Health Policy 123 (2019) 173–181

Contents lists available at ScienceDirect

Health Policy

journal homepage: www.elsevier.com/locate/healthpol

Testing a new taxonomic model for the assessment of medical devices: Is it plausible and applicable? Insights from HTA reports and interviews with HTA institutions in Europe

Sabine Fuchs [a,*], Britta Olberg [a,b], Matthias Perleth [a,b], Reinhard Busse [a], Dimitra Panteli [a]

[a] Berlin University of Technology, Germany
[b] Federal Joint Committee, Berlin, Germany

ARTICLE INFO

Article history:
Received 12 May 2017
Received in revised form 21 February 2018
Accepted 6 March 2018

Keywords:
Health technology assessment
Medical devices
Taxonomy
Prioritisation
Europe

ABSTRACT

Objective: Medical devices (MDs) encompass a broad and heterogeneous range of technologies. While practices vary considerably across countries, MDs often find application in patient care with little or no evaluation of their effectiveness and safety following market approval. A recently proposed taxonomy of MDs considered devices from the viewpoint of Health Technology Assessment (HTA). The aim of the work presented here was to test its plausibility and applicability by considering real-world HTA practices.
Methods: HTA reports on MDs from European institutions were collected in a systematic manner and the evaluated devices and/or related procedures were matched to a position on the taxonomy. Following this, representatives from 16 European HTA institutions were asked about the usefulness of the taxonomy in semi-structured interviews.
Results: 1237 HTA reports (2004–2015) from 33 European institutions were included in the sample. The majority of reports was on technologies from the taxonomic positions initially estimated as having high relevance. Most of the experts interviewed stated that they found the taxonomy useful, particularly regarding its potential to aid in selecting technologies for assessment and to highlight potential methodological particularities per taxonomic position.
Conclusions: Overall, the distribution of identified reports on the matrix confirmed that the initial estimation of the relevance and necessity of HTA provided in the taxonomic model is plausible. In addition, interviews with representatives of European HTA institutions showed that the taxonomy could be useful.

1. Introduction

Europe is one of the biggest markets for Medical devices (MDs), which encompass a broad and heterogeneous range of technologies. They include *'[. . .] any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes: diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease [. . .]'* [1]. To date, several classification systems for MDs have been developed, ranging from those aiming to facilitate market approval by categorizing MDs according to the degree of risk they pose for patients [1,2], to those used for procurement purposes, which take a perspective relevant to pricing and reimbursement decisions [3–5].

The new EU regulation on MDs is in effect since May 2017 (Medical Device Regulation 2017/745, MDR) [1]. In line with its preceding EU framework for the market approval of MDs (Directives 90/385EEC, 93/42/EEC and 98/79/EC), it mainly addresses the issues of risk and functionality for devices aiming to enter the EU market. Depending on its intended purpose and invasiveness, a MD will be assigned to a risk class; different approval requirements apply depending on class. Devices considered to be in conformity with these requirements can be approved for marketing in the EU and bear the CE (Conformité Européenne) marking. Considering device risk is highly important for patient safety, as potential harms increase the higher the risk level, but it is not the only criterion for assessing MDs [6]. While practices vary considerably across countries [7,8], MDs are often applied in patient care with little or no assessment of their effectiveness or safety following market approval [9,10,11]. While device risk is also important from the

* Corresponding author at: Berlin University of Technology, Department of Health Care Management, Straße des 17. Juni 135, H 80, 10623 Berlin, Germany.
E-mail address: sabine.fuchs@tu-berlin.de (S. Fuchs).

https://doi.org/10.1016/j.healthpol.2018.03.004
0168-8510/© 2018 Published by Elsevier B.V.

Classification criteria of EU-Directives according to risk aspects:		Classification according to the relevance of product & service and reimbursement characteristics (includes OECD Classification of Health Care Functions) + **HTA logic** (color coding)											
		Diagnostic Technologies 409						Therapeutic Technologies 987					
		Assistive technology devices (directly used by patients) A1		Artificial body parts (implanted by medical procedure) B1		Medical devices for the assistance of medical professional C1		Assistive technology devices (directly used by patients) A2		Artificial body parts (implanted by medical procedure) B2		Medical devices for the assistance of medical professional C2	
		Example	*No.*	*Example*	*No.*	*Example*	*No.*	*Example*	*No.*	*Example*	*No.*	*Example*	*No.*
93/42/EEC	I		0			Ophthalmo-scope	9	Wrist splint; Insoles	27			Wound dressing	11
93/42/EEC	IIa	Home blood pressure monitor	13			MRI; Ultrasound	177	Hearing aids	18	Grommets; Dentures	12	TENS device	128
93/42/EEC	IIb					X-ray imager	79	Insulin pumps	17	Intraocular lenses; BAHAs	91	Endovenous laser therapy	341
93/42/EEC	III			Pulmonary artery pressure monitor	1	OCT using catheter	1	Silver dressings	4	Stents; TAVI	126	Intracoronary brachy-therapy	128
90/385/EEC	IV			ICD: heart monitor unit	7					ICD: defibrillator unit	84		
98/79/EC	V	Glucose strip; Pregnancy test	7			HPV test; Genetic tests	115						

