Improvement measures based on the actual operation of health technology assessment: focusing on the evaluation period and conformity with guidelines

Shunsuke Takada
DP20403

Graduate School of Pharmaceutical Sciences
Department of Clinical Medicine (Pharmaceutical Medicine)
Kitasato University

5-9-1 Shirokane, Minato-ku, Tokyo, 108-8641, Japan

Abstract

Many developed countries are facing increasing medical expenditures because of the growing elderly population and sophistication of medical technology. Under these circumstances, health technology assessment (HTA) has been introduced in these countries to pursue efficient medical care within limited financial resources. In Japan, a cost-effectiveness evaluation system has been in operation since 2019 for adjusting drug price after being marketed. During the short history of the cost-effectiveness evaluation system in Japan, several concerns have been highlighted. This study analyzed and evaluated HTA in the United Kingdom (UK), or appraisals, which has an extensive record, to obtain suggestions for addressing the concerns in Japan's cost-effectiveness evaluation system.

One of the concerns is regarding the evaluation timeline. The timeline for the evaluation is fixed, regardless of the characteristics of the targeted medicines, and Japan's cost-effectiveness evaluation system does not necessarily assume the existence of medicines that require more time for cost-effectiveness analysis or discussions on their appropriateness. Therefore, we evaluated the characteristics of the medicines that affected the time required for appraisal, using publicly available documents from National Institute for Health and Care Excellence (NICE) —the HTA agency in the UK—the European Medicines Agency, and the European Commission.

The results indicated that the appraisals for orphan medicinal products (OMPs) extended the time between regulatory approval and appraisal completion, which may result in limited access to them. This is because they take more time for discussion at NICE after manufacturers' evidence submission. The appraisals for anti-cancer medicines also required more time for discussion; however, the preparation in advance of the appraisal, such as determining the analysis framework, did not result in access restrictions. In the case of OMPs' appraisals, considering similar preparation in advance would be useful for timely and appropriate decision-making.

Another concern is regarding conformity to analytical guidelines of the cost-effectiveness analysis. With regard to the utility values, Japan's analytical guideline prioritizes values measured by specific methods for appropriate populations. However, it is ambiguous whether this condition should be applied in all cases. Therefore, using NICE's publicly available documents for appraisals of anti-cancer medicines, we ascertained whether the utility values in manufacturers' evidence submissions conformed to the guideline. Thereafter, we evaluated whether their conformity would bias acceptance of manufacturer-proposed utility values by NICE. If they were not accepted, we also examined

the reasons.

The results indicated that manufacturers' conformity to analytical guidelines for utility values was not related to acceptance of manufacturer-proposed utility values by NICE. However, the reasons for non-acceptance differed depending on conformity to guidelines. When manufacturers conformed to them, the main reasons for non-acceptance were discrepancies from the actual medical practice in the UK and inappropriate data adjustment. In the case of non-conformity, the main reason was the quality of the information sources. This suggests that conformity to guidelines may have inherent concerns. When manufacturers evaluate quality-of-life in clinical trials with the intention of utilizing them in cost-effectiveness analysis, the quantity and quality of the data should be confirmed in advance and their appropriateness should be discussed before initiation of trials.

This study address two unique concerns in Japan's cost-effectiveness evaluation system; however, the strategies for overcoming them are identical. Manufacturers and HTA bodies, including the HTA agency, should mutually collaborate and prepare for cost-effectiveness evaluation from the clinical development stage of each medicine to make timely and high-quality decisions.

Table of Contents

List of Tables

List of Figures

Abbreviations

AA	accelerated assessment
B	unstandardized partial regression coefficient
C2H	Center for Outcomes Research and Economic Evaluation for Health
CDF	Cancer Drugs Fund
CI	95 percent confidence interval
CSIMC	Central Social Insurance Medical Council
Di	Cook's distance
EC	European Commission
EORTC	European Organization for Research and Treatment of Cancer
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
EQ-5D	EuroQol 5 Dimensions
ERG	evidence review group
FACT	Functional Assessment of Cancer Therapy
FACT-G	Functional Assessment of Cancer Therapy General
FAD	final appraisal determination
FS	final scope
G-BA	Gemeinsamer Bundesausschuss
GDP	gross domestic product
HAS	Haute Autorité de Santé
HRQOL	health-related quality-of-life
HTA	health technology assessment
ICER	incremental cost-effectiveness ratio
IMF	Innovative Medicines Fund
INAHTA	International Network of Medical Technical Assessment Institutions
IQWIG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
MA	marketing authorization
MTA	multiple technology assessment
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OECD	Organisation for Economic Co-operation and Development
OMP	orphan medicinal product
QALY	quality-adjusted life-year
RQ	research question
SF-36	MOS 36-item short-form health survey
SMC	Scottish Medicines Consortium
STA	single technology assessment
TA	technology appraisal
TAC	technology appraisal committee
TTO	time trade-off
UK	United Kingdom
VPAS	voluntary scheme for branded medicines pricing and access

1. Introduction

1.1. Rising Medical Expenditures

Medical expenditures have been increasing in countries worldwide. Among the Organisation for Economic Co-operation and Development (OECD) member states, the annual total medical expenditures per capita, adjusting for purchasing power parity, averaged $3,061 in 2011 and $4,086 in 2019—an increase of $1,026 over nine years. In Japan, the annual total medical expenditure per capita was $3,740 in 2011 and $4,691 in 2019—an increase of approximately $950 during the same period. Thus, in the OECD member states, total medical expenditures per capita are on the rise; there is a similar trend in Japan [1].

The ratio of total medical expenditures to gross domestic product (GDP) as an indicator of the impact of medical expenditures on financial conditions has also been increasing. Among the OECD member states, total medical expenditures as a proportion to GDP averaged 8.6% in 2011 and 8.8% in 2019—an increase of 0.2 points over nine years. In Japan, the proportion of total medical expenditures to GDP was 10.6% in 2011 and 11.0% in 2019—an increase of 0.4 points during the same period. Thus, among the OECD member states, the financial burden of rising medical expenditures has been mounting; there is a similar trend in Japan [1].

Two factors are implicated as causes of rising medical expenditures in Japan. The first is an increase in the proportion of the elderly population. As summarized in Figure 1-1, the total population in Japan peaked at 128,080,000 in 2008, and thereafter, began to decline. As of April 1, 2022, the total population of Japan was 125,190,000—a slight decrease compared to the peak. However, the population ratio by age group changed during this period. As of April 1, 2008, the population aged 65 or older accounted for 21.8% of the overall population of Japan; it accounted for 29.0% as of April 1, 2022—an increase of 7.2 points over 14 years [2].

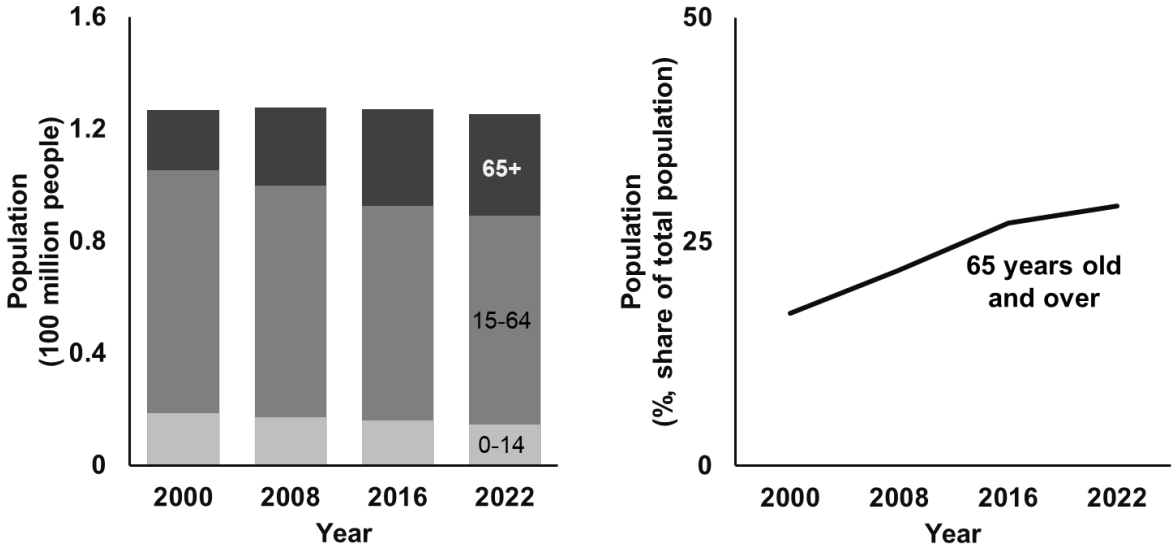

Figure 1-1 Trends in Japan's Total Population and the Proportion of the Population Aged 65 or Older [2]

The annual total medical expenditures per capita in Japan differs by age group. According to the 2019 government statistics, the annual medical expenditures were smallest among those aged 20–24 years—approximately 86,000 yen. In contrast, among the 65–69 years age group, the expenses were approximately 502,000 yen—5.8-times higher compared to the 20–24 years age group. Furthermore, for those in the 75–79 years category, the annual medical expenditures were approximately 790,000 yen—415,000 yen higher than the expenses for the 20–24 years old. An increase in the proportion of people aged 65 years or older, who require more medical services under conditions where the total population is largely constant, is a significant factor behind the surge in medical expenditures [3].

The second factor is innovation in medical technology. Figure 1-2 presents factor analyses of the annual growth in medical expenditures. One factor behind the increase in medical expenditures was the increase in the proportion of the elderly population; however, that factor alone accounts for only approximately half of the growth. Between 2011 and 2015, medical expenditures increased by approximately 2.5% on average, compared to the previous year. Of the increase, 1.4% is explained by factors other than the increase in the proportion of elderly population, increase or decrease in population, and revision of medical fees. Examining the “Other in all instances” factor, hospital inpatient costs accounted for 0.2%, hospital outpatient costs accounted for 0.4%, pharmacy costs were 0.7%, and dentistry costs accounted for 0.1%. Pharmacy costs—approximately 20% of medical expenditures—make the largest contribution to the “Other in all instances” category [3]. Moreover, breaking down the pharmacy costs factor, fees

for medical service professionals accounted for 0.1% and medicine costs were 0.6%, with medicine costs, primarily, raising medical expenditures [4].

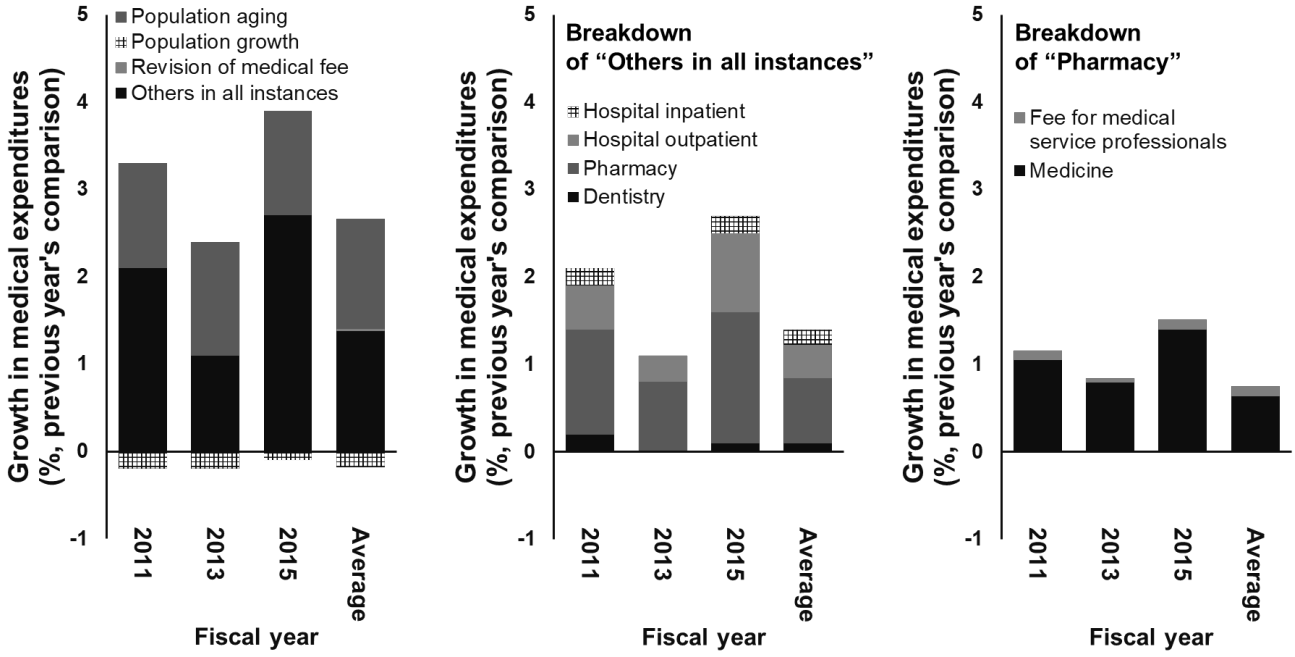

Figure 1-2 Factor Analysis of Growth in Medical Expenditures [4]

In the background of increase in medicine costs is the application of medicines utilizing advanced science and technology. The humanized anti-human IL-6 receptor monoclonal antibody tocilizumab: the first antibody drug to be created in Japan, was put into practical use in 2005 [5]. With the development of science and technology ever since, not only monoclonal antibody medicines, but other modalities such as antibody drug conjugates, nucleic acids, and therapeutic cells have also been put into practical use as medicines. Medicines utilizing these innovations are often priced higher than conventional medications [6], and when new medicines are used by many patients, the resulting financial impact becomes a concern.

