Francisca Crispi, Huseyin Naci, Eva Barkauskaite, Leeza Osipenko, Elias Mossialos

Assessment of devices, diagnostics and digital technologies: a review of NICE medical technologies guidance

Article (Accepted version) (Refereed)

Original citation:
Crispi, Francisca and Naci, Huseyin and Barkauskaite, Eva and Osipenko, Leeza and Mossialos, Elias (2018) Assessment of devices, diagnostics and digital technologies: a review of NICE medical technologies guidance. Applied Health Economics and Health Policy. ISSN 1175-5652 (In Press)

DOI: https://doi.org/10.1007/s40258-018-0438-y

TITLE PAGE

Authors:

Francisca Crispi [1], Huseyin Naci [1], Eva Barkauskaite [2], Leeza Osipenko [2], Elias Mossialos [1]

Title:

Assessment of Devices, Diagnostics and Digital Technologies: A review of NICE Medical Technologies Guidance

Authors affiliations and address:

[1.] LSE Health, Department of Health Policy, London School of Economics and Political Science.

Houghton Street, London, WC2A 2AE, UK.

[2.] NICE Scientific Advice, National Institute for Health and Care Excellence
10 Spring Gardens, London, SW1A 2BU, UK.

Corresponding Author:
Huseyin Naci. LSE Health. London School of Economics and Political Science, 20 Houghton St, London WC2A 2AE, England.

Telephone number: +44 20 7955 6874

Funding:

No financial support was received.

Conflict of Interest and Disclaimer:

FC, HN, and EM have no conflicts of interest to report. EB was an employee of NICE Scientific Advice for the duration of this study, and LO currently leads NICE Scientific Advice. The opinions expressed in this article are those of the authors and do not reflect the view of NICE.

Acknowledgements: The authors would like to thank Christopher Pomfrett and Bernice Dillon from NICE for their valuable inputs on an earlier version of this manuscript.

Word count: 5,663 (excluding Abstract, References, Tables and Figures).

ABSTRACT

Background: The Medical Technologies Evaluation Programme (MTEP) of NICE in England aims to evaluate medical devices that are deemed to be cost-saving or cost-neutral and produce medical technology guidance (MTG) to encourage their adoption.

Objective: To review the MTGs since MTEP's inception in 2009 until February 2017.

Methods: One researcher assessed all published MTGs and extracted data on the clinical and economic evidence supporting each technology. The NICE Committee's decision outcome for each assessment was also recorded. A qualitative analysis was performed on technologies that were not supported for adoption to identify the main drivers of the decision.

Results: 31 MTGs were reviewed. The committee fully supported the medical devices in 14 MTGs, 11 were partially supported and 6 not-supported. 58% of the MTGs had no RCT data available and the main source of evidence came from non-experimental studies. There was no statistically significant difference in the average number of RCTs and non-experimental studies between the fully supported, partially supported and not supported technologies. Whilst all the fully supported MTGs demonstrated cost-saving results only 50% of the not-supported MTGs did. The sponsor estimated a higher average cost-saving than the EAC in most of cases (20/31). The qualitative evaluation suggests that the main drivers for negative decisions were the quantity or quality of studies, and cost incurring results in the economic evaluation.

Conclusions: The main drivers of the decision-making process are the quality and quantity of the submitted evidence supporting the technologies, as well as the economic evaluation results.

Key Words: Medical Technologies, Technology Assessment, NICE, Health Technology Assessment, NHS.

Key points for decision makers:

- From its inception in 2009 until February 2017, NICE developed 31 Medical Technologies Guidance (MTG), fully supporting the adoption of 14 technologies and partially supporting 11.
- The decision to recommend a technology relied on the quantity and quality of the available clinical evidence, as well as the results of the economic evaluation analysis.
- Most supported MTGs were not supported by randomized controlled trials, and instead reied on other sources of evidence such as non-randomized studies, case reports, and expert opinion.

1.1 ACRONYMS

CG	Clinical Guidelines (Programme)
DAP	Diagnostics Assessment Programme
EAC	External Assessment Centre
HST	Highly Specialised Technologies
HTA	Health technology assessment
IP	Interventional Procedures
MTAC	Medical Technologies Advisory Committee
MTEP	Medical Technologies Evaluation Programme
MTG	Medical Technology Guideline
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
RCT	Randomised Controlled Trial
TAP	Technology Appraisal Programme

1.2 INTRODUCTION

1.2.1 Background

Health technology assessment (HTA) has been increasingly adopted worldwide as a mechanism for efficient resource allocation in health care. HTA is defined as "the systematic evaluation of the properties and effects of a health technology, addressing the direct and intended effects of this technology, as well as its indirect and unintended consequences, and aimed mainly at informing decision making regarding health technologies." [1,2]. The National Health Service (NHS) in the United Kingdom, an early adopter of HTA, holds leadership position in this field. The HTA programme in the UK was launched in 1993, aiming to foster the use of systematic reviews to guide the adoption of health technologies in the NHS [3]. The HTA programme preceded the National Institute for Health and Care Excellence (NICE, originally named National Institute of Clinical Excellence), which was set up as an independent public body in 1999 [3]. NICE's role is to improve outcomes for people using the NHS and other public health and social care services, by producing evidence-based guidance and advice for health and social care practitioners and the public.