Fig. 1. Taxonomy for medical devices and number of technologies identified during the plausibility testing including actual examples from the HTA report pool. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Notes: B1/III: originally grey field (see taxonomy in the Online Appendix); active implantable devices and in vitro diagnostics were assigned separate rows (IV and V); estimation of relevance and necessity of a full-fledged HTA for technologies in different taxonomic positions (cells) including 'green' (high), 'yellow' (intermediate) and 'red' (low), 'grey' cells: no devices, and therefore no assessments, available
Abbreviations: BAHA: Bone Anchored Hearing Aid; HPV: Human Papillomavirus; ICD: Implantable Cardioverter Defibrillator; MRI: Magnet resonance imaging; OCT: Optical Coherence Tomography; TAVI: Transcatheter Aortic Valve Implantation; TENS: Transcutaneous Electrical Nerve Stimulation

viewpoint of HTA [12], other characteristics of the device, such as its purpose or mode of application, can also be decisive for its evaluation. For example, outcomes following the application of MDs used within procedures, such as stents, also depend on the operator and other circumstances; these devices can therefore not be assessed the same way as a medical aid (e.g. glasses) would. Given the broad spectrum of MDs illustrated above, both the necessity and the data requirements for assessment can differ.

In 2015, Henschke and colleagues [13] proposed a taxonomy of MDs that considers devices from the viewpoint of HTA arguing that existing MD classifications would not be fully suitable for HTA purposes as they are based on different underlying logics. They created a matrix in table format, based on relevant aspects from existing MD classification schemes (see Online Appendix and Fig. 1). To reflect the element of risk, risk classes described in the EU-Directives 90/385/EEC, 93/42/EEC and 98/79/EC, which were in effect at that time, were assigned different rows. In addition to risk, roles/functionalities of different device types were taken into account and assigned separate columns, differentiating between assistive technology devices (directly used by patients, e.g. wheelchairs), artificial body parts (implanted by a medical procedure, e.g. stents) and MDs for the assistance of medical professionals (e.g. PET/CT scanner). The matrix also incorporates a distinction between the diagnostic or therapeutic nature of devices, which can be crucial for HTA purposes. Based on this model, the authors considered the relevance and necessity of a full-fledged HTA for technologies in different taxonomic positions and used traffic light color-coding (see Fig. 1) to denote their estimations. 'Grey' cells in the matrix correspond to positions for which no devices, and therefore no assessments, would be expected. The authors suggest that this model can be used both to prioritise devices for assessment and as a guide towards tailoring assessment methods to achieve best results depending on the taxonomic position.

To provide policy-makers and researchers with a realistic orientation guide for approaching the assessment of MDs, the work presented in this article aimed to test the plausibility and applicability of the described taxonomic model by considering real-world HTA practices. It encompasses two components: (i) a quantitative plausibility test of the model based on HTA reports produced by European institutions across a range of medical indications, which have been addressed in the reports, and (ii) a qualitative evaluation of its usefulness and applicability in daily practice based on interviews with HTA experts from a sample of the same institutions.

2. Methods

2.1. Quantitative testing of the taxonomy ('plausibility')

In this part of the study, the ability of the taxonomy to correctly classify MD and MD-related technologies was tested. For this purpose, we used the following approach: In previous work, we had systematically identified European HTA institutions and their publicly available reports of MDs [8]. Approximately 85% (n = 40) of European HTA institutions conduct MD-specific assessments that are publicly available. These reports were collected for the present work and screened for eligibility. Reports were considered eligible if they focused on a MD (be it alone or within a procedure), were based on systematic review methodology, and were carried out between the years 2004 and 2015. Information on the detailed inclusion and exclusion criteria is presented in the Online Appendix. The full available documentation for included reports was downloaded and inventoried in Microsoft Excel.

The technologies evaluated in included reports were classified according to their risk class and functionality elements included in the model and assigned a taxonomic position on the matrix (Fig. 1). Each report was screened by two researchers; discrepancies were solved by involving a third person.

To describe the medical indications addressed in included HTA reports, we used the System Organ Class (SOC) classification provided by the Medical Dictionary for Regulatory Activities [14].

2.2. Qualitative evaluation of the taxonomy (‘applicability’)

This part of the study aimed to test the potential usefulness of the taxonomy for HTA institutions. Semi-structured interviews with persons responsible for MD assessment at 16 European HTA institutions were carried out to explore challenges as well as future ideas about HTA of MDs (the full methodology and main results have been published separately [7]). At the end of each interview, participants were also asked for their input on the following questions concerning the taxonomic model:

1. Could your institution use the developed taxonomy as a guidance?
2. Which are helpful aspects?
3. Do you have any suggestions to refine or improve the taxonomy?

Participants had received the interview guide and the developed taxonomy in advance. The aims and rationale of the taxonomic model were also shortly explained during the interview. We used directed content analysis; the final coding framework combined deductive and inductive themes. To illustrate frequencies of relevant results we use the terms *some (10–30%)*, *many (31–65%)* and *most (66–100%)* to depict how many out of all interviewed institutions mentioned each theme. Additionally, we considered two institutional variables for the analysis: whether the MD assessments produced by each institution are legally binding for decision-making and whether the institutions are themselves responsible for the selection of technologies (or are commissioned by the relevant decision-maker to perform specific assessments). We considered these two variables only for the answers on the usefulness of the taxonomy (question 1) and present the results in tables. Quotes are used to further illustrate the main findings.

3. Results

3.1. Plausibility testing of the taxonomy (quantitative analysis)

3.1.1. General characteristics of identified technologies/reports

After collecting publicly available reports on MDs and applying the inclusion and exclusion criteria (see Online Appendix), 1237 reports from 33 institutions in 17 countries were included in the pool used for the classification. The reports addressed 1376 technologies and were of different types (e.g. ‘technical reports’, ‘systematic reviews’) and length. A detailed illustration of the number of considered reports per institution and the website link to their reports are provided in the Online Appendix.