Nivolumab, which is an anti-programmed death receptor-1 monoclonal antibody, was listed in Japan's Drug Price Standard in 2014 for the indication of malignant melanoma. Considering the limited number of patients utilizing that medicine, the drug price of the 100 mg/10 mL strength was calculated to be 729,849 yen using the cost accounting method [7]. Subsequently, other indications including non-small cell lung cancer, renal cell cancer, and Hodgkin lymphoma were added to this medicine, and the number of patients eligible for nivolumab increased. Consequently,

the increased health expenditures associated with nivolumab have become a financial burden. To address this issue, a new mechanism was introduced to flexibly adjust prices of medicines that have experienced market expansion above a certain scale [8]. The factors responsible for the surge in medical expenditures include the increase in the practical use of expensive medicines utilizing innovative technology and rise in the number of patients treated because of the expansion of indications.

Given the background of the rising proportion of the elderly population and advances in medical technology, it is likely that rising healthcare expenditures is not a temporary phenomenon. These background conditions are found not only in Japan, but in other developed countries as well, and are recognized as risks to the sustainability of healthcare [9, 10]. Health technical assessment (HTA) has been introduced worldwide to manage changes in medical expenditures, while promoting advances in science and technologies. HTA has been proposed as a means to continue to deliver more efficient medicines using limited resources.

1.2. Implementation of Health Technology Assessment in Foreign Countries

The International Network of Medical Technical Assessment Institutions (INAHTA), in which public agencies from various countries participate, defines HTA as "a multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its lifecycle. The purpose is to inform decision-making in order to promote an equitable, efficient, and high-quality health system" [11]. This definition aims to evaluate medicines from a wide range of perspectives and is a broader concept than clinical trials by manufacturers to confirm the efficacy and safety of medicines and regulatory review. However, there is a narrower interpretation of HTA focusing on its economic impact [12]. Currently, many countries have established specialized agencies for HTA—the HTA agencies—that evaluate cost-effectiveness based on the effectiveness obtained from newly introduced medicines and the associated required costs. Based on the result of the cost-effectiveness analysis, the agencies ascertain the propriety of insurance reimbursement or set the price of medicines.

The National Institute for Health and Care Excellence (NICE), established in the United Kingdom (UK) in 1999, is a representative of the HTA agencies. NICE assesses the cost-effectiveness of certain medicinal products determined to have a significant financial impact, and makes recommendation for their use under the National Health

Service (NHS)—a process known as appraisal. In the appraisal process, a manufacturer first assesses the cost-effectiveness of a medicine in accordance with the analysis guidelines. Next, a technology appraisal committee (TAC)—an internal NICE organization—reviews the medicine after critical examination by the evidence review group (ERG), consisting primarily of experts from universities. NICE makes recommendations regarding reimbursement, considering factors other than cost-effectiveness, including end-of-life care. In principle, there are three types of NICE judgements: recommend for use under NHS; not recommend for use; or recommend for a smaller group of patients than originally stated by the marketing authorization (MA), described as optimized [13].

HTA in NICE is associated with limited access to medicines. Therefore, the system is constantly amended based on feedback from stakeholders, such as patient groups. For example, the Cancer Drugs Fund (CDF) was established under NHS England in 2011 to address issues related to limited access to expensive anti-cancer medicines [14]. CDF targets anti-cancer medicines that are not recommended for use under NHS. Beginning in 2016, a new system was established to ensure the sustainability of the fund, clarifying CDF budgets and benefits. Under the new system, NICE can make new decisions to recommend medicine use within the CDF budget [15]. In November 2021, the idea of an Innovative Medicines Fund (IMF), which expanded the concept of CDF beyond anti-cancer medicines, was suggested; its implementation is underway [16].

As of April 2022, a total of 1,032 medicines were covered by appraisals, of which 859 (83%) were recommended for some form of use. As shown in Figure 1-3, among 392 anti-cancer medicines covered by appraisals, 191 were recommended for use under NHS, 61 were recommended for limited patient populations, and 51 were recommended for use within CDF. In total, 303 anti-cancer medicines (77%) were recommended for some form of use [17]. A new pricing agreement between the Association of the British Pharmaceutical Industry and the Department of Health and Human Services of the UK was reached in January 2019, and all new medicines were eligible for appraisal after April 2020 [18]. It is expected that the number of NICE appraisals will continue to increase in the future.

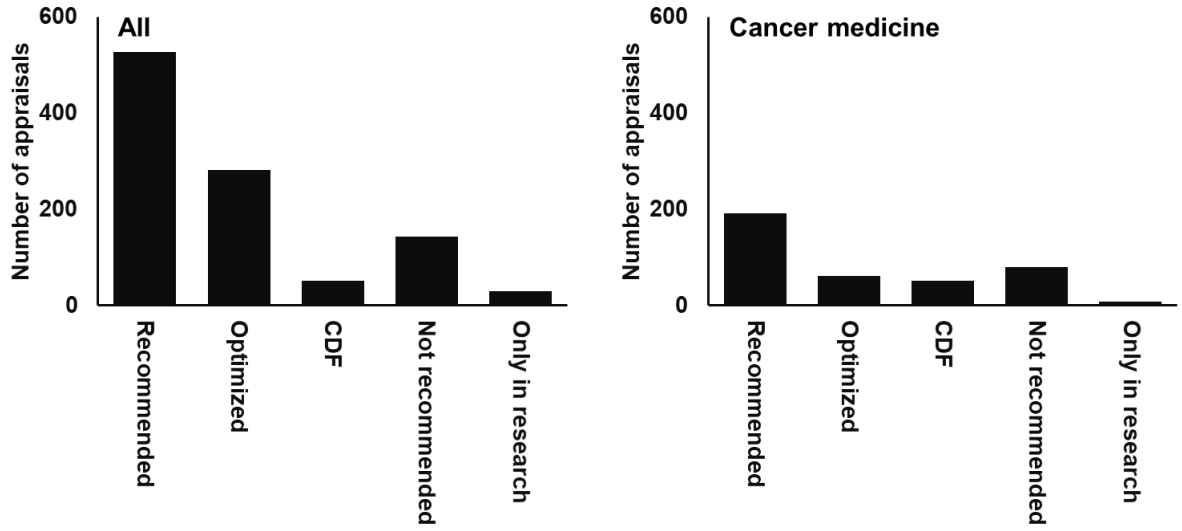

Figure 1-3 Results of Appraisals in NICE

The HTA concept also exists in countries other than the UK. In France, HTA is required when manufacturers negotiate medicine prices for public insurance. If manufacturers wish to set higher reimbursement prices, in certain cases they conduct cost-effectiveness analyses of the medicinal products. The HTA agency in France—the Haute Autorité de Santé (HAS)—reviews manufacturers' cost-effectiveness analyses and validates them by issuing one of three assessments: minor, important, or major issues found. Results of the cost-effectiveness analyses are not reflected in reimbursement prices, but manufactures are forced into stringent price negotiation when the HAS recognizes major issues, even if the medicinal product has obvious health benefits compared to the previous standard of care [19].

In Germany, the HTA agency—the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWIG)—functions in the context of the broader HTA definition. Medicines are reimbursed at the price set by the manufacturers for six months after their launch. In parallel, the Gemeinsamer Bundesausschuss (G-BA)—the decision-making body for the public insurance systems—assesses the additional health benefits of the medicines. Added benefits are assessed in terms of impact on life expectancies, symptoms, quality-of-life, and adverse events, and do not include an economic perspective. IQWIG assesses added health benefit based on the protocol developed by G-BA and reports the results. If there are no additional benefits, the price may be lowered to reference prices of similar medicines. However, if there are additional benefits, price negotiations are held between the insurer federation and the manufacturers. If no agreement is reached after the negotiations, arbitration led by a third-party is conducted; cost-

effectiveness will only be assessed when the arbitration ends in failure [20]. As of April 2022, there have been no instances requiring cost-effectiveness evaluation, that is, there have been no cases of the HTA in the narrow view.

The narrower view of HTA, determining whether or not to reimburse for medicines or to set their prices, is implemented in Sweden and other Nordic countries, in Eastern Europe, and in the Netherlands [21, 22]. Canada, Australia, and other non-European countries have similar mechanisms [23, 24]. The HTA is employed to provide medicines efficiently and continuously in countries worldwide.

1.3. Implementation of Health Technology Assessment in Japan and Its Challenges

In Japan, a cost-effectiveness evaluation system for certain medicines and medical devices was implemented in April 2019. In discussions regarding system design, three priorities were established: not hindering access to pharmaceuticals, considering financial impact, and positioning the new system as a supplement to the existing drug pricing system [25]. These aims are strongly reflected in the current system.

In HTA in other countries, the propriety of the reimbursement is often ascertained based on the results of cost-effectiveness evaluation. Restriction of access to medicines may occur if reimbursement is denied. To avoid this issue, under the HTA system in Japan, a portion of a medicine's price is calculated based on the standard price-setting system in Japan, and is adjusted based on the results of a cost-effectiveness evaluation after being marketed. HTA in the UK targets all new medicines, including medicines for expanded indications, whereas the Japanese system targets a few types of medicines—those with a significant financial impact. Notably, there is a shortage of personnel who can conduct cost-effectiveness analyses at both manufacturers and universities. Furthermore, given that this system is positioned as a supplement to the existing medicine-pricing system, the range of drug-price adjustments is limited. In principle, the price-adjustment target is limited to a proportion of the product price: the utility premium. When the price of a medicine is calculated using the cost accounting method, the portion of the operating profit is also subject to price adjustment. To avoid excessive price adjustments, the upper limit of the price adjustment range is set after considering the overall price of the medicine [26, 27].

The evaluation workflow in Japan is similar to that in the UK. First, at the general meeting of the Central Social Insurance Medical Council (CSIMC), the targeted medicines are selected based on their financial impact. The

manufacturers discuss the framework for cost-effectiveness analyses with the Center for Outcomes Research and Economic Evaluation for Health (C2H)—the HTA agency in Japan. Thereafter, manufacturers conduct the cost-effectiveness analysis of the targeted medicine based on analytical guidelines. Their analysis is critically examined by academic groups comprising university researchers. Based on the analyses by manufacturers and reviews by academic groups, the overall results are summarized by the cost-effectiveness expert committee organized by experts in clinical medicine, statistics, medical economics, and bioethics. The results are finalized after approval by the general meeting of CSIMC, and medicine prices are adjusted accordingly [26].

Several years have passed since the system was implemented. As of April 2022, evaluation of seven medicines has been completed, and price adjustments have been made; however, many issues have emerged in that process.

The first concern is the timeline for cost-effectiveness evaluation. Under the Japanese system, manufacturers conduct analyses within 270 days, regardless of the nature of the targeted medicine. The critical review period for the academic groups is set at 90 or 180 days, depending on whether re-analysis of evidence submitted by manufacturers is needed [28]. HTA in the UK defines a standard timeline for appraisal completion. However, the duration may vary depending on the nature of the medicine as specified in the analytical guidelines—a difference from the Japanese system [13]. When it is difficult for manufacturers to comply within the specified period, they are required to report this at the general meeting of CSIMC. If the report is not approved, the cost-effectiveness of the targeted medicine is considered to be the lowest possible regardless of the results of the analyses. This rule motivates manufacturers to quickly analyze; however, lack of time can impact the quality of analysis. Early completion of the critical review component is often required at the CSIMC meeting, which motivates academic groups to complete their reviews as soon as possible. Moreover, insufficiency of review period for academic groups raises concern about the quality of their deliverables.