NICE is currently responsible for the evaluation of new health technologies. The Medical Technologies Evaluation Programme (MTEP) identifies and selects medical technologies that would benefit from national evaluation. Set up in 2009 [4], the MTEP aims to critically appraise the evidence supporting medical devices and diagnostics and produce guidance to encourage their adoption, promoting a rapid incorporation of beneficial medical technologies in the NHS and "collaborative research" between sponsors and the NHS [5]. The MTEP routes the selected technologies to 6 different technology evaluation programmes. [5]: Technology Appraisals (TAP), Highly Specialised Technologies (HST), Diagnostics Assessment (DAP), Medical Technologies Guidance (MTG), Clinical Guidelines (CG) and Interventional Procedures (IP). The TAP evaluates technologies such as drugs, devices, systems of care and others, prioritising these technologies based on criteria such as the target population, disease severity, and economic impact [6]. The HST evaluates treatments for ultra-orphan conditions. The DAP is focused on the evaluation of diagnostic technologies that could improve health outcomes but are likely to imply higher costs for the NHS [7]. The CG programme provides evidence-based recommendations for services and activities in diverse settings, from the clinical management of conditions, medicines, social care, to interventions oriented to improve health at the community-level [8]. The IP programme evaluates interventional procedures in terms of efficacy and safety [9]. Technologies are selected for MTG if they are likely to be cost saving or cost neutral, and can be evaluated as a single technology and in a short time scale [5]. This study focuses on the MTGs produced by the MTEP.

Medical devices differ from other medical interventions, such as pharmaceuticals [10,4,11]. In terms of effectiveness, medical devices may change their effectiveness through time due to modifications, they are user-dependent and have a "learning curve", which depends on the technology uptake [12]. Also, the effectiveness of the diagnostics depend on subsequent treatment and the medical devices may be used for different medical conditions in different settings – with different effectiveness. Regarding clinical evidence, it is usually limited, generally lacking randomised studies. This may be because the product life spans of devices are generally shorter than for pharmaceuticals, which may imply a negative incentive for evidence production [11]. The benefits at a health system level depend on "organizational factors" (setting, staff) and costs may involve running as well as procurement costs. Finally, medical devices prices are usually more dynamic than other interventions [13].The MTG was established in recognition of these special characteristics of the medical devices, which made them less suitable for the established pathways [4].

1.2.2 Medical Technologies Guidance

For medical technologies routed to the MTEP for MTG , the MTEP produces an adoption recommendation in a Medical Technology Guidance (MTG) [5]. However, unlike TAP

recommendations, a positive adoption recommendation from MTEP does not carry a funding mandate for NHS England.

The current process of the MTEP programme is described in *Figure 1*. The MTAC, composed of around 25 voluntary members including clinicians with expertise with the respective medical technologies, scientists, healthcare regulations experts, and representatives with experience in the medical technologies industry, identifies and selects medical technologies [5]. The sponsors (usually product developers) present their medical technology for evaluation to NICE and then completes a submission based on the defined scope, providing clinical evidence and an economic evaluation model [5]. The External Assessment Centres (EACs), which are typically academic bodies commissioned by NICE, check or assess the clinical evidence and economic evaluation presented by the sponsor [5]. Also, expert advisers who offer additional information and clinical opinions to the committee [10]. The public may attend the committee meetings and comment at the consultation on the draft guidance.[5]. Finally, the MTAC evaluates all these inputs and generates the MTG.

The MTEP has a Research Commissioning workstream, which is part of NICE's Research Recommendations process [14]. The Research Commissioning workstream facilitates research to address the gaps identified by the Medical Technologies Advisory Committee (MTAC) in the guidance recommendations [15]. Evidence generated is expected to inform future reviews of the MTGs.

1.2.3 Previous research

Several previous studies have reviewed NICE's MTEP programme. Cowles et al in 2017 compared different NICE HTA programmes, appraising their impact on the allocative efficiency in the NHS. This study showed that MTEP differs from other NICE guidance-producing programmes in important ways. First, the manufacturer prepares the evidence report (with a review by the EAC), which differs from the DAP, CG and TAP streams where the evidence is presented by the EAC or the evidence review group (ERG). Second, unlike other programmes, the MTEP does not have an appeal process. Third, MTEP does not use a reference case (unlike IP, TAP and HST programmes) [16]. Green et al in 2014 also compared the MTEP with other HTA programmes. By comparing the methods guides, the authors reported that in the TAP, RCTs are stated as the most valid source of evidence, and in the MTEP no preferred source is defined. Also, unlike the TAP process, expert advisors are considered in the MTEP as part of the evidence synthesis process, which Green et al 2014 found to be potentially unreliable [17]. Chapman et al in 2014 did an initial assessment of the programme, through an evaluation of the MTEP aims and initial outcomes. The authors concluded that some barriers persist. For instance, the small manufacturers are underrepresented in the programme submissions. Also, they found that the programme has not effectively reduced the evaluation timeline compared to other programmes [4]. Finally, Alshreef et al in 2016 reviewed the economic evaluations of 12 MTGs, identifying the criticisms raised by the EAC to the sponsor's submission and the committee's considerations. They found that the sponsor's identification and measurement of costs and consequences accounted for 42% of the EAC's criticisms [11].

