are used to make stop/go decisions, e.g. risk profile to high" (I) The majority of models handle risk in simple ways, e.g. one-way sensitivity, scenarios, "risk burndown", or a prevalent project management tool like probability of an event multiplied by impact (I+R)
A3	Difficult to choose the correct variables (KPIs). But make sure that you have only a few (I) + Set milestones before the project begins and track over time if they are achieved (I) Use energy on decomposing the size of the target or patient group (I)

#	**Findings from interview and review (I = interview, R = review)**
A4	Always include information on what input/data is based on and who has delivered the input (I) One model has their own method with an "evidence-mapping" exercise at the core - adapted GRADE (R)
A5	Tracking changes and fixed evaluation templates used on all IMTs (I)

(I) = interview, (R) = review; Key Performance Indicators (KPIs); Innovative Medical Technologies (IMTs)

Table A2 Process components for early assessment in health organisations

#	**Findings from interview and review (I = interview, R = review)**
P0	The in-house analyst compares the project to similar projects in their library of previous reviews and collects various relevant information (R) The evaluator avoids personal contact with the entrepreneur (R) Important to involve experienced people in the assessments (I) Stakeholder analysis and milestones are often applied (I) + One model uses a portfolio/project management approach (R)
P1	Establish dedicated prioritising committees for IMTs, which separate the IMT development processes and involved staff from the stop/go decision. They should: 1) meet at regular times, and 2) consist of senior management and external and internal people who must all be without vested interest in the IMTs (I)
P2	Cognitive biases are generally not handled analysis-wise (R) Nurture a culture with critical reviewing assumptions behind the early assessments; a trial/error mentality, where focus is on minimising optimism bias (I) Optimism bias in assessments is common and some counter this bias by using historical knowledge databases, by inviting the devil's advocate to test assumptions, using pessimistic assumptions, etc. (I) Critical questioning and challenging of assumptions are central, i.e. testing/falsification and using external eyes on your work (I)

#	**Findings from interview and review (I = interview, R = review)**
P3	The reported average success rate is 10% across organisations, 15% in the public vs. 8% in the private (I). One study reports a commercial success rate for early stage ventures of 11% (R) Private sector seems to make more active use of the success rate valuing IMTs (I) One study uses the following five fixed categories when reporting the assessment reached (R): A—recommended for development, B—may go forward, but need to collect more data, C—recommended to go forward, returns likely modest, D—doubtful, further development not recommended, E—strongly recommend to stop further development
P4	Most models are not really iterative, but more than half of the studies advise it (R) Iterative models dominate (I) + Updates ranging from monthly, to every six months to once in a year (I) + Use of a phase model in the developmental process is prevalent and often standard models are used (I)

(I) = interview, (R) = review; Innovative Medical Technologies (IMTs)

Table A3 shows how principles and guideline are connected, i.e. which steps in the guideline draw on which principle(s).

Table A3 Overview of how the principles translate into the steps in the guideline and the templates

Step	**Use the following principle(s)**
Step 1+3	Principle A1, A2, A3, A4, and P2
Step 2	Principle A1, A4, and P1
Step 4	Principle P2
Step 5	Principle P3
Step 6	Principle A4 indicates that the evaluation plan is mandatory, i.e. once the evaluation plan is agreed to by the committee in step 2, it must be followed
Templates	Principle A1, A2, A4, and A5 All three EARTH templates have a fixed format; keep track of past updates and progress over time, document important changes, and support updates at gates
All steps / templates	Principle A0: Transparency and ease of communication Principle P0: Although not directly described in the following guideline and templates, the statements within P0 are important considerations assumed applied in all relevant steps in the guideline for early assessment Overall, EARTH is structured like a stage-gate process (see Figure 1) supporting updates and decision gates (principle P4)

Appendix 2 – Participants in early assessment

Example of members in the four groups of individuals, A-D, involved in early assessment:

A. The neutral analyst heads the evaluation team which consist of development people / innovation consultants, clinicians pro/con the IMT, etc. The neutral analyst is an in-house analyst from an HTA unit or a similar independent and neutral entity in the hospital, who avoids any direct personal contact with the manufacturer or entrepreneur.