The second concern is compliance with analytical guideline. C2H has developed technical guidelines for cost-effectiveness analysis [29] based on HTA in foreign countries; therefore, they are generally not unique in their description. However, information availability about utility value, or health-related quality-of-life (HRQOL) scores, which are significant parameters of cost-effectiveness analyses, significantly differs between other foreign countries and Japan. Therefore, concern regarding compliance with analytical guideline also exists.

Analytical guidelines in Japan stipulate that utility values should be measured using a preference-based measure; thus, utility values measured using a tool called EuroQol 5 Dimensions (EQ-5D) should be prioritized. Estimation of utility values from other sources is also acceptable if EQ-5D values do not exist. Utility values measured in Japan prioritize the data from foreign countries. However, in Japan, the number of existing research studies that measure HRQOL is lower than in other countries [27]. Under these circumstances, it is tempting to use the EQ-5D data from relevant clinical trials of the medicines designated for cost-effectiveness analysis to meet the first priority criteria. However, during cost-effectiveness analyses by manufacturers, the follow-up period for each subject is often inadequate, especially in clinical trials of anti-cancer medicines. Additionally, the EQ-5D data obtained from these subjects are not considered to be sufficiently informative. In these cases, there is no definitive answer to the question of whether a limited information source that meets the conditions of priority should be used, or whether other abundant sources that do not meet priority criteria should be used.

1.4. Research Purpose

The purpose of this research is to obtain ideas for addressing the concerns regarding the cost-effectiveness evaluation system in Japan from HTA systems in foreign countries. HTA in the UK was the primary subject of the research, for two reasons. First, NICE has completed more than 1,000 medicine appraisals, and there are many cases to be analyzed. Second, documents prepared by manufacturers and HTA agencies are widely disclosed in the NICE appraisal process, and various types of information can be utilized for the analysis. To achieve the study objectives, the following research questions (RQ) were posed, and answers sought through analysis of appraisals in NICE.

RQ 1: Are there any characteristics of medicines that impact the appraisal-completion timeframe?

RQ 2: Is manufacturer conformance to analytical guidelines associated with TAC's acceptance of manufacturers' proposed utility values?

2. Literature Review

This chapter comprehensively reviews existing studies relevant to RQ1 and RQ2 to clarify what has been elucidated and what remains unknown. The review will provide insights to develop novel research plans.

2.1. Existing Studies Relevant to RQ1

Several studies about the time spent on HTA agencies' decision-making were inspired by concerns regarding restricted access to medicines through HTA implementation. For example, Stoykova et al. studied the time lag between when efficacy evidence was obtained and when cost-effectiveness evidence was obtained. Thirty NICE appraisals that were completed by October 2001 were investigated. The study identified a time lag of 3.2 years from when exploratory evidence on effectiveness was obtained to when the first evidence of cost-effectiveness was obtained. It is difficult to fully eliminate this time lag, as cost-effectiveness analysis is usually based on efficacy evidence. However, the study determined that NICE could reduce the time lag by actively leading decision-making under uncertainty [30]. This study did not directly investigate restricted access to medicines through the implementation of HTA; however, it did increase risk awareness.

Several studies have focused on the period from MA of a medicine to decision-making by the HTA agency, exploring if any restricted access to new medicines occurred based on HTA implementation. Akehurst et al. compared the time from the MA of 12 anti-cancer medicines to decision-making by HTA agencies among eight European countries. In Germany, the period was 119 days—the shortest period among the eight countries; in Spain, it was 713 days—the longest. These results should be interpreted with caution because each country implements HTA differently; however, these findings demonstrated that HTA agencies take considerable time for decision-making [31].

Ford et al. studied the differences between NICE and the Scottish HTA agency known as the Scottish Medicines Consortium (SMC). They compared the time from MA of medicines to decision-making by HTA agencies using appraisals that had been completed by August 2010. Results indicated that the decisions about reimbursement made by the two agencies were identical. However, they identified that the HTA agency decision-making time was shorter for SMC rather than for NICE. One of the reasons for NICE's delayed decision-making was discussions involving many stakeholders [32]. Similarly, Varnava et al. found that SMC and the Welsh HTA agency known as the All Wales

Medicines Strategy Group had shorter duration from MA to the HTA agency's decision-making compared to NICE [33]. Many studies have focused on the time required for HTA agencies to make decisions. However, the primary objective of these studies was to describe the time period; they did not consider the factors impacting the time period.

Several other studies have focused on factors affecting the time spent on decision-making by HTA agencies. Barham et al. investigated the factors affecting the time taken from the initiation of appraisals in NICE to the publication of decision-making guidance. Evaluation of 18 appraisals completed by 2008 revealed that the single technology assessment (STA) process introduced by NICE in 2005 reduced the appraisal period of anti-cancer medicines compared to the multiple technology assessment (MTA) process [34].

Casson et al. examined factors impacting the time taken from the development of an analytical framework, known as the scope document in NICE, to the release of decision-making guidance. The study covered the appraisal of non-anti-cancer medicines, differing from the study by Barham et al. Casson et al. focused on 196 appraisals completed by February 2010, revealing that the STA process reduced the period for appraisals compared to the MTA process—same as Barham et al.'s results. The Casson et al. study also revealed that decisions by NICE tended to be delayed in both STA and MTA processes, compared to the standardized timeline. They quantitatively demonstrated that publication of decisions tended to be delayed because of manufacturers' appeals for NICE's decision-making. Furthermore, they identified that appraisals of anti-cancer medicines took longer to yield decisions compared to other medicines [35]. Walton et al. quantitatively demonstrated that frequent NICE committee meetings to organize discussion points delayed its decision making [36].

Concerns have been raised about restricted access to medicines because of HTA implementation. Several studies have investigated the time taken from MA of medicines to decision-making by HTA agencies. Many researchers have attempted to determine the factors affecting decision-making time. A few studies have investigated if systems introduced to reduce the time required for appraisal, such as the STA process in NICE, are performing as expected. Others have quantitatively demonstrated the effects of factors that could be expected to be involved in prolonging the appraisal processes, such as the number of committee meetings.

However, research on the following two questions has been limited. First, whether the characteristics of medicines themselves affect the time-to-decision-making in NICE? Previous studies have focused on whether elements of the

appraisal processes, such as appeals by manufacturers or newly introduced systems including the STA process, affect NICE time-to-decision-making. There have been very few studies on the nature of medicines themselves, including application types or relevancy to orphan medicinal product (OMP) or anti-cancer medicines.

Second, whether there are factors affecting the time from MA to NICE decision-making? Many studies have examined the factors affecting the duration of the NICE appraisal process. For manufacturers who intend to introduce medicines to the market earlier, it is crucial to anticipate the period from MA of their medicines to decision-making by HTA agencies, as any knowledge about that time period provides significant insights for considering a launch strategy.

2.2. Existing Studies Relevant to RQ2

Prior to the publication of NICE's guidelines regarding the analytical method—the method guide—in 2004, only one study on utility values, referred to by studies regarding cost-effectiveness analyses, had been conducted. Brauer et al. investigated the process of obtaining utility values in 306 studies of cost-effectiveness analyses published between 1998 and 2001. They identified that 36% of the studies estimated utility values using direct elicitation, including the standard gamble method, time trade-off (TTO) technique, and rating scale method; 25% estimated utility values based on clinical judgment and expert opinions. Only 23% of the studies estimated utility values using preference-based outcome measures that were to be recommended later in NICE's method guide. However, the proportion of studies referring to utility values measured by preference-based outcome measures gradually increased during Brauer et al.'s study period, which was likely due to the simplicity of measurement [37].

The method guide recommended that manufacturers use EQ-5D results obtained in the relevant clinical trials of appraised medicines as the source for utility values in cost-effectiveness analyses. Subsequently, NICE published three technical support documents to supplement the method guide. The first document summarized the technical considerations for mapping techniques that could be employed when manufacturers cannot follow the recommended approach [38]. The second discussed possible options to estimate utility values when manufactures cannot follow the recommended approach. For example, it suggested the option of using general preference-based outcome measures other than EQ-5D, using direct elicitation such as TTO, or using interviews with patients and experts [39]. The third

document outlined points to remember when manufacturers employ utility values from various information sources in cost-effectiveness analyses. The issues emphasized by the document included utility value measurements for patients prior to taking medicines, utility values decrease over time, and uncertainty of these values [40].

After the method guide was published, several studies examined sources of utility values in cost-effectiveness analysis studies and HRQOL assessment conducted in clinical trials, expecting their use in HTA. Longworth et al. investigated the implementation of EQ-5D to assess HRQOL for several diseases. They clarified that EQ-5D could sensitively detect changes of HRQOL in skin manifestations and several cancers, but could not detect changes in hearing impairment. However, they identified that EQ-5D could detect HRQOL changes in hearing impairment by adding queries to the EQ-5D questionnaire [41].

Reed et al. focused on breast cancer and examined the status of HRQOL assessments conducted in clinical trials until October 2011; the number of studies evaluating HRQOL has been increasing year by year. They concluded that the European Organization for Research and Treatment of Cancer (EORTC) was the cancer-specific HRQOL measure used most frequently, followed by the Functional Assessment of Cancer Therapy (FACT); only one study used EQ-5D. They concluded that EQ-5D was not the best method to capture HRQOL changes resulting from the progression of breast cancer and associated medication, and proposed the use of cancer-specific measures in HTA [42].

Adlard et al. conducted a comprehensive review of studies of cost-effectiveness analyses for pediatric diseases and investigated those studies' sources of utility values. Forty-three studies published from 2004 to 2010 were identified, 11 of which used EQ-5D from relevant clinical trials as the source of utility values, in line with the NICE method guide. However, the 11 studies utilized modified EQ-5D for pediatric respondents—studies in which adult EQ-5D questionnaires were presented to pediatric patients and studies in which adults responded on behalf of pediatric patients. Adlard et al. proclaimed the need for standardization of methodologies for cost-effectiveness analyses of pediatric medications [43]. Studies by Longworth et al., Reed et al., and Adlard et al. summarized the sources of utility values in cost-effectiveness analyses or HRQOL measures conducted in clinical trials.

Several other studies analyzed the sources of utility values in NICE appraisals. Tosh et al. analyzed the sources of utility values for 39 appraisals conducted between 2004 and 2008, and identified that only 56% of manufacturers' analyses followed the method guide in terms of utility values. When manufacturers did not conform to the method

guide, they could not conduct the HRQOL assessment in relevant clinical trials, or it was difficult to utilize the utility values from relevant trials because of inappropriate validation [44].

Rose et al. also analyzed the sources of utility values, focusing on 12 appraisals of medicines for breast cancer conducted between 2006 and 2017. They revealed that more than half of the manufacturer analyses included in the appraisals did not follow the NICE method guide. As noted by Reed et al., this choice likely reflects the current feeling that EQ-5D is not the best method for capturing HRQOL changes because of progression of breast cancer and associated medication [45]. Hill focused on 40 appraisals of pediatric medicines and analyzed the sources of those appraisals' utility values. Just as Adlard et al. indicated, a wide variety of sources were used in manufacturers' analyses included in appraisals [46]. These studies indicate that utility values have been obtained from a variety of sources in NICE's appraisals.

Evidently, many studies on HRQOL assessment have been conducted since the publication of the NICE method guide. Multiple studies have been conducted to analyze HRQOL measures used in clinical trials and other studies to assess the sources of utility values in cost-effectiveness analyses, focusing on specific disease areas. Several studies have focused on the utility values cited in manufacturer analyses that were then included in NICE appraisals. These studies have shown that clinical trials have utilized a variety of measures to assess HRQOL, and that EQ-5D—recommended by the method guide—is not always used. Study results have also revealed that manufacturer analyses included in NICE appraisals do not always conform to the method guide regarding utility values.

Few studies have assessed whether manufacturer-proposed utility values estimated by a variety of measures are acceptable to the NICE TAC. It is crucial for manufacturers who bring appraisals to NICE to ascertain whether their utility values can be passed to the committee. This information may aid manufacturers' consideration of development strategies, including HRQOL assessment in clinical trials.