1.2.4 Objectives

To the best of our knowledge, no study so far has reviewed the full sample of published MTG to date (2017), aiming to evaluate the evidence supporting MTAC assessments, and the main drivers behind the decision on adoption of the technologies. This study addresses this gap in the literature.

This work is based on a review of published MTGs. It summarizes the final recommendations and cases for adoption of the evaluated guidance reports and characterises the clinical evidence and economic evaluation presented in the guidance reports.

1.3 METHODS

All NICE MTGs published between 2009 to February 2017, from the NICE electronic database, were included. When information was missing in the guidance reports, data from the associated EAC Report and the sponsor's submission were extracted.

1.3.1 Data extraction

A standardised electronic data extraction form was created in Microsoft Excel. Information about each guidance was extracted. Each guidance was assessed by a single reviewer and afterwards discussed among all the investigators.

1.3.2 Parameters evaluated

The guidance reports were reviewed to extract both general information about the clinical area and technology and detailed information on the clinical and economic evidence supporting it *(Supplement 1).*

First, the clinical area of each MTG was evaluated and categorised. For the clinical care setting, the following categories were considered: Hospital, Secondary Care, Primary Health Care and Community Level. Regarding the committee's suggestion on routine adoption, 3 categories were adopted by merging the 5 categories of recommendations defined in the methods guide [4]: Fully Supported (FS) (which refers to: Case for adoption is fully supported), Partially Supported (PS) (which refers to: Case for adoption is partially supported; or Case for adoption is partially supported and technology has potential to provide significant patient or healthcare system benefits) and Not Supported (NS) (which refers to: Case for adoption is not currently supported but technology has potential to provide significant patient or healthcare system benefits; or Case for adoption is not supported and technology does not have potential to provide significant patient or healthcare system benefits). When the medical technology was categorised as "Partial Support", a description of partial support was noted.

Also, information about the Research Commissioning workstream was extracted. Some of the partially-supported and not-supported MTGs are selected for Research Recommendations [10]. The decision to select MTGs for further research typically depends on the nature of the evidence gap, information about ongoing research, ethical or practical aspects of conducting additional research, and the likely costs and benefits of undertaking studies. The percentage of MTGs selected for research recommendations (section 1 of the MTG) and the number of studies that were completed was documented. Further, any subsequent reviews of the MTGs (if available) was described.

Second, the clinical evidence submitted by the sponsor, studies added or rejected by the EAC, and the final evidence base available on the technology were evaluated. The evidence was categorised as RCT, non-experimental (such as case-control, cohort and cross-sectional studies) and unpublished/abstract studies. To capture all the relevant evidence, the unpublished/abstract RCTs were included in the RCT count. Only patient-based studies were included (and *in vivo* studies were excluded). Ongoing studies were not incorporated.

Third, a reduced form of the CHEERS checklist was used to extract information on economic parameters [18]. Both from the sponsor's submission or the EAC report, best and worst cost-saving (or cost-incurring) values per patient in Great British Pounds (GBP) were included. These values were extracted from the guidance, and were standardised to represent the best and the worst-case scenarios. In some reports, these worst and best-case values could arise from comparing the technologies to different comparators or in different sub-populations. For example, in MTG17, the minimum value is the cost saving when comparing the technology, Debrisoft, to a comparator Hydrogel while the maximum value is the cost saving when comparing it to other comparator, Larvae. The time horizon used in the economic evaluations for MTGs was also extracted. When the models developed by the sponsor and EAC used different time horizons (as was the case in MTGs 3, 6, 10, 28), a common time horizon used in both the sponsor and the EAC economic models was extracted. The changes made by the EAC in the sponsor's

submission were assessed, focusing on the following areas: Clinical Pathway, Model clinical inputs, Costs, and Sensitivity Analyses. When the EAC made no changes to the submitted models, the sponsor's values for cost-savings were extracted. Also, the NHS economic impact of adopting the technology, the value of the presented annual cost saving in GBP for the NHS resulting from adopting the technology was extracted.

1.3.3 Qualitative Evaluation

An in-depth qualitative evaluation of the not-supported guidelines was performed. The main criticisms made by the EAC on the sponsor's presented clinical and economic evidence were summarised. This was followed by an independent appraisal of each not-supported MTG, based on the qualitative and quantitative information together. The aim of this secondary evaluation was to identify what were the main drivers underlying the negative outcome of the decision process.

1.3.4 Statistical Analysis

A Kruskall-Wallis test was performed to analyse if there were statistically significant differences in the average number in each type of clinical evidence and the number of participants per RCT between the 3 categories of MTG recommendations.