Out of all devices assessed, 71% (n = 987) had a therapeutic and 29% (n = 409) a diagnostic purpose. Among the latter, 30% (n = 122) were in vitro diagnostics. Moreover, a broad range of indications was captured by reports: oncological (22%, n = 270) and cardiovascular indications (21%, n = 252) represented the largest shares, followed by musculoskeletal and connective tissue disorders (11%, n = 139). A more detailed breakdown can be found in the Online Appendix.

The majority of included reports assessed MDs in the context of a procedure. Individual brand name products were less frequently the focus of assessment, unless this was embedded in the HTA programme design. This was the case for the National Institute for Health and Care Excellence (NICE) Medical Technology Evaluation Programme (MTEP) in the United Kingdom, and the HTA tracks of the Emilia Romagna or Veneto regions in Italy (4,7%, n = 58; for the three institutions).

The most frequently addressed technologies within included reports (Table 1) were radiofrequency ablation (RFA; 3.2%, n = 44), positron emission tomography/computed tomography (PET/CT; 1.7%, n = 24) and brachytherapy (1.5%, n = 21). As can be seen in Table 1, the most frequently addressed technologies mainly represent therapeutic technologies with a high(er) risk class.

3.1.2. Insights related to the plausibility of the taxonomy

The distribution of the technologies addressed in the reports among the cells of the taxonomic model is presented in Fig. 1.

The majority of reports in the sample addressed technologies from the green cells (62.9%, n = 879) whereas relatively few technologies could be assigned to the red cells (3.6%, n = 51). Yellow cells corresponded to 466 technologies (33.5%) in the sample.

Some reports assessed more than one technology (4%, n = 49), for example PET and PET/CT, and sometimes these technologies belonged to different taxonomic positions (2.7%, n = 33). Twenty reports assessed only one technology falling in more than one cell and these were counted twice in Fig. 1 (n = 1396). Two examples are implantable cardioverter defibrillators (B1/IV and B2/IV), compris-

Table 1
Most frequently assessed technologies in the HTA report pool.

Medical device/procedure	No. of reports included	Taxonomic cell (column/row)
Radiofrequency ablation (RFA)	44	C2/IIb*, C2/III*
Positron emission tomography/Computed tomography (PET/CT)	24	C1/IIb
Brachytherapy	21	C2/IIb*, C2/III*
Positron emission tomography (PET)	20	C1/IIa
Computed tomography (CT)	18	C1/IIb
Magnet resonance imaging (MRI)	17	C1/IIa
Robot-assisted surgery	13	C2/IIb
Drug eluting stent (DES)	12	B2/III
Transcatheter Aortic Valve Implantation (TAVI)	10	B2/III
High-intensity ultrasound (HIFU)	10	C2/IIb
Sacral nerve stimulation (SNS)	7	B2/IV
MitralClip	7	B2/III
Intraoperative radiotherapy (IORT)	7	C2/IIb

Notes: *depends on indication addressed in report.

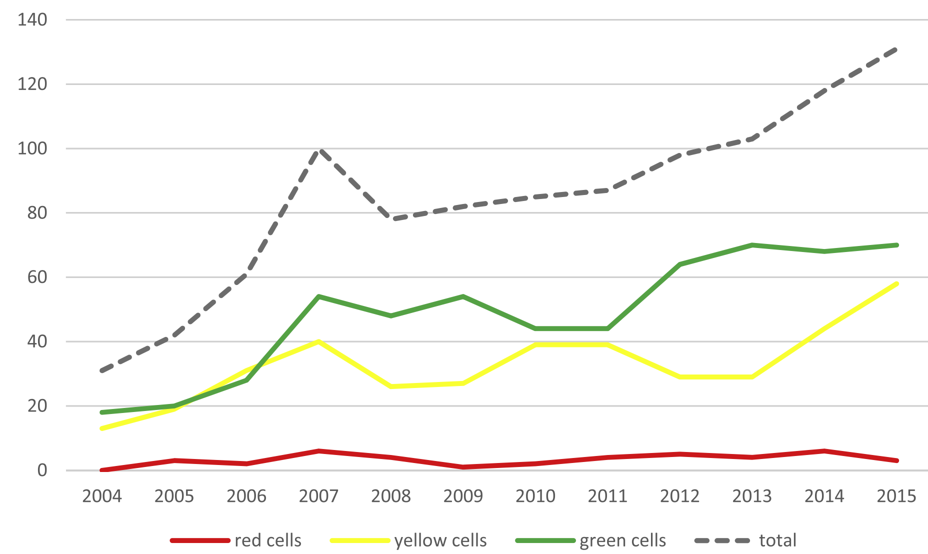

Fig. 2. Reports (n = 1237) by 33 European HTA institutions, total and per color tier used in the taxonomic model (green n = 879, yellow n = 466, red n = 51), 2004–2015. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ing a diagnostic and a therapeutic component, and silver dressings (A2/II and C2/II), which can be used by health professionals and patients alike. For one originally grey cell (B1/III) we identified one report (evaluating the pulmonary artery pressure monitor) whereas no report in our sample could be matched to one originally red cell (A1/I).

With regard to report production over time (see Fig. 2), the number of HTA reports per year increased overall between 2004 and 2015, showing an additional peak in 2007. HTAs on high risk class devices show a strong upwards trend starting in 2011, whereas HTAs assigned to yellow cells decreased in the same year but also started to increase in number as of 2013. These trends might be associated with new EU requirements for market approval of MDs introduced in 2007 and effective as of March 2010. The number of HTAs on low risk MDs (red cells) was substantially lower than those for the green and yellow cells and generally steady over the regarded time span.