3. Research 1: Medicine Characteristics Affecting the Time to Guidance Publication by NICE

3.1. Introduction

In England, NICE plays a role in appraising new medicines that have received regulatory approval, with a view to make recommendations regarding their cost-effective use in the NHS. In the STA process, which covers a single medicine for a single indication, NICE develops a final scope (FS) after topic selection, which defines diseases, patients, and medicines covered by the appraisal. The submitting company discusses with NICE on how the decision problem will be addressed, and then, submits evidence. An ERG, which is independent of the NICE and the submitting company, is in charge of evaluating the manufacturer's evidence submission to identify its strengths and weaknesses. Based on the evidence submitted by the manufacturer and the evaluation of the ERG, one or more appraisal committee meetings are held to develop a final appraisal determination (FAD) document. Finally, NICE publishes technology appraisal guidance [47].

According to the standard timeline, NICE spends approximately 40 weeks to make a recommendation, which is generally after MA [30, 47]. This slow approach to the publication of guidance was criticized, especially considering medicines for life-threatening diseases or diseases with low treatment satisfaction, because patients have limited access to new medicines during the time gap between MA and guidance publication [48, 49]. Following the criticism, NICE took several measures to accelerate the process of guidance publication. For example, it established the STA process that reduced the time for guidance publication [50]. Since 2015, more than 300 guidance documents have been published, and more than 80 percent of them were achieved via the STA process [51, 52]. Furthermore, the recent appraisal process, which came into effect in 2018, established a policy that enabled cancer medicines to have shorter timelines for FS development [53].

Previous studies have shown that introducing the STA process contributed to speeding up the process of guidance publication [32, 34, 35]. Moreover, one study suggested that cancer medicines prolonged the appraisal process post FS development [35]. To the best of our knowledge, there are no studies on the various appraised medicine characteristics that affect the speed of guidance publication by NICE. We believe that the time gap between MA and guidance publication, rather than the total time NICE spends during the appraisal processes, is noteworthy, for the following two reasons. Firstly, the importance of implementing more and faster NICE appraisals for new medicines

and delivering a faster adoption of the most clinically and cost effective medicines is emphasized in the 2019 voluntary scheme for branded medicines pricing and access (VPAS), which announced that all new medicines will undergo an appropriate NICE appraisal by April 2020 [54]. Secondly, the time gap between MA and guidance publication better reflects the accessibility to new medicines; however, there is limited evidence regarding this time gap [32, 34, 55].

The aim of the present study was to investigate various appraisals or medicine characteristics that affect the time gap between MA and guidance publication by NICE, and explore the factors influencing this time gap by focusing on the cost-effectiveness analyses included in these appraisals.

3.2. Methods

3.2.1. Data Sources

We used publicly available documents from the websites developed by NICE, the European Medicines Agency (EMA), or the European Commission (EC). These included the FS, the first appraisal consultation document or the FAD issued by NICE, the annual reports issued by the EMA, and the Union Register of medicinal products developed by the EC.

3.2.2. Inclusion and Exclusion Criteria

We considered all technology appraisals (TAs) designated as STAs and completed by July 2020 for inclusion. However, we excluded appraisals completed before August 2016, as they followed a different process from the recent appraisal process. We also excluded appraisals if they were among the following: 1) terminated appraisals, 2) appraisals for medical devices, 3) appraisals that reviewed previous appraisals, and 4) appraisals that had been replaced by subsequent reviews. Additionally, we excluded appraisals if MA had been granted before any scope documents were published by the NICE because such cases follow a considerably different process as compared to the standard process [53].

3.2.3. Key Dates and Periods

From each TA report, we extracted the dates of the following: 1) FS publication, 2) FAD publication, 3) validation

of MA application by the EMA, and 4) MA. Based on the dates, we calculated the following time periods: 1) total number of months between MA and FAD publication (MA to FAD), 2) total number of months between validation of MA application and FS publication (VAL to FS), and 3) total number of months between FS publication and FAD publication (FS to FAD). The relationship among these three time-periods are summarized in Figure 3-1. We assumed that the MA to FAD period represented the period wherein patients had limited access to new medicines. After identifying medicine characteristics that were associated with longer or shorter MA to FAD period, we subsequently evaluated their influence on the VAL to FS and FS to FAD periods to explore the reason for this association, while focusing on the appraisal processes. We assumed that the VAL to FS period reflected how quickly NICE undertook topic selection and finalized the FS and that the FS to FAD period represented how long NICE spent time on discussions before FAD publication. VAL was used as an alternative indicator of the start date of the topic selection stage by NICE because the start date itself is not publicly available. We understand that it is possible to evaluate medicine characteristics affecting the length of the topic selection and scope development stage based on the period from VAL to FS.

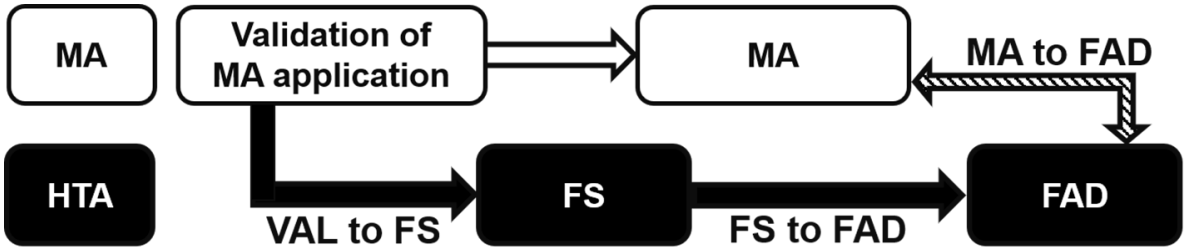

Figure 3-1 Key Periods Associated with the Evaluation Processes in MA and HTA

Abbreviation: FAD, final appraisal determination; FS, final scope; HTA, health technology assessment; MA, marketing authorization; VAL, validation of marketing authorization application.

3.2.4. Appraised Medicine Characteristics Associated with Key Periods

We identified the appraised medicine characteristics associated with key periods—MA to FAD, VAL to FS, and FS to FAD—based on literature review and considering the underlying appraisal processes. For example, it is logical that fewer appraisal experiences in a certain disease category could prolong the time required to provide the FS and FAD. Several studies have suggested that the appraisals of cancer medicines are complex and uncertain, given that

they tend to be associated with a high incremental cost-effectiveness ratio (ICER) gap between manufacturers and ERGs. Thus, such appraisals could take longer than usual to be completed [56, 57]. Therefore, we included the following variables: 1) year of appraisal completion, defined as the publication year of the FAD (2016–18, 2019–20); 2) application type (initial application, extension application); 3) total number of previous appraisals in the same disease category; 4) cancer medicines (no, yes); 5) OMPs designated by the EMA (no, yes); and 6) accelerated assessment (AA) granted by the EMA (no, yes). Disease categories were based on the "conditions and diseases" classification formulated by the NICE [58].

3.2.5. Factors Influencing the Period from FS to FAD

Once we had identified medicine characteristics that were associated with longer or shorter periods from MA to FAD, and their association with longer or shorter VAL to FS or FS to FAD periods, we used nine detailed variables regarding cost-effectiveness analyses to explore the reason why the identified characteristics were related to longer or shorter VAL to FS or FS to FAD periods. The variables regarding cost-effectiveness analyses were as follows: 1) medicine cost (10,000 pounds/month or less, more than 10,000 pounds/month); 2) total number of comparators specified in the FS (two or less, more than two); 3) ICER gap between the manufacturers' ICER bid in their initial evidence submission and the ERG's ICER described in its initial report, both of which were based on the list price of appraised medicines (20,000 pounds/quality-adjusted life-year [QALY] or less, more than 20,000 pounds/QALY); and 4) innovative technology acknowledged in the TA (non-innovative, innovative). Furthermore, factors regarding clinical trials included in the cost-effectiveness analyses were as follows: 5) number of subjects (500 or less, more than 500); 6) time from MA validation to approval, which represents the speed of the regulatory approval process (300 days or less, more than 300 days); 7) phase (others, phase 3); 8) double-blinded randomized control trial (DBRCT) (no, yes); and 9) comparators in the clinical trials (not specified in the FS as a comparator in the cost-effective analyses, specified in the FS).

3.2.6. Statistical Analysis

Multiple linear regression analysis was used to examine the association between the characteristics of appraised

medicines and the MA to FAD period. In the analysis, we used Cook's distance (Di) to test for highly influential observations, which were defined as having a Di larger than 0.5 and conducted a sensitivity analysis by excluding these data. Subsequently, we performed the same analysis to examine the association between these characteristics and the VAL to FS and FS to FAD periods. In addition, linear regression analyses were used to identify associations between the factors in the cost-effectiveness analyses and the FS to FAD period. We conducted a univariable analysis to narrow down the variables to be used in the multivariable analysis. As there were a few preliminary findings of the association, baseline variables ($p < 0.10$) in the univariable analysis were included in the multivariable analysis. For each analysis, we calculated the unstandardized partial regression coefficient (B) and 95 percent confidence interval (CI). We chose a complete–case analysis because the proportion of missing data was low. Variance inflation factors were calculated to assess multicollinearity between variables; factors greater than ten were considered to represent multicollinearity. All statistical analyses were conducted using StatsDirect ver. 3.3.3 (StatsDirect Ltd., Cheshire, UK). Values were considered statistically significant at $p < 0.05$.

3.3. Results

3.3.1. Overview of the Investigated STAs

One hundred and sixteen appraisals met the criteria for analysis, and an overview is presented in Table 3-1. Of the appraisals, 64 percent (74/116) were completed between 2016 and 2018, and 60 percent (69/116) were derived from the initial MA application. NICE had experienced less than ten appraisals in the same "conditions and diseases" for approximately 39 percent (45/116) of the appraisals and more than twenty appraisals for 30 percent (35/116). Furthermore, 62 percent (72/116) of the appraisals were cancer medicines and 22 percent (25/116) were OMPs; only 5 percent (6/116) were granted AA. Table 3-1 summarizes the factors influencing the FS to FAD period. In case of factors in the cost-effectiveness analyses, the monthly medicine cost was more than 10,000 pounds in 21 percent (24/116) of the appraised medicines. The total number of comparators specified in the FS was more than three in 58 percent (67/116) of appraisals. Moreover, the ICER gap between the manufacturer and ERG was more than 20,000 pounds/QALY in 40 percent (37/93) of appraisals, whereas only 14 percent (16/116) of medicines were referred to as innovative, and their values were not fully captured by their ICER. In terms of the factors regarding clinical trials

included in the cost-effectiveness analyses, the total number of subjects was more than 500 in 52 percent (60/116) of the appraised medicines, the time from validation of MA application to approval was more than 300 days in 50 percent (53/106) of appraisals, 85 percent (99/116) were of phase 3 trials and 54 percent (63/116) were of DBRCTs, and the comparators of 43 percent (50/116) of the trials were specified in the FS as comparators of the cost-effectiveness analyses.

Table 3-1 Summary of 116 STAs Investigated

	Total* (N=116)	2016-2018 (N=74)	2019-2020 (N=42)
Characteristics of appraised medicines			
Application type – no. (%)			
Initial application	69 (59.5)	45 (38.8)	24 (20.7)
Extension application	47 (40.5)	29 (25.0)	18 (15.5)
Previous appraisal – no. (%) †			
<10	45 (38.8)	32 (27.6)	13 (11.2)
10 to 20	36 (31.0)	26 (22.4)	10 (8.6)
>20	35 (30.2)	16 (13.8)	19 (16.4)
Cancer medicine – no. (%)			
No	44 (37.9)	28 (24.1)	16 (13.8)
Yes	72 (62.1)	46 (39.7)	26 (22.4)
OMP – no. (%)			
No	91 (78.4)	59 (50.9)	32 (27.6)
Yes	25 (21.6)	15 (12.9)	10 (8.6)
Accelerated assessment – no. (%)			
No	110 (94.8)	69 (59.5)	41 (35.3)
Yes	6 (5.2)	5 (4.3)	1 (0.9)
Factors regarding cost-effectiveness analyses			
Medicine cost			
≤ 10,000 pound/month	92 (79.3)	62 (53.4)	30 (25.9)
> 10,000 pound/month	24 (20.7)	12 (10.3)	12 (10.3)
No. of comparators in the FS			
≤ 2	49 (42.2)	27 (23.3)	22 (19.0)
> 2	67 (57.8)	47 (40.5)	20 (17.2)
ICER gap between the manufacture and the ERG#			
≤ 20,000 pound/QALY	56 (60.2)	35 (37.6)	21 (22.6)
> 20,000 pound/QALY	37 (39.8)	22 (23.7)	15 (16.1)
Innovative technology			
Non-innovative	100 (86.2)	62 (53.4)	38 (32.8)
Innovative	16 (13.8)	12 (10.3)	4 (3.4)
Factors regarding clinical trials included in cost-effectiveness analyses			
No. of subject			

≤ 500	56 (48.3)	36 (31.0)	20 (17.2)
> 500	60 (51.7)	38 (32.8)	22 (19.0)
Time to approval[#]			
≤ 300 days	53 (50.0)	32 (30.2)	21 (19.8)
> 300 days	53 (50.0)	32 (30.2)	21 (19.8)
Phase			
Others	17 (14.7)	10 (8.6)	7 (6.0)
P3	99 (85.3)	64 (55.2)	35 (30.2)
Double-blinded randomized control trial			
No	53 (45.7)	38 (32.8)	15 (12.9)
Yes	63 (54.3)	36 (31.0)	27 (23.3)
Comparator			
Not specified in the FS	66 (56.9)	36 (31.0)	30 (25.9)
Specified in the FS	50 (43.1)	38 (32.8)	12 (10.3)

* Percentages may not total 100 because of rounding.
† Previous appraisal means the total number of previous ones in the same "condition and diseases" categories. [58]
# Ninety-three or 106 appraisals were available because of some missing entries.
Abbreviation: ERG, evidence review group; FAD, final appraisal determination document; FS, final scope; ICER, incremental cost-effectiveness ratio; OMP, orphan medicinal product; QALY, quality-adjusted life year.