1.4 RESULTS

1.4.1 MTG characteristics

31 MTGs and their associated reports were assessed, published between March 2011 and February 2017. A summary of the general characteristics of the MTGs is presented in *Table 1*. Regarding the committee's recommendation on adoption, 14 technologies were fully supported (45%), 11 were partially supported (36%) and 6 were not supported (19%) *(Figure 2).* On only 1 MTG (MTG11), the committee's suggestion was unclear, and thus, was categorised as having partial support. The majority of the reasons behind the partial support were restrictions for a certain patient sub-population. The majority of technologies were targeted for tertiary care, i.e. hospital setting (77%).

6 MTGs (19%) contained research recommendations (MTG 1, 5, 20, 21, 22 and 31). 5 (16%) led to development of research protocols by External Assessment Centres (MTG5, 20, 21, 22 and 31). Recommended research on MTG5 was completed and led to a publication [19], which was considered by NICE in 2016. NICE decided not to update MTG5 based on this new information. At the end of our study period, recommended research studies for MTG20, 21, and 22 were underway. Recommended research for MTG31 was in design stage.

1.4.2 Clinical evidence appraisal

A summary of the clinical evidence for each guidance is presented in *Table 2*. The evidence submitted by the sponsor consisted of RCTs, non-experimental studies such as case-control studies, cross sectional studies, technical reports and technology registries. Abstracts and unpublished studies submitted by the sponsor were also reviewed. After its evaluation, the EAC added an average of 2 studies per MTG (Range: 0 – 14) and rejected an average of 9 studies per MTG for further evaluation (Range: 0 - 190). Overall, the EAC added 59 and rejected 278 studies. Several reasons were provided for rejecting studies: population was outside of the scope (MTG29 and 30), studies had overlapping cohorts (MTG27); there was double reporting of studies (MTG14, 17 and 16); studies did not assess the device being evaluated (MTG3); studies did not report outcomes or interventions defined in the scope (MTG12 and 22); small studies with limited follow-up (MTG16 and 26); and that studies did not address the decision problem (MTG21 and 13).

Regarding the quality of the evidence, the majority of the MTGs did not present any RCTs for the committee's evaluation (58%). The proportion of MTGs with one or more RCTs was non-significantly lower for the fully-supported MTGs compared to those that were either not supported or partially supported *(Figure 3).* Also, the number of participants per RCT was not higher among fully-supported technologies. On average, 29% of the fully-supported MTGs had 1 or more RCTs, less than the partially-supported (55%) and the not-supported (50%) MTGs. The fully-supported, partially-supported and not-supported MTGs did not have statistically significant differences regarding the number of RCTs ($p= 0.7164$), number of participants per RCT ($p= 0.227$), non-experimental ($p= 0.192$) or unpublished or abstract studies per guideline ($p= 0.536$). The fully supported MTGs had more than twice as many non-experimental studies as the not-supported MTGs *(Figure 4).*

1.4.3 Economic evaluation appraisal

The economic evaluations submitted by the sponsors utilised different modeling approaches *(Table 3, extended version in Supplement 2)*, most common being the decision tree, 20 (63%) followed by the Markov model, 5 (16%). The information regarding the model type was not available in 6 MTGs. Minimum and maximum cost-saving (or cost-incurring) values per patient calculated by the sponsor and the EAC are presented in the *Supplement 3.* The changes made by the EAC to the economic evaluation presented by the sponsor were assessed in 4 areas: clinical pathway, clinical and demographical inputs, costs and sensitivity analysis. In 15 MTGs the EAC suggested changes to the clinical pathway, in 22 MTGs changes were proposed to the model clinical inputs, in 23 MTGs changes were proposed to the costs in the model and in 20 to the sensitivity analyses conducted in the economic evaluation. The minimum and maximum cost-saving estimated by the sponsor and the EAC were extracted from the

MTGs. A summary of the difference between the average values estimated by the sponsor and the EAC is presented in *Figure 5*. Despite the large variation among MTGs, the sponsor estimated a 57.9% higher average cost saving than the EAC, with a higher value in 20 cases (65%).

Regarding the 6 not-supported MTGs, 2 were estimated to be cost-incurring on average (MTG 6 and 22). In one case (MTG 21) the EAC deemed it impossible to conduct an economic evaluation due to a lack of evidence to populate the economic model. In the 3 remaining cases (MTGs 5, 20 and 31), the EAC economic evaluation showed cost savings, which on average were between £8.5 and £2,244 per patient. In contrast, almost all of the fully-supported MTGs had estimated cost savings in their economic evaluations. The only exception was MTG 15, which had a slightly cost-incurring average value (-£7.1) per patient. Nevertheless, the experts included in the MTG 15 evaluation agreed that the cost-incurring scenario was unlikely, and that in the most probable clinical scenarios the technology was cost-saving.