B. The prioritising committee (senior management, head of innovation, clinicians, external members, patients, etc.)

C. An internal review group with other relevant stakeholders at the hospital (perhaps the entrepreneur, staff from economics, innovation, IT department, relevant clinical departments, etc.)

D. The external review group (experts from other hospitals, professional business developers from the private or public sector, academia, etc.)

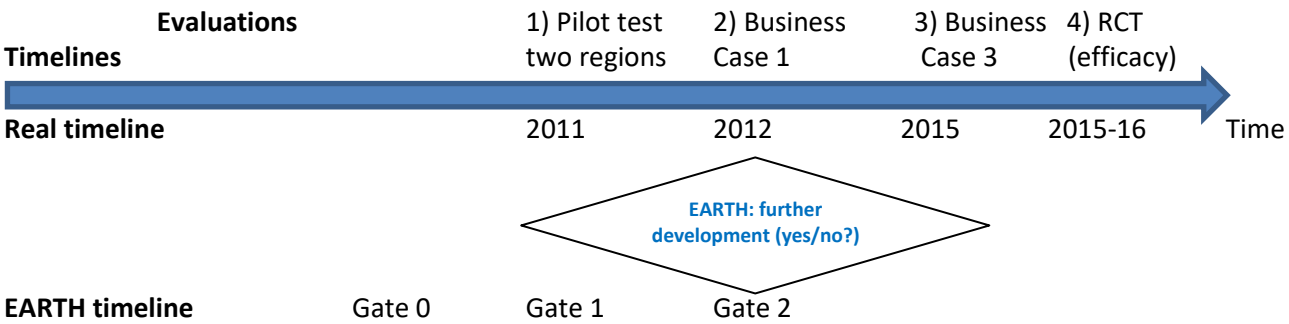
Appendix 3 – An early assessment

Background

The usefulness and applicability of EARTH, i.e. the guideline and templates, are illustrated with a somewhat imperfect example. A national project on telemonitoring of diabetic foot ulcer patients constitutes the case. This case is imperfect since no early assessment has taken place for this particular IMT. In this project, consultations at the outpatient clinic at the hospital were substituted with teleconsultations in the patient's own home carried out by a specialised nurse using telemedicine for consulting hospital experts. In the region of southern Denmark two out of three consultations were converted to telemonitoring [58]. The goals of the project were cost reduction and improved wound healing.

Based on a so called business case with heavy assumptions as substitute for evidence - business case 1 in figure A4 below - it was decided to turn the above program into national policy by promising large health gains and economic gains. The fact that a later business case - business case 3 in figure A4 - drastically reduced the expected gains and that a RCT and a cost-effectiveness analysis [59] finished in 2015 and 2016 did not show clinical or economic effects did not change the national policy. There is no doubt that the whole process would have been improved by an early assessment process. Therefore, a "counterfactual" early assessment is developed. The idea is to show what an early assessment would have looked like in hindsight and critically ask: What would the decision outcome have been if early assessment had been carried out with EARTH? Explicitly would a yes or a no have been issued in 2012 at the "diamond" in figure A4?

Figure A4 Overview of the actual national assessments of the case and the corresponding "counterfactual" EARTH timeline



Data sources: 1) An internal KPI report on preliminary pilot data from the RCT and a test in the Region of Zealand, n=16 [60]. 2) A national evaluation ("business case") performed by a private consultancy firm [61]. 3) A national evaluation ("business case") performed by a private consultancy firm [62]. 4) An RCT of the IMT performed by a public hospital, n=374 [58, 59].