3.3.2. Association of the Appraised Medicine Characteristics with Key Periods

The median value of the MA to FAD period was 5.5 months (Figure 3-2). The periods from VAL to FS and FS to FAD among the appraised medicine characteristics are shown in Figures 3-3 and 3-4, respectively.

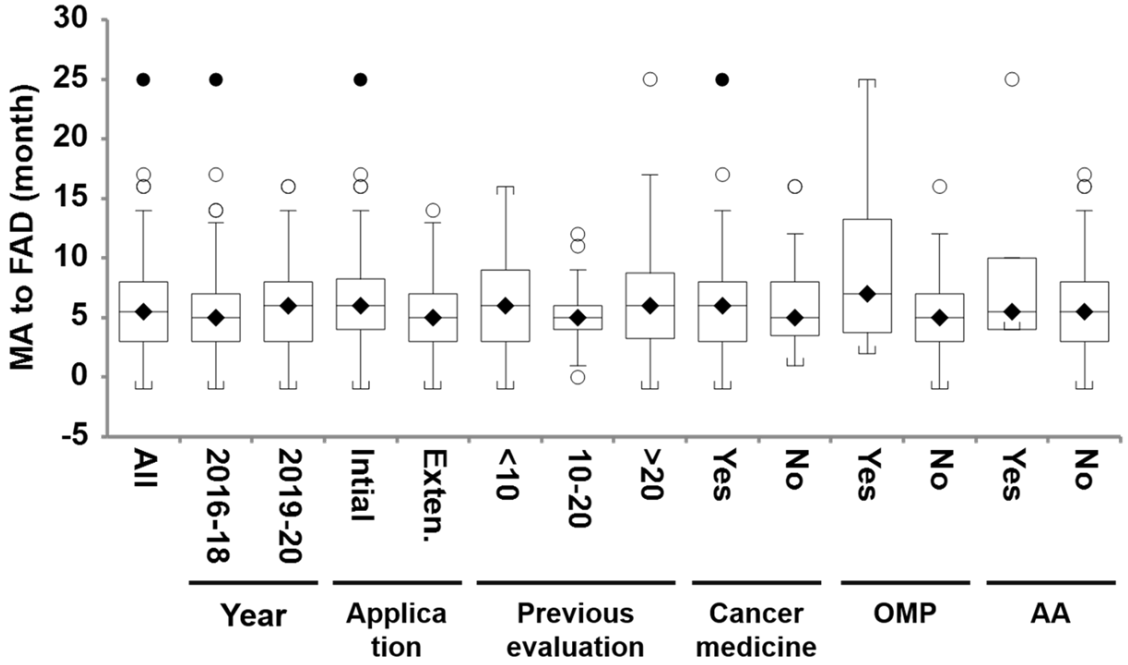

Figure 3-2 Box-and-Whisker Plot of the MA to FAD Period among Apprised Medicine Characteristics

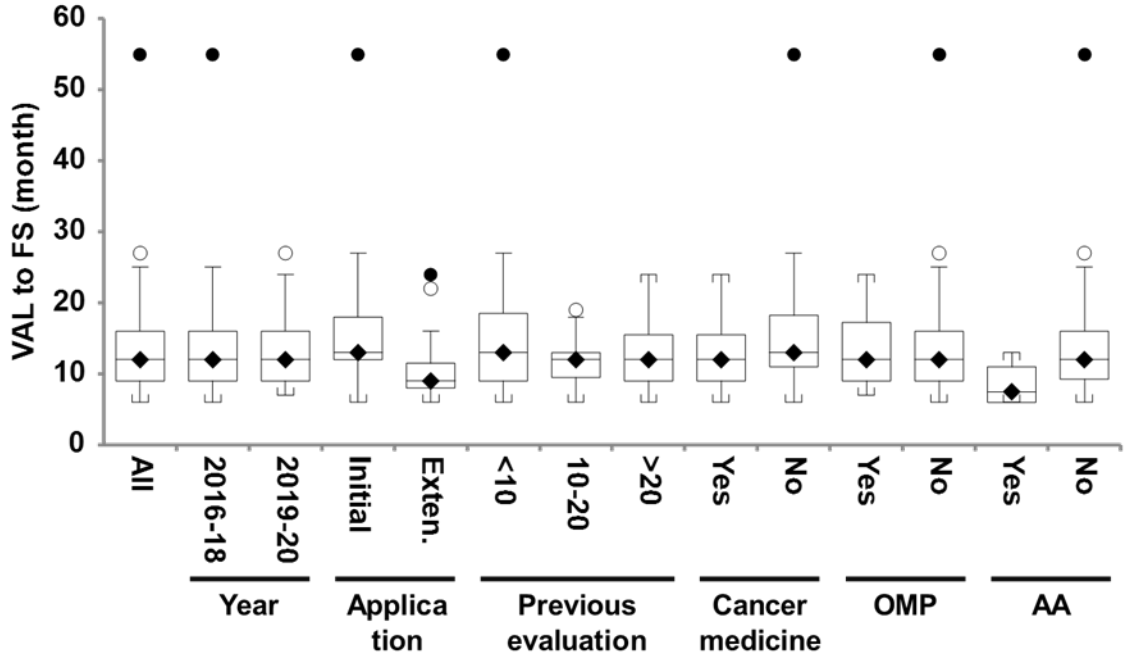

Figure 3-3 Box-and-Whisker Plot of the VAL to FS Period among Apprised Medicine Characteristics

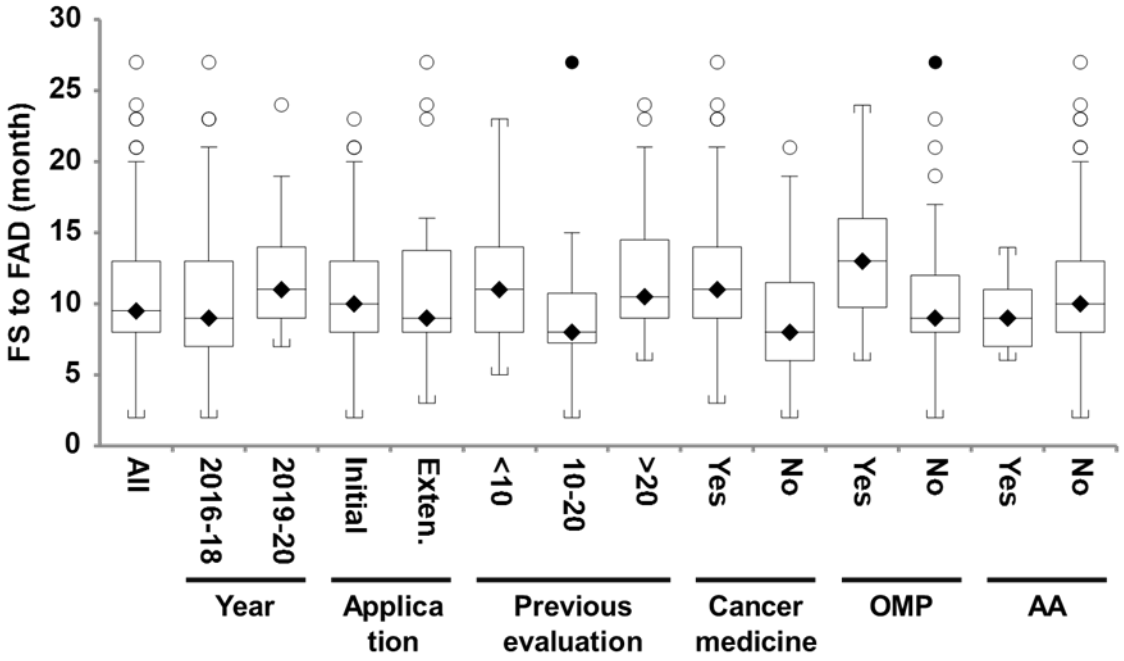

Figure 3-4 Box-and-Whisker Plot of the FS to FAD Period among Apprised Medicine Characteristics

The upper and lower whiskers are the upper or lower quartiles plus 1.5 times the interquartile distance. The horizontal lines that split the boxes in two represents median values, which are also expressed as the black diamonds on the boxes. The white and black circles denote outliers of 1.5 and 3 times the interquartile range, respectively.

Abbreviation: AA, accelerated assessment; FAD, final appraisal determination; MA, marketing authorization; OMP, orphan medicinal product.

There were positive associations between OMPs and the MA to FAD period (B = 3.042, 95 percent CI = 1.100–4.984). This means that OMPs were associated with a 3.042-month increase in the MA to FAD period. Cancer medicines and extension applications were negatively associated with the VAL to FS period (B = −3.405, 95 percent CI = −6.343 to −0.467 and B = −5.908, 95 percent CI = −8.564 to −3.252, respectively). There were positive associations between cancer medicines or OMPs and the FS to FAD period (B = 2.366, 95 percent CI = 0.464–4.268 and B = 2.833, 95 percent CI = 0.727–4.940, respectively) (Table 3-2). There was no multicollinearity between variables.

We identified TA429, which was in scope development for more than one year, as a highly influential observation [59]. We conducted a sensitivity analysis excluding this observation and obtained a result consistent with the main findings (Appendix Table 1).

Table 3-2 Multivariable Analysis of Appraised Medicine Characteristics Associated with Key Periods

Characteristics (no. of appraisals)	MA to FAD		VAL to FS*		FS to FAD	
	Unstandardized coefficients month, (95%CI)	P value	Unstandardized coefficients month, (95%CI)	P value	Unstandardized coefficients month, (95%CI)	P value
Completion year						
2016-2018 (74)	Reference		Reference		Reference	
2019-2020 (42)	0.194 (-1.365, 1.753)	0.805	-0.850 (-3.402, 1.702)	0.510	1.202 (-0.489, 2.893)	0.162
Application type						
Initial (69)	Reference		Reference		Reference	
Extension (47)	-0.572 (-2.150, 1.005)	0.474	-5.908 (-8.564, -3.252)	< 0.001	-0.049 (-1.759, 1.662)	0.955
Previous appraisal						
1 appraisal added	-0.015 (-0.077, 0.047)	0.631	-0.005 (-0.105, 0.095)	0.919	-0.033 (-0.100, 0.035)	0.340
Cancer medicine						
No (44)	Reference		Reference		Reference	
Yes (72)	0.302 (-1.453, 2.056)	0.734	-3.405 (-6.343, -0.467)	0.024	2.366 (0.464, 4.268)	0.015
OMP						
No (91)	Reference		Reference		Reference	
Yes (25)	3.042 (1.100, 4.984)	0.002	-0.242 (-3.352, 2.868)	0.878	2.833 (0.727, 4.940)	0.009
AA						
No (110)	Reference		Reference		Reference	
Yes (6)	2.478 (-0.877, 5.833)	0.139	-4.019 (-9.390, 1.351)	0.141	-1.202 (-4.840, 2.436)	0.514

* 109 appraisals were available, because there were some missing entries in the European Medicines Agency's validation date. Among the 109 appraisals, the 67 were completed from 2016 to 2018, the 69 were initial application, the 68 were cancer medicines, the 25 were OMPs, and the 6 were granted accelerated assessment.
Abbreviation: AA, accelerated assessment; CI, confidence interval; FAD, final appraisal determination; FS, final scope; MA, marketing authorization; OMP, orphan medicinal product; VAL, validation of marketing authorization application

3.3.3. Association of Factors in the Cost-Effectiveness Analysis with the FS to FAD period

In the preceding analysis, we found a positive association between OMPs and the MA to FAD period and that OMPs were also associated with a longer FS to FAD period compared to non-OMPs, while there was no association with longer or shorter VAL to FS periods. Based on the results, in the subsequent analysis, we focused on the association of detailed factors regarding cost-effectiveness analysis with the length of the FS to FAD period to explore

the reason why OMPs were related to a longer FS to FAD period. The univariable analysis revealed that the ICER gap between manufacturers and ERGs and the time from MA to approval were associated with an increase in the FS to FAD period. The FS to FAD period negatively associated with the following attributes: total number of comparators specified in the FS, DBRCT, and clinical trials in which the comparators were specified in the FS. Among them, the ICER gap was not used in the multivariable analysis, because it had a large number of missing entries. The multivariable analysis showed that independent factors associated with a shorter FS to FAD period were DBRCT and clinical trials in which the comparators were specified in the FS (B = −2.183, 95 percent CI = −3.733 to −0.633 and B = −2.637, 95 percent CI = −4.267 to −1.006, respectively) (Table 3-3). There was no multicollinearity between variables.