Only 5 MTGs included an analysis of the NHS overall cost saving incurred by adopting the technology (MTG 25, 28, 29, 30, and 32) into the healthcare system. Of those, 2 had full support and 3 partial support. The calculated potential annual cost-saving to the NHS ranged between £1,480,000 and £129,000,000.

1.4.4 Key drivers of negative decisions

a. Committee criticisms of the clinical evidence presented in the not-supported MTGs

The committee's evaluation of the clinical evidence was qualitatively assessed *(Supplement 4).* Overall, low quality or quantity of the clinical evidence was a dominant issue, specifically raised in all of the not-supported MTGs. Some of the specific reasons raised were small numbers of patients, lack of appropriate comparators, lack of evidence on long-term clinical outcomes, and lack of effect on the selected outcomes, low methodological quality and high risk of bias. For instance, in MTG20, the committee raised concerns about uncontrolled confounding in one of the studies. Another oft-cited issue raised by the committee in 4 of the not-supported MTGs was the lack of comparative evidence. Less frequently, the committee commented on the applicability of the presented evidence from outside of the United Kingdom (e.g., MTG 1 and 21).

3 of the not-supported MTGs had RCTs. Despite the availability of RCTs in MTG 31 and 21, the RCTs did not evaluate the outcomes deemed as important by the committee. In MTG 5, the committee judged that the evidence did not consider a sufficient number of patients, ultimately lacking the power to demonstrate a statistically significant difference between the intervention and control groups.

b. Committee criticisms of the economic evaluation presented in the not-supported MTGs

Generally, most MTGs referred to the uncertainties in the presented cost models and the need for further research *(Supplement 5)*. The majority of the not-supported MTGs were subject to committee criticism regarding model clinical inputs. In MTG 5, the committee reported problems with the cost calculations, as well as the level of uncertainty associated with model inputs. In MTG 21, due to the lack of clinical evidence, some relevant scenarios were deemed as not feasible to assess. In the remaining 3 MTGs (6, 20 and 22), the models did not consider all the relevant costs and resources involved (i.e., in MTG 20 the model included limited information on treatment costs and resource implications of having the disease; in MTG 22 the model did not capture potential savings from differences in diagnostic accuracy between current practice and the medical technology), which affected the results. Erroneous assumptions were criticized in some MTGs. For instance, in MTG 5 the clinical evidence was judged not to support the assumptions in the economic model, and in MTG 21 there was uncertainty about the validity of the assumptions. Also, the clinical pathway was criticized in some models. For example, in MTG 5 some important comparative alternatives were omitted. In other MTGs, the treatment pathway used in the model omitted stages that were considered relevant by the clinical experts and therefore, the model was not generalizable to the UK clinical practice. Other criticisms focused on the lack of clarity of the models submitted by the sponsor (MTG 6). In some cases, such as MTG 6, there was wide variability in results due to parameter uncertainty (i.e. the difference between current practice and the medical device

ranged from a cost saving of £195 to a cost increase of £536) and hence, it was not possible to draw certain conclusions about estimated cost savings associated with the technology.

c. Main driver of decision-making

Following an assessment of the most common criticisms of the not-supported MTGs, each MTG was evaluated individually by considering both the qualitative and quantitative evidence.

In the case of Parafricta Boots (MTG 20), despite an EAC-estimated cost-saving ranging from £595 to £2799 per at-risk person, the committee decided that there was insufficient evidence to support the claimed clinical effectiveness of the technology. The available evidence included 2 case series, 2 unpublished studies and 1 non-randomised trial. Regarding the latter, it was not clear if the results were biased by unassessed confounding factors.

Mist (MTG 5) was also considered to be cost-saving according to its economic evaluation results. However, the quality of the published studies was deemed as insufficient to support the adoption of the technology, as the committee remained uncertain about clinical effectiveness. Although the available evidence included 2 RCTs and 8 non-experimental studies, the negative decision was due to the low number of participants and no comparators being included in the studies. Also, for chronic wounds there were no studies showing long-term outcomes.

In the case of Ambulight PDT (MTG 6), the negative decision was primarily based on the lack of sufficient clinical evidence as well as the economic evaluation model that did not provide certain cost-saving results. No RCTs were presented in support of the technology, and the evidence was based only on 3 sources, 2 of which were unpublished data.

In Humigard (MTG 31), even though the clinical evidence included 6 RCTs with 541 patients in total, evidence did not conclusively demonstrate an effect on reducing adverse outcomes (incidence of surgical site infections, length of stay in post-operative recovery and total hospital stay). A review of the RCTs suggested that most reported effectiveness in terms of pain reduction, with only 2 addressing other outcomes (core and wound temperature).

In the case of Re-cell (MTG 21), although the available clinical evidence included 3 RCTs and 13 unpublished sources, there was only limited information on the primary outcomes of interest, which subsequently impeded the development of a cost model by the EAC.

Finally, no RCTs were presented for Vibratip (MTG 22) and the committee found the studies to be of low methodological quality. Therefore, there was a high level of uncertainty surrounding the diagnostic accuracy of the technology.