3.1.3. *Dealing with unforeseen difficulties*

When MDs consisted of more than one components with differing risk classifications, a decision had to be made regarding their proper taxonomic assignment. We chose to assign such MDs to just one cell (highest class) to stay in line with the borderline manual of the EU [15], which was in effect at that time. For example, the manual suggests that bone-anchored hearing aids be categorised as class IIb; they consist of the sound processor (class IIa according to classification rule 9 of Annex IX of Directive 93/42/EEC) and the implant (class IIb according to classification rule 8 of the same Directive). Additionally, there were some technologies where the main functionality (column A, B or C) was not unequivocal. For instance, innovative medical devices for managing blood glucose levels (e.g. continuous subcutaneous insulin infusion, CSII) can be used directly by patients (column A) but can also be a support tool for medical professionals towards optimising their patients' treatment (column C). Hence, in some cases the technology was assigned to more than one cell (see also the example of silver dressings, above). Finally, there are devices where only one component is (subcutaneously) implanted, such as sensor-augmented insulin pumps (SAPs), consisting of insulin pumps (CSII) and real-time continuous glucose monitoring devices; these could be considered either as implants (column B) or MDs for patients to use (column A). We assigned SAPs to column A.

3.2. *Potential applicability of the taxonomy (qualitative analysis)*

3.2.1. *Usefulness of the taxonomy*

Overall, most of the interview partners stated that the taxonomy is useful but to different degrees. Only four said that it could be used directly within their institution's daily work, while six thought that it would not be directly applicable in their context. Four interviewees generally saw the taxonomy as a helpful tool but without specifying further. One supported the general usefulness of a unified taxonomy for HTA purposes as follows:

> *'I do agree that it would be very appropriate to deliver a new classification. I am not sure the classification delivered would be used by our organization but it would clearly help to have a unified taxonomy.'*

Two interviewees stated that the taxonomy was not useful or that they were not sure about its added value; one of them explained:

> *'Honestly my first reflection, what is the added value for us to have a different kind of taxonomy, we do an HTA on a specific device, and I do not immediately see that added value of having this taxonomy.'*

With regard to the two institution-specific variables, the role of the interviewee's institution within the prioritisation process does not seem to influence their views on the usefulness of the taxonomy, as no difference could be observed (see Table 2). In contrast, the taxonomy was considered useful mainly for other stakeholders in the system (such as the Ministry of Health), more often among participants from institutions whose assessments are not binding for decision-making.

3.2.2. *Helpful aspects and limitations of the taxonomy*

Many interviewees pointed out the helpful or important aspects of the taxonomic model but also its limitations. These insights are summarised in Table 3.

The color-coding linked to the relevance of HTA was the element most frequently mentioned as helpful. The following quote clearly illustrates this theme:

> *'So we have pretty much the same reasoning behind what you have from here. It has to have an economic impact on the healthcare system but it also has to have a big impact on the patient's survival or quality of life*, [. . .]

Table 2
Answers on the usefulness of the taxonomy depending on a) role of the institution's assessments in coverage decision-making in the respective jurisdiction and b) institution's role in the prioritisation process.

	Useful			Not useful	
	. . . for institution	. . . for other stakeholder*	not specified		
Overall analysis	4	6	4	2	16
a) Role of institution's MD assessments in coverage decision-making in resp. jurisdiction					
Binding	1	1	0	1	**3**
Not binding	1	3	4	1	**9**
Not clear**	2	2	0	0	**4**
	4	**6**	**4**	**2**	16
b) Institution's role in the prioritisation process					
Responsible	2	3	2	1	**8**
Not responsible	2	3	2	1	**8**
	4	**6**	**4**	**2**	16

Notes: * including MoH, sickness funds; **depends on either a specific programme or on who has commissioned the report (regional vs national health service).

One interviewee added that even the yellow cells in the taxonomic model, corresponding to groups of devices for which the relevance of a full evaluation is case-specific, make sense:

'[. . .] for the in vitro diagnostics, I think you've got pregnancy tests in yellow. That would be – actually that works, because that's case specific [. . .], something like a pregnancy test and over-the-counter test, they would not normally, you wouldn't do an evaluation. But something [. . .] that a patient will use at home [. . .] they would do an assessment on that. So that does actually work.'

Another participant gave a more detailed explanation on how the taxonomy could be useful for tailored methodological considerations:

'Have sort of a generic evaluation, what you do with diagnostic things, what you do with therapeutic things and then by the subcategories, directly used by the patients and so and then perhaps by the risk categories, if there are specific things.'

Only half of the interviewees commented on perceived limitations of the taxonomic model. Despite the fact that the traffic light color-coding was the most frequently mentioned helpful aspect, the operationalisation used in the model was at the same time also the most frequently cited limitation, as one participant pointed out:

'[. . .] surgical sutures with antibiotics [. . .] this is a device that from a technical point of view, it's really simple, boring maybe for an engineer, but it has huge economic impact.'

Furthermore, individual additional points were raised regarding important elements not considered in the taxonomic model, such as the speed of modification of MDs (incremental innovation).

3.2.3. Suggestions for refinement

Several interviewees made suggestions for the improvement and/or refinement of the taxonomic model, for example including additional rows/columns for prognostic and screening devices within the diagnostic part of the matrix, renaming the category 'Artificial body parts' (as it also actually encompasses the modification of organs, not only their replacement), clarifying that row IV does not correspond to a risk class defined according by the EU regulatory framework, explaining the rows and performing further reliability tests.