Table 3-3 Univariable and Multivariable Analyses of Factors Influencing the FS to FAD Periods

Factors (no. of appraisals)	FS to FAD		
	Univariable	Multivariable (N=106)	
	P value	Unstandardized coefficients month, (95%CI)	P value
Factors regarding cost-effectiveness analyses			
medicine cost			
≤ 10,000 pound/month (92)	0.524		
> 10,000 pound/month (24)			
No. of comparators in the FS			
≤ 2 (49)	0.092	Reference	
> 2 (67)		-1.474 (-2.993, 0.045)	0.057
ICER gap between the manufacture and the ERG			
≤ 20,000 pound/QALY (56)	0.009	NA*	
> 20,000 pound/QALY (37)			
Innovative technology			
Non-innovative (100)	0.289		
Innovative (16)			
Factors regarding clinical trials included in cost-effectiveness analyses			
No. of subject			
≤ 500 (56)	0.201		
> 500 (60)			
Time to approval			
≤ 300 days (53)	0.064	Reference	
> 300 days (53)		0.698 (-0.792, 2.188)	0.355
Phase			
Others (17)	0.163		
Phase 3 (99)			
Double-blinded randomized control trial			
No (53)	0.035	Reference	
Yes (63)		-2.183 (-3.733, -0.633)	0.006
Comparator			
Not specified in the FS (66)	0.021	Reference	
Specified in the FS (50)		-2.637 (-4.267, -1.006)	0.002

* ICER gap was not used in the multivariate analysis because it had large number of missing entries.
Abbreviation: ERG, evidence review group; FAD, final appraisal determination; FS, final scope; ICER; incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

The multivariable analysis, including the ICER gap, is shown in Appendix Table 2, in which similar trends to those observed in the main findings were obtained. For OMPs, clinical trial designs are shown in Table 3-4 and their time

between the validation of MA application and approval is shown in Appendix Table 3.

Table 3-4 Clinical Trial Designs referred to by the Cost-Effectiveness Analyses in the OMPs' appraisals

	Total	DBRCT	
		Yes	No
OMP, n (%) *	25 (100)	9 (36)	16 (64)
Non-OMP, n (%) *	91 (100)	54 (67)	37 (33)

* Percentages may not total 100 because of rounding.
Abbreviation: DBRCT, double-blinded randomized control trial; OMP, orphan medicinal product.

	Total	A clinical trial's comparator is specified in the FS as a comparator in cost-effectiveness analyses	
		Yes	No
OMP, n (%) *	25 (100)	8 (32)	17 (68)
Non-OMP, n (%) *	91 (100)	42 (46)	49 (54)

* Percentages may not total 100 because of rounding.
Abbreviation: FS, final scope; OMP, orphan medicinal product.

3.4. Discussion

In the present study, we assessed the appraised medicine characteristics that affect the speed of guidance publications by NICE. Among the 116 STAs, 25 had OMP designations. The OMPs were associated with a longer MA to FAD period than non-OMPs, and they had a positive association with the FS to FAD period. In terms of clinical trial-related factors included in each cost-effectiveness analysis, non-DBRCTs and comparators of trials not specified in the FS as the comparator in the cost-effectiveness analyses were associated with a prolonged FS to FAD period.

Earlier studies that assessed the association between the appraisal processes and the time to guidance publication by NICE showed that introducing the STA process or no appeals from manufacturers improved the speed of guidance publication [32, 34]. However, in these studies, little attention was paid to the appraised medicine characteristics that contributed to this. In addition, in these studies, only one of the periods was considered—the time to guidance publication by NICE or the time gap between MA and guidance publication. In contrast, in the present study, we included characteristics of both appraisal processes and appraised medicines. Considering a broad range of characteristics, we evaluated NICE appraisal processes (represented by the VAL to FS and FS to FAD periods) along

with the time gap between MA and guidance publication (represented by the MA to FAD period).

The present study showed a positive association between OMPs and the MA to FAD period. This may be explained by the association of longer FS to FAD period with OMPs than with non-OMPs, because there was no association between OMPs and other processes related to health technology assessment and MA, including the period between VAL and FS or between validation of MA application and approval. This longer FS to FAD period with OMPs would have partially been caused by their clinical trial designs, which were included in the cost-effectiveness analyses; 64 percent (16/25) of the trials with OMPs were non-DBRCTs, and 68 percent (17/25) used comparators not specified in the FS as a comparator in the analyses.

The prolonged period from FS to FAD in non-DBRCTs was probably due to the complicated evaluation of the added health benefits or potential biases. Of the non-DBRCTs, 26 percent (14/53) were non-randomized and were single-arm trials, which made it difficult to estimate the additional effectiveness of the appraised medicines. In most cases, published data were referred to; however, manufacturers and ERGs were required to carefully evaluate the heterogeneity between study populations and generalizability of the results to patients in the NHS [60–62].

In non-DBRCTs, biases derived from subjective outcome evaluations are inevitable because of their non-blinded setting. Utility data, which are essential for cost-utility analysis, are usually collected via questionnaires from patients, which would lead to a bias [63, 64]. To confirm their reliability, manufacturers and ERGs also should refer to other data sources, leading to a long FS to FAD period.

For the cost-effectiveness analyses, no information was available on comparators when comparators not specified in the FS were used in the clinical trials. In such cases, external references are necessary to evaluate the added health benefits of appraised medicines. For example, when only placebo-controlled trials are available, manufacturers usually conduct network meta-analyses, connecting the appraised medicines and comparators via placebo data [65, 66]. In the appraisal of ustekinumab for the treatment of ulcerative colitis, the manufacturer extracted relevant information from more than ten trials to evaluate the added health benefits of ustekinumab with five comparators. In this appraisal, the FS to FAD period was 12 months, which is longer than the median value (10 months) of 116 STAs. These processes supposedly require time, because NICE carefully discusses imbalances among the extracted trials; this may lead to a long FS to FAD period [67].

Both OMPs and cancer medicines were associated with a longer FS to FAD period than other medicines; this was consistent with a previous study finding, that is, cancer topics prolonged the time to guidance publication [35], whereas cancer medicines had no association with long MA to FAD periods. This was probably because the VAL to FS period was approximately 3.4 months shorter with cancer medicines than with non-cancer medicines. This can partially be explained by the new appraisal process, which came into effect in 2018, and showed that cancer medicines had a shorter topic selection stage, than non-cancer medicines [53]. Unlike cancer medicines, OMPs are not considered to shorten the VAL to FS period.

This study had some limitations. We could not consider the discounts offered by manufacturers, because discount information was not publicly available. These discounts were assumed to improve the ICER of the appraised medicines, probably leading to a shortened FS to FAD period. We could not take into account whether budget impact tests were implemented in parallel with each appraisal. Commercial discussions between manufacturers and the NHS England after such tests will allow NICE to plan potential changes to the timelines of appraisals, probably leading to a lengthened FS to FAD period. We did not deal with the quality of the cost-effectiveness analyses submitted by the manufacturers; this might affect the length of the FS to FAD period as well. It was impossible to quantify their quality, but we confirmed that almost all manufacturers conducted analyses using the submission template formulated by NICE. We could not take into consideration the scientific advice offered by NICE because the process is not disclosed. Such scientific advice would help manufacturers to understand the perspective of decision makers, which might shorten the STA process.

In summary, the findings of the present study suggest that OMPs are associated with a longer time between MA and guidance publication by NICE than non-OMPs; this may be attributed to the prolonged FS to FAD period. Limited access to new medicines for orphan diseases, along with low treatment satisfaction, negatively affects patients. To address this issue, increased efforts are needed to shorten NICE appraisal process for OMPs, which will help the VPAS commitments to assess all medicines as rapidly as possible and achieve their fast adoption.

4. Research 2: Acceptability of Manufacturer-Proposed Utility Values for NICE Cancer Medicine Appraisals

4.1. Introduction

In England, NICE performs appraisals of new medicines in terms of their cost-effectiveness for the NHS, and makes recommendations for the NHS based on these appraisals. Through examination of evidence submitted by medicine manufacturers and evidence review groups, the TAC determines and publishes TA guidance, which represents the TAC's final recommendation regarding the technology in question [53].

The cost-effectiveness of an appraised medicine is typically expressed in terms of cost per healthy year gained, or cost per QALY gained, when compared to a comparator medicine. QALYs are calculated by estimating the number of years of life a patient has left after receiving the treatment in question, and weighting each year using a HRQOL score [68]. According to the NICE's guide concerning the TA method (henceforth referred to as "the NICE method guide"), measurement of changes in HRQOL should be based on direct self-reports from patients, and the utility of these changes should be determined by comparing the reported HRQOL with public preferences using a choice-based method. NICE encourages the use of the EQ-5D, which has been evaluated in relevant clinical trials for measuring the HRQOL. When EQ-5D data are not available, NICE suggests that utility values be estimated by mapping other HRQOL measures or health-related benefits observed in relevant clinical trials to the EQ-5D, or that EQ-5D data be obtained from existing literature [69].

Some studies have suggested that the EQ-5D lacks sensitivity to changes in health [70, 71]; therefore, it seems logical that manufacturers, who naturally seek to underline the HRQOL-improving effect of their test medicines, would prefer to use more specialized scales in their clinical trials. In fact, it is much more common for manufacturers to employ disease-specific measures for the health condition of interest than use a general scale. For example, cancer-specific measures that are directly relevant and sensitive to cancer symptoms are used to evaluate the HRQOL of patients with cancer. Two common cancer-specific HRQOL measures are the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) and the Functional Assessment of Cancer Therapy General (FACT-G) [72, 73].

A descriptive study has examined the HRQOL measures used in clinical trials of health technologies relating to the treatment of breast cancer and pointed out a marked heterogeneity in terms of which measures were used [42].

Other studies have focused on the sources of utility values used in manufacturers' cost-utility analyses in NICE TAs and elucidated lacking or poor compliance with the NICE method guide for HRQOL [42–45]. However, to our knowledge, no previous study has investigated whether the TAC accepted manufacturer-proposed utility values and the reasons for non-acceptance.

The aim of the present study was to investigate whether the TAC's acceptability of manufacturer-proposed utility values is dependent on the information sources; this was examined by focusing on STAs of cancer medicines, which represent over half of all NICE technology appraisals conducted in the past five years.

4.2. Methods

4.2.1. Data Sources and Eligibility Criteria

We examined documents that are publicly available from the NICE's website. The files in question included the manufacturers' evidence submission as part of their initial appraisal consultation documents and the TA guidance prepared by NICE. Our inclusion criterion was any STA for a cancer medicine completed between January 2011 and December 2020 because, in cases of multiple technology appraisals, manufacturers' evidence submission often did not include the information sources of manufacturer-proposed utility values. We excluded appraisals if they were: 1) terminated before completion, 2) appraisals of medical devices, 3) appraisals that reviewed previous appraisals, or 4) appraisals that had been replaced by subsequent reviews. We also excluded appraisals for which the economic model considered health states other than pre-progression, post-progression, and death during cancer treatments because manufacturers need to estimate more utility values when their economic model was more complicated. For example, in the appraisal of trastuzumab emtansine for adjuvant treatment of human epidermal growth factor receptor 2-positive early breast cancer, the manufacturer developed a Markov model with seven health states and estimated six utility values [74]. Estimating more utility values may elevate the risk of non-acceptance by the TAC independent of the information sources of manufacturer-proposed utility values.