1.5 DISCUSSION

1.5.1 Summary of main findings

The review of the MTGs showed that the decision outcome in almost half of the 31 published MTGs to date fully supported the technology while the remaining MTGs recommended partial or no support. The majority of the fully-supported technologies were recommended in a tertiary care setting. Out of the 31 MTGs, about one fifth led to research recommendations.

1.5.2 Clinical Evidence

Interestingly, a lower proportion of fully-supported MTGs had 1 or more RCTs, as compared to that in partially-supported and not-supported MTGs. According to our statistical analysis, there was no difference in the number of randomized and non-randomized studies between the fully-supported, partially-supported and not-supported medical technologies. Nevertheless, this could be influenced by the low number of units. Although the design of a study provides only partial information about its rigor and

relevance to a given decision problem, randomized controlled trials are widely accepted as the gold standard for assessing the effectiveness of health care interventions.

The majority of the completed MTGs (58%), irrespective of their outcome, did not include RCTs for the committee's evaluation. However, this is not entirely surprising: unlike other programmes, the methods guide for MTEP does not state that RCT is the preferred source of evidence. This flexibility of not having RCTs as the preferred option may be warranted considering the fact that the majority of medical devices are developed by small and medium enterprises, as well as the complexity of conducting robust RCTs for medical devices. Drummond et al. (2009) previously pointed out the "additional challenges" associated with the generation of evidence on medical technologies. These are related to ongoing product modifications, which may have an impact on efficacy, the associated "learning curve" regarding the uptake of new devices, and the common unfeasibility of blinding, which may increase the risk of bias. Previous research also suggested that the efficacy of a medical device depends not only on the device but on the users characteristics, which might act as key confounders [20]. A concerted effort involving academic, industry, and regulatory experts is needed to evaluate the feasibility of designing and conducting RCTs of medical technologies. For example, Schnell-Inderst et al. have recently recommended wider use of adaptive trial designs to overcome the complexities of generating rigorous evidence on medical technologies [21]. In the absence of RCTs, the quality of the evidence supporting the use of many medical devices may be associated with increased uncertainty. Also, given the low number of MTGs that had RCTs, statistical methods to systematically combine treatment effects from randomised and not randomised studies, such as Bayesian methods, could be considered, utilising bias-adjusted evidence synthesis [22,21].

Overall, there was a high degree of consistency in the type and detail of information reported in the MTGs. However, in some MTGs (e.g., MTGs 2 and 6) it was unclear which study was submitted by the sponsor and which was added or rejected by the EAC. Furthermore, although MTGs provide a summary of the supporting studies, in some cases it was not possible to identify the type of the study, and determine whether it was a poster abstract or a peer-reviewed paper. In the future, MTGs should aim to have more homogeneity in their reporting, in order to increase comparability and facilitate the public evaluation of the process. However, it is also appreciated that MTG production is an evolving process in the changing landscape of the NHS and NICE.

During our study period, almost one fifth of completed MTGs had research recommendations. While no MTGs were updated with new evidence generated from the Research Commissioning workstream, this was because the majority of studies were underway. The extent to which completed studies inform future reviews of MTGs should be evaluated in the long term.

1.5.3 Economic Evaluation

The assessment of supported and not-supported medical devices revealed differences regarding results of the economic evaluation. The majority of MTGs with full support had results that suggested the technology was cost-saving, but only 3 out of the 6 not-supported MTGs reported cost-saving results per patient. The analysis also revealed the impact of the EAC assessment on the evaluation and the results of economic evaluation. The EAC often made suggestions or changes in several components of the economic model, which materially changed the results in 26 of the reviewed MTGs. Of these 26, in 20 of the MTGs the sponsor estimated a higher cost-saving value than what was proposed by the EAC. This is expected, as the manufacturer's evaluation tend to be higher than external review groups. In the majority of these 20 MTGs, the EAC made changes regarding the inputs, either clinical (16) or costs (17). For example, in MTG 2 the EAC stated that the hourly cost used by the sponsor was too high, in MTG 3 the sponsor used inaccurate cost data, in MTG 11 the sponsor did not include the VAT and in MTG 19 the EAC found an error in the hourly nursing costs. In 14 of those MTGs the EAC did further sensitivity analysis, and in 10 MTGs suggested changes to the clinical pathway.

The reported cost-saving values in guidance documents were not always comparable between MTGs as they were expressed in different metrics. For example, in MTGs 7 and 14, the cost-saving was

calculated “per operation theatre”, while most of the other MTGs calculated the cost-saving per patient. The feasibility of adopting a common metric across all MTGs should be explored, as this would facilitate more meaningful comparisons on the cost impact of these technologies on the NHS.

The MTEP methods guide states the following as one of the objectives of the programme: “to evaluate the impact of the technology on the healthcare system, alongside its clinical benefits for individual patients” [10]. Accordingly, all MTGs had a section regarding the NHS considerations, focusing on the healthcare system impact. However, information on NHS economic adoption impact was absent in 26 MTGs. Although the methods guide does not explicitly state the adoption impact to be a requirement for the MTGs, this parameter would be valuable for policymakers in the future when evaluating the economic implications of these technologies. The 5 MTGs that did incorporate this information were MTG 25, 28, 29, 30 and 32.