4. Discussion

4.1. Main findings

The taxonomic model developed by Henschke et al. [13] combines different device classification schemes that have been used for different purposes in an effort to provide a foundation for considering MDs from an HTA perspective. The aim of the present work was to test its plausibility and applicability for daily practice. According to our results, the distribution of the identified HTA reports on the model generally confirms that the taxonomy is plausible: The majority of reports in the sample addressed technologies from the green cells, where HTA was considered of high relevance, and relatively few reports were available for the red cells considered of low relevance. Only one report was identified for one grey cell where no reports were expected. In addition to the plausibility testing, opinions from experts working at 16 European HTA institutions indicated that the taxonomy is generally seen as useful, primarily for commissioning institutions, decision-makers or other stakeholders (e.g. sickness funds) and mainly for institutions were the assessment process is not binding. The traffic light color-coding used to denote the relevance of HTA for different device groups was the taxonomic element most frequently considered useful, particularly in the context of prioritising devices for assessment; however, the same concept was also perceived as a limitation for certain systems. Suggestions for refinement included for example adding rows/columns for prognostic and screening devices.

4.2. The taxonomy and its elements: comparison of the results to the literature

To our knowledge, the tested taxonomy is the first and only model of MDs from a viewpoint of HTA. As mentioned in the introduction, EU regulation on the market approval of MDs focuses primarily on the elements of device risk and functionality. The new regulation 2017/745 reinforces the importance of risk consideration: to receive market approval, certain device classes (e.g. implantable class III devices) are to be subject to a new assessment procedure which will involve the newly created Medical Device Coordination Group (designated by Member States) and the Assessment Committee for Medical Devices (designated by the European Commission) while high-risk devices will be assigned to specialised Notified Bodies. The EU further strengthens its risk-dependent requirements for market approval by suggesting in the instruction for borderline cases [15] that technologies consisting of more than one component with differing risk classifications should be categorised in the highest risk class.

From the viewpoint of HTA, risk can play an important role in the prioritisation of technologies for assessment. In some systems, this is already reflected in legislation: for example Germany introduced early benefit assessment for high-risk MDs in 2016 (§137 h Social Code Book V). Furthermore, based on the overall distribution of identified HTA reports in our study, European HTA institutions appear to focus on higher risk class MDs already, particularly for devices used by medical professionals in therapeutic interventions.

Table 3
Usefulness of the developed taxonomy as a guidance for institution – helpful/important aspects and limitations.

Theme	Summary of theme	Quote(s)	Frequency*
1) Helpful/important aspects			
Prioritisation (traffic light color-coding with regard to relevance of HTA)	- Helpful with selecting technologies for assessment - Relevance of an HTA or full-fledged HTA depending on taxonomic position (green = high; red = low; yellow = case-specific) - Particularly HTA of MDs with high risk (green fields)/high economic relevance - Consideration of volume effects	*'I thought maybe originally that it might be very useful in the topic selection phase, because quite a lot more technologies come through there, and just in terms of we have thought about how can we categorise them.'* *' Usually, [the high-risk MDs] are the more expensive ones and you also have to take volume effects into account.'*	Many (6)
Separation between diagnostic and therapeutic MDs	- Taxonomy takes diagnostic and therapeutic nature of devices into account	*'We have different types of medical devices for treatments, for screening, also for prognosis, and according to your taxonomy you split it into diagnostic and therapeutic.'* *'The separation in diagnostic and therapeutic MDs is helpful while you could differentiate the methodological approaches.'*	Some (5)
Potential for tailored methodological consideration	- Usefulness seen in potential methodological considerations depending on specific taxonomic position	*'[. . .] but you have different approaches for diagnostics and therapeutics, yes. I think it looks interesting. It would be interesting to see even more in the next stage, what the implications for the methods would be'* *'I can imagine that for all implants it is useful to have a long-term perspective on the patients and their outcomes'*	Some (5)
Purpose-based classification of devices in three groups	- Categorisation of MDs in three groups based on their purpose (1. MDs for patients to use, 2. Artificial body parts and 3. MDs used by medical professionals) seen as helpful to consider certain aspects	*'I also think it's good to separate things that are directly used by patients.'*	Some (5)
Risk potential/invasiveness	- Risk categories according to EU directives taking the invasiveness of MDs into account	*The scheme, I can understand that: for the green cells with high-risk MDs, it is rather indicated to make an evaluation [. . .].'*	Some (3)
Unified taxonomy from HTA perspective	- Common taxonomy for HTA community seen as useful	*'Fortunately this taxonomy is of utmost importance for HTA's world.'*	Some (2)
2) Limitations			
Prioritisation (traffic light color-coding with regard to relevance of an HTA)	- MDs from the red cells of interest (e.g. high volume & thus high economic impact)	*'To take the color coding for example of the thermometry example again. We think we would have to assess thermometry, and it's because it's used in lots of different ways in different places, and it's low cost and high volume, and there's lots of critical outcomes related to it.'*	Some (4)
Potential for tailored methodological consideration	- Uncertainty about usefulness seen in potential methodological considerations depending on specific taxonomic position	*'Unsure yet, if the taxonomy could be a helpful methodological support tool'*	Some (3)
Complexity of MDs	- Complexity of MDs could not be considered in one taxonomy	*'I don't know to which extent, how it would be used in the case of diagnostics [. . .] I'm thinking about the example of genetic diagnostic technologies, the diagnosis is in the beginning not so important. But finally if you have to decide on if it is positive or negative then you have to establish a treatment that will influence the final outcomes. I don't know how you are going to manage that.'*	Some (2)
Combines already existing classification	- No new classification, just based on existing ones	*'You crossed, matched several classification systems to have a new way to classify but it's not a new classification system, it's just a taxonomy.'*	Some (2)
Artificial construct	- Not for practical use, it's too theoretical	*'I think it is too artificial, because in medicine you don't think in boxes or categories, you think of patients, that they get better outcomes.'*	Some (1)
Modification level of MDs	- Modification level of MDs is an important component which is not considered by the taxonomy	*'What's exciting is the distinction between Incremental and breakthrough innovation [. . .] So if I read here cardiac stent, then I would ask, do those incorporate all stents or just drug eluting stents or from when it is an antibody-coated stent? So that would be something different, the degree of modification is in part decisive.'*	Some (1)

Note: *Categorisation of themes in some (10–30%), many (31–65%) and most (66–100%), see [7].