4.2.2. Data Extraction from Manufacturers' Evidence Submissions

For each appraisal, we collected data regarding the HRQOL measures from the manufacturers' submitted evidence.

First, we examined the "clinical effectiveness evidence" section of the manufacturers' evidence submission and identified the clinical trials from which evidence was used in the economic model (such clinical trials are henceforth referred to as "main trials"). If two or more clinical trials were listed, we selected the one from which data were used to estimate effectiveness of the appraised medicine as the main trial.

Second, again examining the "clinical effectiveness evidence" section, we identified the HRQOL measures applied in the main trials. HRQOL measures were classified into four categories: 1) EQ-5D, 2) EORTC QLQ, 3) FACT, and 4) others, respectively. The EORTC QLQ and FACT categories included both general measures (such as the FACT-G and EORTC QLQ-C30) and cancer-specific measures (such as the FACT-Breast or EORTC QLQ-BR23). Instruments other than the EQ-5D, EORTC QLQ, and FACT were classified as "others."

Third, we identified the information sources of manufacturer-proposed utility values for pre- and post-progression states. The information sources were classified into three categories: 1) EQ-5D, 2) mapping other measures to the EQ-5D, and 3) using existing literature or TA guidance. If the manufacturers' evidence submission adopted a time-to-death approach to estimate the patients' utility values, we interpreted this as using the same information source for both pre- and post-progression states.

Forth, we identified the median follow-up period at the time of manufacturers' evidence submission and the frequency of EQ-5D measurements in the main trials to explore the reason for the TAC's acceptance of the manufacturer-proposed utility values. The EQ-5D measurement frequency was classified into four categories: 1) at least once every two weeks, 2) at least once a month, 3) at least once every two months, and 4) less than once every two months. If two or more categories were applicable to one submission, we chose one that appeared with the highest frequency among them. When the EQ-5D measurement frequency was not constant during the follow-up period, we adopted one immediately after the treatment initiation of the investigational medicines.

4.2.3. Data Extraction from TA Guidance

We examined TA guidance to determine whether manufacturer-proposed utility values for pre- and post-progression states were subject to objection by NICE. We considered the values to be unacceptable for the TAC if the TAC's comments included words such as "inappropriate," "inadequate," "unfit," "irrelevant," and/or

"unacceptable." In contrast, we considered the values to be acceptable if the comments included antonyms of the above-mentioned words, or if there were no comments regarding the manufacturer-proposed utility values.

In cases where the manufacturer-proposed utility values were not accepted by the TAC, we investigated the reason for non-acceptance based on the description in the TA guidance regarding appropriateness of these values. Reasons were classified into three categories by referring to an existing published taxonomy of errors and threats to the credibility of health economic models [75] and based on discussions between the researchers: 1) inappropriate value for the UK population (e.g., using a higher utility value than that for the general UK population), 2) inappropriate data adjustment (e.g., no adjustment for age or gender), and 3) unreliable data source (e.g., using an extremely limited number of subjects). If more than two reasons were mentioned in the TA guidance, we only considered the most discussed one as the cause for non-acceptance in each appraisal because the less-discussed reasons alone were not necessarily sufficient to cause non-acceptance by the TAC.

4.2.4. Statistical Analysis

We categorized manufacturers' evidence submissions into those in which the information source for the utility values was the application of the EQ-5D in the main trials and those in which utility values were obtained through the other methods, respectively, and then used Fischer's exact test to assess the hypothesis that there were differences between these groups in the TAC's acceptance of the utility values. We also compared the reasons for non-acceptance of the manufacturer-proposed utility values stated in the TA guidance from both groups based on the hypothesis that there were differences between them in the reasons for non-acceptance by the TAC.

As exploratory analyses, we extracted manufacturers' evidence submissions in which the information source for the utility values was the application of the EQ-5D in the main trials, and then, used Mann-Whitney U test to assess the hypothesis that the median follow-up period at the time of submission were different between the type of the TAC's judgement (acceptable or unacceptable). We also compared the frequency of EQ-5D measurements in the main trials between them using Fischer's exact test based on the hypothesis that there were differences in the frequency.

These analyses were conducted separately for pre- and post-progression states. All analyses were performed using

StatsDirect ver. 3.3.3 (StatsDirect Ltd., Cheshire, UK). P values less than 0.05 were considered to be statistically significant.

4.3. Results

4.3.1. Characteristics of the Main Trials in Terms of Health-Related Quality-of-Life Measurement Approach

A total of 414 appraisals were completed between January 2011 and December 2020. Among them, 200 STAs were for cancer medicines. We investigated 136 appraisals, after excluding 64, in the present study. The number of STAs for cancer medicines showed an increasing trend over time; between 2011 and 2014, there were approximately five per year, while between 2017 and 2020, there were approximately twenty per year. The number of STAs in which the manufacturers' evidence submission contained main trials including EQ-5D measurements also increased over time, rising from 10 percent during 2011 and 2012 to 82 percent during 2019 and 2020. There were few main trials in which HRQOL was measured only through the EQ-5D; most featured multiple assessments combining both the EQ-5D and cancer-specific measures such as the EORCT QLQ and FACT (Figure 4-1).

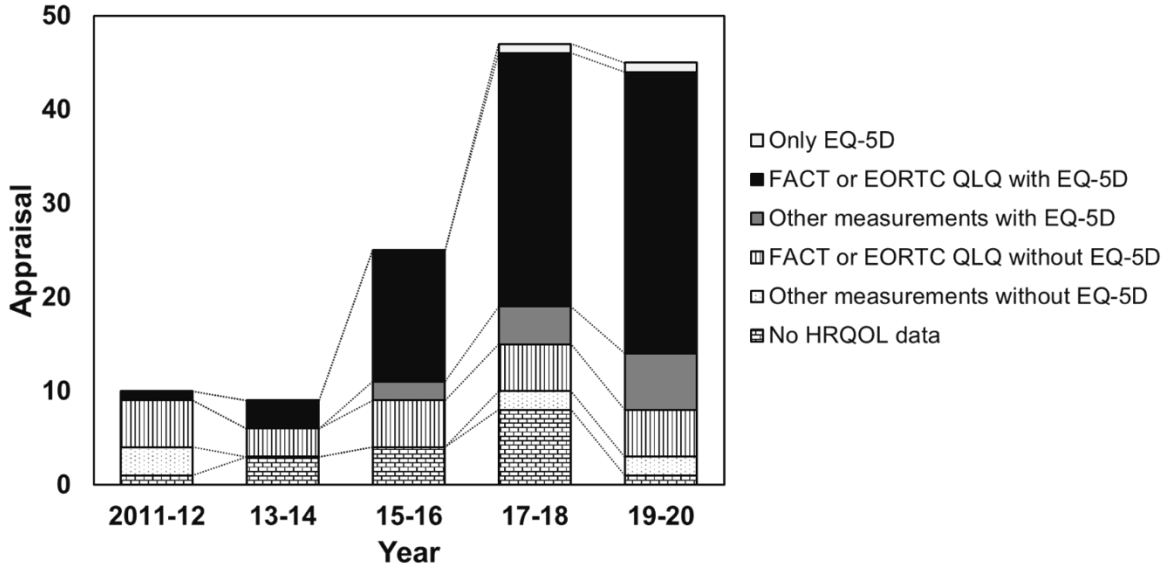

Figure 4-1 Health-Related Quality-of-Life Measurements Used in the Main Trials

Abbreviation: EORTC QLQ: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire; EQ-5D: EuroQol 5 Dimensions; HRQOL: health-related quality-of-life; FACT: Functional Assessment of Cancer Therapy.

4.3.2. Information Sources for Utility Values in the Manufacturers' Evidence Submissions

The information sources of the manufacturer-proposed utility values for both pre- and post-progression states are shown in Figures 4-2. There was an increase over time in the proportion of submissions for which the information source was the EQ-5D. For 2011–2012, the EQ-5D was the information source for 20 percent of the pre-progression state and 10 percent of the post-progression state; however, this rose to 84 percent and 56 percent, respectively, for 2019–2020. When considering the entire research period (i.e., 2011–2020), post-progression utility values, when compared to pre-progression values, were relatively heavily sourced from existing literature or TA guidance rather than the EQ-5D.

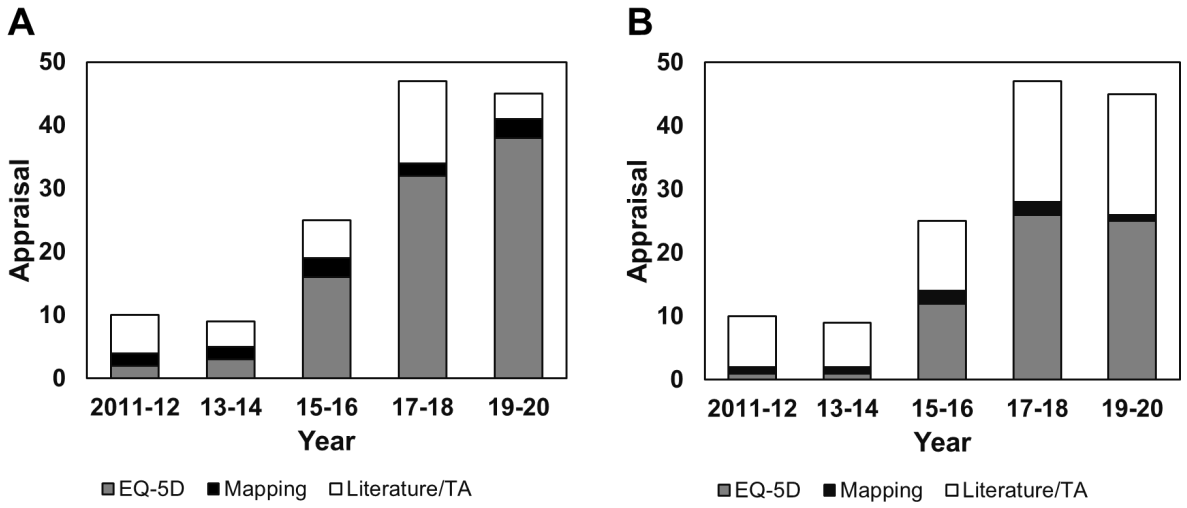

Figure 4-2 Information Sources of Manufacturer-Proposed Utility Values

Note. (A) Utility values for the pre-progression state, (B) utility values for the post-progression state.

Abbreviation: EQ-5D: EuroQol 5 Dimensions; TA: technology appraisal.

Table 4-1 shows the relationship between HRQOL measures used in the main trials and the information sources of manufacturer-proposed utility values. In ninety-one appraisals, manufacturer-proposed utility values for the pre-progression state were sourced through the application of the EQ-5D in the relevant clinical trials. Eighty-seven of them used the EQ-5D in the main trials, while the remaining four used the EQ-5D in relevant clinical trials of appraised medicines other than the main trials. In twelve appraisals, manufacturer-proposed utility values for the pre-progression state were sourced by mapping other HRQOL measures to the EQ-5D using the existing published algorithms. The HRQOL evidence mapped into the EQ-5D were measured in all the main trials. In thirty-three

appraisals, manufacturer-proposed utility values for the pre-progression state were sourced from existing literature or TA guidance. Forty-five percent of them measured HRQOL using the EQ-5D or other measures in the main trials, but used existing literature or TA guidance to estimate manufacturer-proposed utility values. For the post-progression state, compared to the pre-progression state, a higher number of appraisals (72 percent) fell into this category.

Table 4-1 Relationship between Information Sources for Manufacturer-Proposed Utility Values and Health-Related Quality-of-Life Measurements Performed in the Main Trials

Method used in manufacturers' evidence submission	Total n (%)	HRQOL measurements in the main trials, n (%)		
		Evaluated		Not evaluated
		Including EQ-5D	Not including EQ-5D	
Pre-progression				
EQ-5D	91 (100)	87 (96)	4 (4)[a]	0 (0)
Mapping other measures to the EQ-5D	12 (100)	0 (0)	12 (100)[b]	0 (0)
Existing literature/TA guidance	33 (100)	2 (6)	13 (39)	18 (55)
Post-progression				
EQ-5D	65 (100)	62 (95)	3 (5)[a]	0 (0)
Mapping other measures to the EQ-5D	7 (100)	0 (0)	7 (100)[b]	0 (0)
Existing literature/TA guidance	64 (100)	27 (42)	19 (30)	18 (28)

Abbreviation: EQ-5D: EuroQol 5 Dimensions; HRQOL: health-related quality-of-life; TA: technology appraisal.
[a] These indicate the appraisals where manufacturers used EQ-5D carried out in the relevant clinical trials other than the main trials as the information sources of manufacturer-proposed utility values.
[b] For the pre-progression state, manufacturers mapped eight EORTC QLQ, two FACT, and two MOS 36-Item Short-Form Health Survey (SF-36) to the EQ-5D, and for the post-progression state, they mapped four EORTC QLQ, one FACT, and two SF-36 to the EQ-5D.