Figure A4 shows the timeline for the actual assessments performed for the case in the first row below the blue arrow. A total of four evaluations were completed within the years 2011-2016. As mentioned above a "go" was given to nationwide implementation based on the two evaluations from 2011 and 2012. Would a counterfactual EARTH assessment based on the 2011 and 2012 evaluations labelled Gate 1 and Gate 2 in EARTH have led to further development of the IMT? The 2011 and 2012 evaluations mirror the typical situation for carrying out EARTH: Very little hard data and a high degree of uncertainty. The EARTH assessment ideally covers Gate 0, Gate 1 and Gate 2 and the 'diamond decision' on further development is based on data obtained at Gate 1 and Gate 2, while Gate 0 covers the initial work of the prioritising committee regarding whether to initiate test of the IMT and agreeing to an evaluation plan.

The counterfactual EARTH assessment

We now walk through the steps of the EARTH model using the guideline (see Figure 3). An early assessment is performed with two updates based on the 2011 and 2012 evaluations. Actual templates are needed for this task and templates that adhere to the EARTH principles are used in reporting the EARTH assessment below.

Step 1 and 2 (Gate 0): In a counterfactual EARTH assessment, the first step would be to establish a prioritising committee for the IMT who establishes the strategic fit and assuming they decide on initiating test of IMT then agree to an evaluation plan provided by the evaluation team. The evaluation plan should contain suggested KPIs, risks and desired evidence development on data during the course of the two planned updates. Based on the original aim of the project, four KPIs are formulated by the evaluation team at Gate 0 in an initial impact case: two clinical KPIs and two economical ones. Normally the target or goal of the KPIs would be explicitly stated in a Gate 0 column. However, specific goals were not developed early on for this project and thus Gate 0 is left empty in Table A5.

Table A5 Template for the impact case for the diabetic ulcer case

Impact case in EARTH (NA = data not available)					
KPIs (impact)	**Gate 0**	**Gate 1**		**Gate 2**	
Clinical aspects					
K1) No. ulcers treated with tele-monitoring in Denmark (N)	NA	NA		12,240	
K2) Wound healing increased	NA	NA		10 % faster healing time	
K3) Mortality unchanged	NA	3 in TM, 0 in SM (16.7% vs. 0%)		NA	
*Evidence lvl. (actual|planned)	-	2	2	1	2
Economic aspects					
K4) Annual economic saving	NA	Possible (but no numbers)		DKK 330m	
K5) Time for visits decreased	NA	Double resource use		To 59% (from 116 to 69 visits)	
*Evidence lvl. (actual|planned)	-	1-2	2	1	2

A "traffic-light"-approach is used (green-yellow-red colour coding) to indicate if Key Performance Indicators (KPIs) and evidence level are developing in a positive direction or according to plan, i.e. goal or expectation. NA means data not available. However, due to the many "NA's" in the table colour coding (comparisons) was not possible to perform regarding all cells.

*) When deciding and assessing the evidence level, an evidence hierarchy developed for innovations is used with a 1 to 5 scale, where a higher number indicates stronger evidence [48]. Actual evidence level at a given gate is compared to the plan, i.e. the evaluation strategy decided in step 2. Hence, the two numbers for each column.

We assume that the evaluation team suggests (and the prioritising committee agrees) to monitor the IMT with two updates, i.e. the pilot study and the business case, Gate 1 and Gate 2, respectively, and to aim for an evidence level of minimum 2 on KPIs at both gates. The main result of the impact case is summarised in Table A5 (at the time when two updates has elapsed). The data sources for gate 1 and 2 are the small pilot study from 2011 and the initial business case from 2012 in Figure A4. The small pilot evaluation analysed data from 16 patients on a few KPIs, e.g. time spent on consultations and mortality. The initial business case with liberal use of numerous assumptions and based on expert opinions alone contained expectations about clinical and economic effects. The strength of the evidence was low in both of the above mentioned evaluations. This is reflected in the template in Table A5 where the actual and planned evidence level is shown in the row below each group of KPIs (hence the two numbers per column for each gate).