This is not surprising, as it is in line with the remit of many European HTA institutions, which mainly encompasses the assessment of high risk MDs or MDs with a high economic impact [7]. Nevertheless, experts working at some of these institutions also stressed the importance of assessing lower risk class devices when utilization (volume) is high in the respective country. Based on our HTA report pool, only 10 out of 33 institutions produced reports on such devices. This might be also attributed to the fact that in some countries different institutions are responsible for the assessment of lower risk class MDs or medical aids, which may not have been captured in our pool.

To reconcile these notions, a differentiated approach regarding the extent of HTA is conceivable: for lower risk class devices, a more simplified assessment encompassing relevant elements might be sufficient, whereas high risk class devices need a full-fledged HTA.

In addition to risk and functionality, the purpose-based categorisation separating assistive technology devices from artificial body parts and devices for the assistance of medical professionals within the taxonomy is another important component from an HTA perspective. For the most part, devices falling into the latter two categories are used within procedures and the outcomes of their application depend on the operator as well as other circumstances. It is known that changes in operator performance (learning curves) affect outcomes regarding both effectiveness and costs and can lead to bias in the results if not explicitly considered in HTAs [16]. Thus, the assessment of such MDs needs to consider additional dimensions, which are not relevant in the evaluation of assistive technology devices mainly consisting of one product. On the other hand, learning effects are also relevant in the first category, as outcomes here depend on the patients using the devices (e.g. an inhaler) correctly. Thus, sufficient patient training and correction of misuse should also be emphasized in HTAs on assistive technology devices where appropriate. However, this is not always a simple task: methodological approaches towards and hurdles in considering learning curves in HTAs have already been outlined in several articles [7,17,18].

Furthermore, differentiating between column B and column C devices is important because all MDs in column C serve as a temporary tool to assist medical professionals while column B devices modify or replace the function of organs or tissue. Consequently, for the latter a long-term perspective of effects should be considered in evaluation [17]. To address these issues, specific methodological tools are required to guide the evaluation of corresponding MDs. EUnetHTA developed a guide for the assessment of therapeutic MDs, which takes into account aspects deriving from context and user dependence, which are of greater importance in the evaluation of devices compared to that of drugs; the guide also considers the issue of incremental development of MDs [19]. A recently published article by Schnell-Inderst and colleagues [20] specifically developed recommendations for systematic reviews of comparative effectiveness of therapeutic MDs as part of HTA. In previous work [7,8] we showed that only a limited number of methodological documents tailored to MDs are available from European HTA institutions. Some of those identified documents [21,22] already consider some of the aforementioned methodological particularities. As already suggested in the literature [7,17], adding these elements to existing institutional documents could make more sense than creating a fully new document for (certain) MDs only.

The separation of diagnostic and therapeutic devices within the taxonomy was found to be a useful and important component, because different methodological approaches need to be adopted depending on the nature of the device. This is in line with European standards and guidelines such as the different iterations of the HTA Core Model® developed by EUnetHTA, including one for diagnostic technologies, which explicitly addresses the issues of clinical utility and clinical validity of diagnostic tests [23] or the methodological guideline for therapeutic MDs already mentioned above [19]. The aforementioned, device-specific methodological documents developed by European HTA institutions set the focus mainly on diagnostic devices or medical (biomarker) tests [8]. Furthermore, the IDEAL framework, which describes the stages in the life cycle of interventional therapy innovations and recommends study designs for each stage, was recently adapted to meet specific criteria for MDs [24].

4.3. *The taxonomy as a whole: Comparison of the results to the literature and further steps*

With its aim of providing decision-makers and researchers with an orientation guide/tool for a systematic identification and selection of MDs and procedures for HTA purposes and identifying potential methodological issues, the taxonomy addresses an important issue and ongoing topic of discussion within the HTA community [20,25–28]. Difficulties in selecting technologies for assessment due to the broad spectrum of MDs have been identified as a procedural challenge [7]. According to a European survey, an explicit process for priority setting was employed by 34% of 47 included HTA institutions (vs. 26% where the commissioning institution prioritises, 10% with no explicit procedure and 30% without publicly available information). In these cases, prioritisation was most commonly based on medical-scientific criteria, on criteria related to the epidemiological significance of the condition as well as on economic criteria. These institutions represent equally independent research entities (academic or functioning as governmental institutions) or national/regional governmental institutions (see the Online Appendix for [8]). This implies that more European HTA institutions prioritise by themselves rather than the commissioning institution indicating that the taxonomy could be used directly by them as a support tool for selecting relevant technologies for assessment.