1.5.4 What is the main driver of decision-making?

Among the not-supported medical technologies, the decision was driven primarily by the quantity or quality of the presented evidence. Although some of the not-supported technologies had RCTs, either their outcomes were not clinically meaningful or had an insufficient number of patients. Some of the non-experimental supporting studies had potential sources of bias or were unpublished. In 3 of the not-supported MTGs the technologies were not deemed as cost-saving, or the economic evaluation could not be performed due to the lack of clinical evidence. 3 of the not-supported medical technologies were cost-saving, in which case the lack of certainty about their effectiveness was the main driver for the negative recommendation.

1.5.5 Further comments

Although there is an initiative of involving patients and public in the MTEP programme through the Patient and Public Involvement Programme PPIP [5], the guidelines do not explicitly detail the involvement of patients in the decision-making process. Through the process, people and organisations may register their interest in a technology, and they may be invited to attend the committee meetings [5]. Also, they may participate and provide input in the consultation process of the draft recommendation. Nevertheless, the guidelines do not explicitly show the extent of patient contribution in the final MTG.

Although technologies selected for MTG are the ones deemed to be cost-saving or cost-neutral, the Methods guide [10] does not clarify what would happen if they are found to be cost -incurring during the evaluation. Even if cost-incurring, they could still represent a beneficial and cost-effective use of NHS resources, and so rejecting them on that basis may not be appropriate. Also, by adopting a cost-consequence approach, the committee is not in a position to evaluate value-for-money.

1.5.6 Strengths and weaknesses of the review

The review has several strengths. This is the first review to critically assess all the published MTGs stemming from the MTEP. Adopting a systematic approach, an overview of both the clinical and economic evidence underpinning the decision process in the MTGs was provided. It also incorporates a qualitative appraisal of the not-supported MTGs, identifying factors that could explain the key drivers of the committee decisions.

There were a number of limitations to this review: the guidelines were reviewed by 1 investigator and then discussed with senior members of the team. Also, a small degree of inconsistency in the reporting of MTGs complicated the appraisal and comparison of the guidance documents. In the economic evaluation analyses, an average result of the economic evaluation was calculated by extracting minimum and maximum values from the MTGs, but this is likely to have limited comparability between the sponsors and EAC values. The qualitative assessment was limited to not-supported MTGs. Conducting a qualitative evaluation for identifying the key drivers of the positive decisions was beyond the scope of this review and should be investigated in future evaluations.

1.5.7 Unanswered questions and future research

A number of MTGs stated that the evidence on the technology will be updated when it becomes available (e.g., MTG 22). NICE is currently updating 2 MTGs (20 and 21). Future investigations should monitor the evolution of the MTGs: as pointed out by Drummond and colleagues, medical technologies are routinely modified, with frequent changes in their clinical and economic implications [20]. Also, it would be interesting to evaluate further the supported and partially supported MTGs to look for other drivers in the decision making process that were not assessed in this review. Finally, future research should evaluate the extent to which medical technologies recommended by the MTEP are adopted in the NHS and their subsequent economic implications.

1.5.8 Conclusions

In conclusion, the MTEP programme has to date published 31 MTGs, supporting the adoption of most technologies, either fully or partially. Most of the MTGs did not have RCTs in the clinical evidence presented for the committees' evaluation, regardless of the recommendation status, which may reflect the feasibility challenges of conducting RCTs for certain medical technologies. According to our review, the EAC played a significant role in the MTEP process. Overall, the EACs rejected 278 and added 59 studies to the sponsors' submissions. Also, it suggested changes to the sponsors' economic model, which had an impact on the final values in 84% of the cases (26/31). The main drivers of the decision in the not-supported MTGs was the lack of quantity or quality of the evidence, as well as economic models with cost-incurring results in half of the cases (3/6).

Data Availability Statement

The extended versions of the evaluated MTGs are available as supplementary material. No additional data are available.

Author Contributions

HN, EB, LO, and EM conceived the study. FC performed the document review, completed all data extraction, and developed the first full draft of the paper. All authors critically reviewed and contributed to subsequent drafts. All authors have read and approved the final version of the paper for submission.

1.6 TABLES *(IN SEPARATE DOCUMENT)*

1. Table 1. Summary of the MTGs
2. Table 2. Clinical Evidence Appraisal
3. Table 3. Economic Evaluation Appraisal

1.7 FIGURES *(IN SEPARATE DOCUMENT)*

1. Figure 1. MTEP pathway
2. Figure 2. Committee's recommendation on Adoption
3. Figure 3. RCTs evaluated by the committee according to Committee's recommendation on adoption
4. Figure 4. Clinical evidence evaluated by the committee according to Committee's recommendation on adoption.
5. Figure 5. Difference in the estimated cost savings by the sponsor and EAC (GBP).