The update cycle (Gate 1): At the first update, i.e. gate 1, we assume data from the pilot study is entered into the impact case and after analysing the data a new KPI, K3, was added because the evaluation team

discovered a possible negative trend in mortality (though very uncertain given that the evidence level behind the data is very low, i.e. only 16 patients). Next, the prioritising committee would meet for the first update to rate the potential of the IMT. Presented with the impact case as it looks at Gate 1 in Table A5, we assume that the prioritising committee was not satisfied with the missing explicit goals for the IMT, the possible negative effect on mortality, and the increased resource use. Thus, for Gate 1, the prioritising committee issued a B-rating (may go forward, but need to collect more data) in the overview template shown in Table A6.

Table A6 Decision part of the template for the EARTH-assessment overview for the diabetic ulcer case

EARTH assessment overview			
Decisi on	**Potential rating (filled out by committee after deliberation of EARTH-assessment at current gate)**		
	Rating of the potential*	Gate 1	Gate 2
		B	C (No-go for further development)

* A, B, C, D or E where: A—recommended for development, B—may go forward, but need to collect more data, C—recommended to go forward, benefits likely modest, D—doubtful, further development not recommended, E—strongly recommend to stop further development.

The update cycle (Gate 2): At Gate 2, the last update according to the plan, we assume data is entered from the 2012 initial business case into the impact case and after analysing data, the prioritising committee now has access to data corresponding to both gates in Table A5. Although "data" in Gate 2 is positive, it is essentially only a formulation of explicit goals for the project based on the lowest evidence available, i.e. expert opinions. Further, the best evidence available (Gate 1) is not supporting or making the goals very plausible. Due to the low evidence level and the negative development in K3 and K5, we find it likely that the prioritising committee would at best issue a C-rating (recommended to go forward, benefits likely modest) at Gate 2.

Step 6: Gate 2 was the last planned update and should the prioritising committee now wish to issue a "go" to usual clinical testing at this point in time it is not possible. The reason is that the planned minimum evidence level of 2, which the committee agreed to at Gate 0, is not reached on any of the KPIs, cf. the

binding evidence threshold. Instead of the evidence level improving, it has actually declined over time. Also, less than a B in potential rating is issued. In summary then, the result of the counterfactual EARTH is a "no-go" from the EARTH assessment regarding further development (or usual clinical testing).

Perspectives

This was an imperfect example. In a more full blown EARTH assessment of this case, the following four changes to analysis and process would most likely have been proposed:

1. A critical questioning procedure (steps 1 and 4 in the guideline in Figure 3) should have been applied and would most likely have resulted in:
    a. Explicit goals at gate 0 for the IMT - the waterfall method alone would have made sure that the size of the patient group was estimated, i.e. K1 in Table A5
    b. Reduced any initial expectations on most KPIs considerably during both updates
2. Starting a RCT right away would NOT have been the approach in EARTH where a clear evaluation and evidence development plan from the onset of the project is a cornerstone (step 1 in the guideline). However, in the case of a decision on a RCT in the early phases of development, EARTH would have planned for the use of any interim data from the RCT to support the final "diamond decision" in 2012 in Figure A4. Interim RCT data would have been a much better support of the 2012 decision than the initial business case based on the lowest evidence available.
3. Broader KPIs in the impact case including organisational aspects
4. Monitoring whether possible reductions in overall risk are observed over time through the use of the risk analysis template

This appendix displayed a minimal version of an EARTH assessment demonstrating the use of the impact case template, i.e. the development in KPIs and evidence for the IMT. Also, the decision part of the overview template was used. Further details on how to populate the EARTH model, including examples of additional KPIs and risk elements, and what a counterfactual risk analysis may have looked like for the case (element three and four above), can be found elsewhere [13].