Based on the results of the qualitative analysis, the taxonomy is mainly seen as potentially useful for other stakeholders (e.g. MoH) and not for the HTA institutions themselves. This is mainly attributed to the fact that those institutions are not responsible for prioritisation. For a better definition of the potential users of the taxonomy and the optimal time for its application within the decision-making process further testing is required involving also other stakeholders (e.g. payers). The potential use of the taxonomic model as a tool to facilitate division of labor among institutions responsible for the assessment of different types of MDs could also be explored in this context.

In contrast, work programmes within HTA institutions were developed such as MTEP at NICE in the UK, establishing their own specific procedure for prioritisation [22]. For those institutions, the taxonomy can serve mainly as a tool to think of certain aspects for the assessment. This potential of the taxonomic model to signpost particular methodological considerations depending on the taxonomic position was considered as a helpful aspect by our interviewees. Beyond the issues already raised above (e.g. long-term effects for column B devices and learning curves for column C), a potential approach for devices with both diagnostic and therapeutic functions suggested by the authors of the taxonomic model was for the two components to be considered separately. We did not identify any reports already using this approach in our sample. There are further aspects that pertain to individual device groups, such as battery life for active implantable devices (see row IV in taxonomy and [29]), which may pose additional challenges to be considered and can be signposted for the corresponding taxonomic positions. Moreover, for some device types (e.g. risk class I devices), it might be appropriate to accept lower level of evidence/non-randomised studies as (part of) the basis for assessment.

The process of assigning the technologies to cells was not always straightforward (e.g., some technologies could have been assigned to more than one cell). This displays on one hand the nature of these devices but on the other hand the taxonomy's limits and thus relative lack in flexibility, as raised by some interview partners. In the same direction, devices were sometimes evaluated in combination in the context of a procedure, while risk classification rules set out by the EU Directives apply to individual products. This also supports further testing of the taxonomy. However, before such an initiative is put into motion, the taxonomy would need to be adapted to reflect the updated version of the EU's regulatory framework (MDR), which succeeded its development. The main change would entail modifying the row assignment and underlying classification criteria (e.g., a new risk classification system for in vitro MDs; a draft updated version and explanatory notes can be found in the Online Appendix). The new classification criteria might have an impact on the distribution of reports across the taxonomy, assuming that even more reports will be assigned to the green cells.

4.4. Strengths and limitations of this work

The independent filtering of reports by two researchers as well as the consultation of one more in undecided cases aimed to reduce the potential risk of errors during the screening process. Previous experience in the area of MD assessment and the language skills of the researchers contributed to the construction of a comprehensive yet manageable report pool. In addition, the systematic approach used to identify the institutions involved in MD assessment as a basis for collecting HTA reports and contacting experts for the interviews, lead to a representative sample of reports and a sample of interviewed institutions with a varied level of experience in MD assessment, allowing for a rounded review of the taxonomic model. In few cases, insufficient information about the evaluated technology was given in the reports; thus, despite our best efforts, we cannot exclude the possibility that some reports may have been assigned to a wrong cell in the taxonomy. Furthermore, for this work no distinction was made between tests with a purely diagnostic compared to a prognostic nature, which also detracts from the granularity of results.

5. Conclusions

The results of the quantitative testing generally confirm that the taxonomy is plausible and the results of the qualitative testing support its usefulness and/or applicability in daily HTA practice. Based on these findings, we propose that a revised version of the taxonomy could i) serve as a helpful tool for institutions tasked with selecting technologies for assessment, especially those who have not yet developed specific and structured programmes to do so, and ii) trigger some fruitful discussions about the prioritisation of MDs itself. The developed taxonomy aimed to provide the conceptual framework for a classification in the logic of HTA. It is to be viewed as a dynamic model that can incorporate changes or be used as a basis for device-specific methodological guidance. Future work will focus on these elements.

Conflict of interest

BO and MP work for the Federal Joint Committee (G-BA), which is the highest decision-making body of the joint self-government of physicians, dentists, hospitals and health insurance funds in Germany. One of its tasks is issuing directives determining the benefit basket of the statutory health insurance funds. SF, RB and DP report no conflict of interest.

Acknowledgements

This research was funded by the European Union's Seventh Framework Programme and undertaken under the auspices of the ADVANCE_HTA project (Grant number 305983; www.advance-hta.eu). The European Commission had no role in the study design, collection and analysis of data, writing of the report or submission of the paper for publication. The authors would like to thank all interviewees for their valuable time and input; the ADVANCE_HTA consortium and our students Lisa Becker, Fabian Berkemeier, Helene Eckhardt, Tessa Creutz and Julian Ramirez for their support.

Online Appendix

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.healthpol.2018.03.004.

References

[1] European Union (EU): Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. Official Journal of the European Union, L 117, Vol. 60, 5 May 2017.

[2] European Union (EU): Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. Official Journal of the European Union, L 117, Vol. 60, 5 May 2017.

[3] OECD/Eurostat/WHO. A System of Health Accounts 2011: Revised edition. Publishing, Paris: OECD; 2017. Available from: https://doi.org/10.1787/9789264270985-en [Accessed 21.12.17].

[4] Busse R. Decision-making. The link between reference pricing and procurement. Brussels; 2008, 15.10.2008. Available from: http://www.mig.tu-berlin.de/fileadmin/a38331600/2008.lectures/Brussel_2008.10.15_rb_EHTI.pdf [Accessed 22.09.17].

[5] Perleth M, Busse R, Schwartz FW. Regulation of health-related technologies in Germany. Health Policy 1999;46:105–26.

[6] Banta D, Kristensen F, Jonsson E. A history of health technology assessment at the European level. International Journal of Technology Assessment in Health Care 2009;25(S1):68–73.