1.8 SUPPLEMENT *(IN SEPARATE DOCUMENT)*

1. Supplement 1. Parameters description
2. Supplement 2 Extended Economic Evaluation
3. Supplement 3. Calculations
4. Supplement 4. Qualitative evaluation: Clinical evidence EAC key criticisms from not supported MTGs
5. Supplement 5. Qualitative evaluation: Economic evaluation EAC key criticisms from not supported MTGs

1.9 REFERENCES

1. WHO: Health technology assessment of medical devices: WHO Medical device technical series. In: World Health Organization (ed.). (2011)

2. HTA Glossary: International Network of Agencies for Health Technology Assessment and Health Technology Assessment international. http://www.htaglossary.net.

3. Raftery, J., Powell, J.: Health Technology Assessment in the UK. The Lancet **382**(9900), 1278-1285 (2013). doi:10.1016/S0140-6736(13)61724-9

4. Chapman, A.M., Taylor, C.A., Girling, A.J.: Are the UK Systems of Innovation and Evaluation of Medical Devices Compatible? The Role of NICE's Medical Technologies Evaluation Programme (MTEP). Applied Health Economics and Health Policy **12**(4), 347-357 (2014). doi:10.1007/s40258-014-0104-y

5. NICE: Medical Technologies Evaluation Programme: Process guide In. London, (2011)

6. NICE: Guide to the processes of technology appraisal In: National Institute for Health Care Excellence (ed.). (2014)

7. NICE: Diagnostics Assessment Programme manual. In: National Institute for Health Care Excellence (ed.). (2011)

8. NICE: Developing NICE guidelines: the manual. In: National Institute for Health Care Excellence (ed.). (2014)

9. NICE: Interventional procedures programme manual. In: National Institute for Health Care Excellence (ed.). (2016)

10. NICE: Medical Technologies Evaluation Programme: Methods guide In: National Institute for Health Care Excellence (ed.). London, (2011)

11. Alshreef, A., Jenks, M., Green, W., Dixon, S.: Review of Economic Submissions to NICE Medical Technologies Evaluation Programme. Appl Health Econ Health Policy **14**(6), 623-634 (2016). doi:10.1007/s40258-016-0262-1

12. Ciani, O., Federici, C., Tarricone, R., Bezerianos, A.: Current and Future Trends in the HTA of Medical Devices. In book: XIV Mediterranean Conference on Medical and Biological Engineering and Computing (2016). doi:10.1007/978-3-319-32703-7_257

13. Rothery, C., Claxton, K., Palmer, S., Epstein, D., Tarricone, R., Sculpher, M.: Characterising Uncertainty in the Assessment of Medical Devices and Determining Future Research Needs. Health Economics **26**(S1), 109-123 (2017). doi:10.1002/hec.3467

14. NICE: Research Recommendations Process and methods guide In.National Institute for Health Care Excellence (ed.).United Kingdom, (2011)

15. Pomfrett, C., Campbell, B., Pugh, P., Campbell, M., Marlow, M.: Medical technology evaluation II: catalysing the development of primary clinical evidence for promising technologies. Paper presented at the 9th HTAi Annual Meeting, Bilbao (2012).

16. Cowles, E., Marsden, G., Cole, A., Devlin, N.: A Review of NICE Methods and Processes Across Health Technology Assessment Programmes: Why the Differences and What is the Impact? Appl Health Econ Health Policy **15**(4), 469-477 (2017). doi:10.1007/s40258-017-0309-y

17. Green, W., Hutton, J.: Health technology assessments in England: an analysis of the NICE Medical Technologies Evaluation Programme. Eur J Health Econ **15**(5), 449-452 (2014). doi:10.1007/s10198-013-0539-3

18. Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A.H., Mauskopf, J., Loder, E.: Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Bmj **346**, f1049 (2013). doi:10.1136/bmj.f1049

19. White, J., Ivins, N., Wilkes, A., Carolan-Rees, G., Harding, K.G.: Non-contact low-frequency ultrasound therapy compared with UK standard of care for venous leg ulcers: a single-centre, assessor-blinded, randomised controlled trial. Int Wound J **13**(5), 833-842 (2016). doi:10.1111/iwj.12389

20. Drummond, M., Griffin, A., Tarricone, R.: Economic evaluation for devices and drugs--same or different? Value Health **12**(4), 402-404 (2009). doi:10.1111/j.1524-4733.2008.00476_1.x

21. Schnell-Inderst, P., Hunger, T., Conrads-Frank, A., Arvandi, M., Siebert, U.: Recommendations for primary studies evaluating therapeutic medical devices were identified and systematically reported through reviewing existing guidance. J Clin Epidemiol **94**, 46-58 (2018). doi:10.1016/j.jclinepi.2017.10.007

22. Schnell-Inderst, P., Iglesias, C.P., Arvandi, M., Ciani, O., Matteucci Gothe, R., Peters, J., Blom, A.W., Taylor, R.S., Siebert, U.: A bias-adjusted evidence synthesis of RCT and observational data: the case of total hip replacement. Health Econ **26** Suppl 1, 46-69 (2017). doi:10.1002/hec.3474