[7] Fuchs S, Olberg B, Panteli D, Perleth M, Busse R. HTA of medical devices: challenges and ideas for the future from a European perspective. Health Policy 2017;121(3):215–29, http://dx.doi.org/10.1016/j.healthpol.2016.08.010.

[8] Fuchs S, Olberg B, Panteli D, Busse R. Health technology assessment of medical devices in Europe: processes, practices, and methods. International Journal of Technology Assessment in Health Care 2016;32(4):246–55, http://dx.doi.org/10.1017/S0266462316000349.

[9] Tarricone R, Torbica A, Ferré F, Drummond M. Generating appropriate clinical data for value assessment of medical devices: what role does regulation play? Expert Review of Pharmacoeconomics & Outcomes Research 2014;14(5):707–18, http://dx.doi.org/10.1586/14737167.2014.950233.

[10] Hulstaert F, Neyt M, Vinck I, et al. Pre-market clinical evaluations of innovative high-risk medical devices in Europe. International Journal of Technology Assessment in Health Care 2012;28(3):278–84, http://dx.doi.org/10.1017/S0266462312000335.

[11] Olberg B, Fuchs S, Panteli D, Perleth M, Busse R. Scientific evidence in health technology assessment reports: an in-depth analysis of european assessments on high-risk medical devices. Value Health 2017;20(10):1420–6, http://dx.doi.org/10.1016/j.jval.2017.05.011.

[12] Krüger LJ, Evers SM, Hiligsmann M, Wild C. Divergent evidence requirements for authorization and reimbursement of high-risk medical devices – the European situation. Health Policy and Technology 2014;3(4):253–63, http://dx.doi.org/10.1016/j.hlpt.2014.08.005.

[13] Henschke C, Panteli D, Perleth M, Busse R. Taxonomy of medical devices in the logic of health technology assessment. International Journal of Technology Assessment in Health Care 2015;31(5):324–30, http://dx.doi.org/10.1017/S0266462315000562.

[14] MedDRA Maintenance and Support Services Organization. Introductory Guide. MedDRA Version 14.0; 2011. Available from: http://www.who.int/medical_devices/innovation/MedDRAintroguide_version14_0_March2011.pdf [Accessed 24.04.16].

[15] European Commission Manual on Borderline and Classification in the Community Regulatory Framework for Medical Devices. Version 1.17 09-2015.

[16] Varabyova Y, Blankart CR, Schreyögg J. The role of learning in health technology assessments: an empirical assessment of endovascular aneurysm repairs in german hospitals. Health Economics 2017;26:93–108, http://dx.doi.org/10.1002/hec.3466.

[17] Tarricone R, Torbica A, Drummond M. Challenges in the assessment of medical devices: the MedtecHTA project. Health Economics 2017;26:5–12, http://dx.doi.org/10.1002/hec.3469.

[18] Taylor RS, Iglesias CP. Assessing the clinical and cost-effectiveness of medical devices and drugs: are they that different? Value Health 2009;12(4):404–6, http://dx.doi.org/10.1111/j.1524-4733.2008.00476_2.x.

[19] European network for Health Technology Assessment (EUnetHTA). Therapeutic medical devices. EUnetHTA; 2015. Available from: https://eunethta.fedimbo.belgium.be/sites/5026.fedimbo.belgium.be/files/WP7-SG3-JA2-GL-therap_medical_devices.pdf [Accessed 15.02.16].

[20] Schnell-Inderst P, Iglesias CP, Arvandi M, Ciani O, Matteucci Gothe R, Peters J, et al. A bias-adjusted evidence synthesis of RCT and observational

data: the case of total hip replacement. Health Economics 2017;26:46–69, http://dx.doi.org/10.1002/hec.3474.

[21] Haute Autorité de Santé (HAS). Methodological choices for the clinical development of medical devices. HAS; 2013. Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/2014-03/methodological_choices_for_the_clinical_development_of_medical_devices.pdf [Accessed 3.08.16].

[22] National Institute for Health and Care Excellence (NICE). Medical technologies evaluation programme methods guide. NICE; 2011. Available from: http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-medical-technologies/Medical-technologies-evaluation-programme-methods-guide.pdf [Accessed 3.09.16].

[23] European network for Health Technology Assessment (EUnetHTA). Joint Action 2, Work Package 8. HTA Core Model ® version 3.0. EUnetHTA; 2016. Available from: www.htacoremodel.info/BrowseModel.aspx. [Accessed 27.03.17].

[24] Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. BMJ 2016;353:i2372, http://dx.doi.org/10.1136/bmj.i2372.

[25] Husereau D, Henshall C, Sampietro-Colom L, Thomas S. Changing health technology assessment paradigms? International Journal of Technology Assessment in Health Care 2016;32(4):191–9.

[26] Kosherbayeva L, Hailey D, Kurakbaev K, Tabarov A, Kumar A, Gutzskaya G, et al. A process of prioritizing topics for health technology assessment in Kazakhstan. International Journal of Technology Assessment in Health Care 2016;32(3):147–51.

[27] Noorani HZ, Husereau DR, Boudreau R, Skidmore B. Priority setting for health technology assessments: a systematic review of current practical approaches. International Journal of Technology Assessment in Health Care 2007;23(3):310–5.

[28] Henshall C, Oortwijn W, Stevens A, Granados A, Banta D. Priority setting for health technology assessment: theoretical considerations and practical approaches. A paper produced by the priority setting subgroup of the EUR-ASSESS Project. International Journal of Technology Assessment in Health Care 1997;13:144–85.

[29] Dean J, Sulke N. Pacemaker battery scandal. BMJ 2016;352:i228, http://dx.doi.org/10.1136/bmj.i228.