4.3.3. The TAC's Considerations of the Information Sources of the Manufacturer-Proposed Utility Values

Table 4-2 shows the TAC's judgements on the manufacturer-proposed utility values by the type of utility values (EQ-5D in the main trial or others). Fisher's exact test revealed no significant differences between the type of utility values in the TAC's judgement (acceptable or unacceptable) for both pre- and post-progression states. For pre- and post-progression states, 67 percent (58/87) and 56 percent (35/62), respectively, of the manufacturer-proposed utility values derived from the application of the EQ-5D in the main trials were accepted by the TAC; meanwhile, 59 percent (29/49) and 57 percent (42/74), respectively, of the utility values derived from other means were accepted.

Table 4-2 The TAC's Judgments on the Manufacturer-Proposed Utility Values

Information sources of manufacturer-proposed utility values		Total n (%)	TAC's judgment, n (%)		p value[a]
			Acceptable	Unacceptable	
Pre-progression state	EQ-5D in the main trials	87 (100)	58 (67)	29 (33)	0.458
	Others	49 (100)	29 (59)	20 (41)	
Post-progression state	EQ-5D in the main trials	62 (100)	35 (56)	27 (44)	1.000
	Others	74 (100)	42 (57)	32 (43)	

Abbreviation: EQ-5D: EuroQol 5 Dimensions; TAC: technology appraisal committee.
[a] Fischer's exact test was conducted to test the hypothesis that there are differences in the proportion of acceptance by the TAC between the "EQ-5D between the main trials" and "others" groups.

In regard to the manufacturer-proposed utility values that were not accepted by the TAC, we compared the reasons for non-acceptance stated in the TA guidance by the type of utility values (EQ-5D in the main trial or others). Table 4-3 presents information on the statistically significant differences consequently found in the reasons for non-acceptance between the type of utility values for both the pre- and post-progression states. Among the manufacturers' evidence submissions that featured the EQ-5D in the main trials as a utility-value source, major reasons for non-acceptance were inappropriate values for the UK population (52 percent for pre-progression state and 41 percent for post-progression state) and inappropriate data adjustment (45 percent for pre-progression state and 52 percent for post-progression state). For the other group, reliability of the data source was a common reason for non-acceptance (50 percent for pre-progression state and 34 percent for post-progression state).

Table 4-3 Reasons for the TAC's Non-Acceptance of Manufacturer-Proposed Utility Values

Reasons for non-acceptance by the TAC		Information sources for manufacturer-proposed utility values, n (%)		p value[a]
		EQ-5D in the main trials	Others	
Pre-progression state	Total	29 (100)	20 (100)	-
	Inappropriate value for the UK population	15 (52)	7 (35)	< 0.001
	Inappropriate data adjustment	13 (45)	3 (15)	
	Unreliable data source	1 (3)	10 (50)	
Post-progression state	Total	27 (100)	32 (100)	-
	Inappropriate value for the UK population	11 (41)	14 (44)	0.014
	Inappropriate data adjustment	14 (52)	7 (22)	
	Unreliable data source	2 (7)	11 (34)	

Abbreviation: EQ-5D: EuroQol 5 Dimensions; TAC: technology appraisal.
[a] Fischer's exact test was conducted to test the hypothesis that there is a difference in the reasons for non-acceptance by the TAC between the "EQ-5D in the main trials" and "others" groups.

4-3-4. The Factors Affecting TAC's Acceptance of the Manufacturer-Proposed Utility Values

Figures 4-3 shows the median follow-up period in the main trials at the time of manufacturers' evidence submissions in which the information source for the utility values was the application of the EQ-5D in the trials. When the TAC accepted the manufacturer-proposed utility values for pre-progression state, the main trial had longer median follow-up period (78.0 weeks) than the unacceptable cases (47.6 weeks); however, there was no difference in median follow-up period between the type of the TAC's judgement about utility values for post-progression state (66.8 weeks for acceptable cases and 53.3 weeks for unacceptable cases).

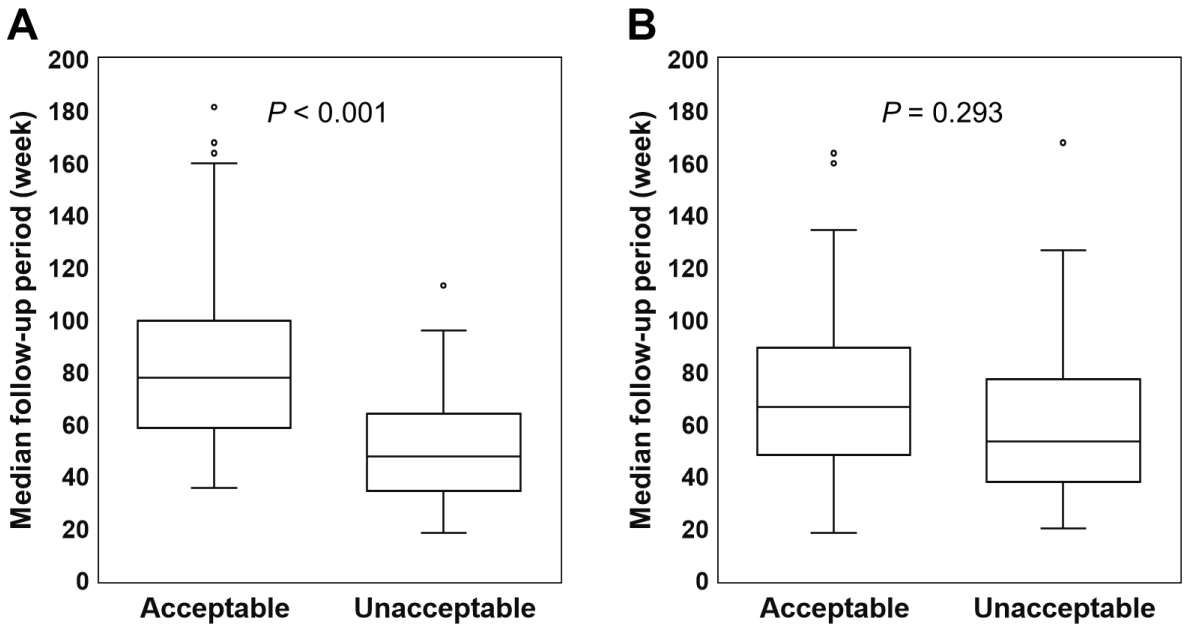

Figure 4-3 Median Follow-up Period of the Main Trials at the Time of Evidence Submissions

The upper and lower whiskers are the upper or lower quartiles plus 1.5 times the interquartile distance. The horizontal lines that split the boxes in two represents median value. The white circles denote outliers of 1.5 times the interquartile range. Mann-Whitney U test was performed to assess difference between the type of the TAC's judgement.

Table 4-4 shows frequency of EQ-5D measurement in the main trials included in the manufacturers' evidence submissions in which the information source for the utility values was the application of the EQ-5D in the trials. Measurement frequency was different between the type of the TAC's judgement about utility values for both pre- and post-progression states. When the TAC accepted the manufacturer-proposed utility values for pre-progression state, 74 percent (43/58) of the main trials measured EQ-5D at least once a month, while 38 percent (11/29) fell into this category in the unacceptable case. When the TAC accepted the manufacturer-proposed utility values for post-progression state, almost all (97 percent, 34/35) of the main trials measured EQ-5D at least once every two months, while 74 percent (20/27) fell into this category in the unacceptable case.

Table 4-4 Frequency of EQ-5D Measurement in the Main Trials

TAC's judgment		Total n (%)	Frequency of EQ-5D measurement, n (%)				p value[a]
			Once every two weeks	Once a month	Once every two months	Lesser extent	
Pre-progression state	Acceptable	58 (100)	6 (10)	37 (64)	10 (17)	5 (9)	0.002
	Unacceptable	29 (100)	0 (0)	11 (38)	11 (38)	7 (24)	
Post-progression state	Acceptable	35 (100)	3 (9)	19 (54)	12 (34)	1 (3)	0.007
	Unacceptable	27 (100)	0 (0)	17 (63)	3 (11)	7 (26)	

Abbreviation: EQ-5D: EuroQol 5 Dimensions; TAC: technology appraisal committee.
[a] Fischer's exact test was conducted to test the hypothesis that frequency of EQ-5D measurement in the main trials was different between the type of the TAC's judgement.

4.4. Discussion

In the present study, we assessed whether the NICE TAC's acceptance of manufacturer-proposed utility values is dependent on the manufacturers' information sources for these values. The number of appraisals for which the EQ-5D was the information source of the manufacturer-proposed utility values increased consistently over the period of 2011 to 2020. The TAC's acceptance of the manufacturer-proposed utility values was not dependent on the manufacturers' information sources, or whether they met the NICE method guide; the primary reasons for non-acceptance by the TAC differed between the manufacturers' evidence submissions that featured EQ-5D-sourced utility values and those that sourced utility values through other means.

Several previous studies have assessed information sources of manufacturer-proposed utility values [43–45]. These studies highlighted that there is variation in the methods manufacturers use to select and incorporate utility values in economic models, and that a large proportion of manufacturers' evidence submissions does not include data that accords with the NICE method guide. However, in these studies, little attention was paid to the TAC's acceptance of the manufacturer-proposed utility values. From the perspective of manufacturers, applying the EQ-5D in main trials seems important for meeting the NICE method guide but, due to the EQ-5D's lower sensitivity in comparison to disease-specific measures, the EQ-5D is not always useful for elucidating the HRQOL-improving effect of test medicines. Thus, in the present study we focused on the relationship between the information sources for manufacturer-proposed utility values and the TAC's acceptance of these values.

The present study showed that, between 2011 and 2012, 20 percent of the manufacturers' evidence submissions at

least partially met the NICE method guide regarding utility values; this percentage is comparable to that reported in a previous study, which showed that between 2004 and 2008, 32 percent of appraisals of cancer medicines satisfied these guides [44] and that between 2019 and 2020, approximately 80 percent of submissions were assumed to meet the guides to some extent. Considering the present finding of an upward trend in the proportion of main trials in which the EQ-5D was applied with other measurement tools, it is conceivable that manufacturers are increasingly attempting to propose utility values that meet the NICE method guide.

The present study found that, among the manufacturer-submitted evidence analyzed, utility values for the post-progression state were less likely to meet the NICE method guide than those for the pre-progression state. In other words, a considerable number of manufacturers ceased to use the EQ-5D as a source of utility values when they considered the post-progression state. This may be explained by the limitations concerning investigating the HRQOL of patients after disease progression. Several manufacturers' submissions mentioned that they collected HRQOL data in the pre-progression state only [76, 77]. Other submissions ceased using the EQ-5D in the post-progression state because they only collected the EQ-5D at the initial point of progression [78] or because their EQ-5D data were highly immature at the time of the preparation of the evidence submission [79]. This indicates that, even in cases when the EQ-5D was applied in the main trials, it was difficult to meet the NICE method guide regarding utility values during the post-progression state. NICE are currently reviewing the method guide to set a hierarchy of preferred methods for measuring HRQOL for when their preferred methods are not available or not appropriate, which will be helpful for manufacturers to estimate utility values during the post-progression state [80].

The present study showed that more than one-third of the appraisals for which manufacturer-proposed utility values were sourced through the application of the EQ-5D in the main trials were not accepted by the TAC. Thus, meeting the NICE method guide is not a sufficient condition for TAC's acceptance. In contrast, more than half of the appraisals for which manufacturer-proposed utility values were not sourced by the EQ-5D in the main trials were accepted by the TAC if manufacturers considered the best available data. A reason for this may be because manufacturers could refer to several completed appraisals for similar cancer types and treatment lines as over half of all NICE technology appraisals conducted in the past ten years are for cancer medicines. Utility values based on precedent appraisals would be at a low risk of non-acceptance by the TAC because they have been already discussed within NICE. These

two factors may contribute to the finding that the TAC's acceptance of the manufacturer-proposed utility values was not dependent on whether the utility values were sourced through the application of the EQ-5D in the main trials.

The present study categorized the reasons the TAC did not accept manufacturer-proposed utility values into three groups, and the submissions in question were differentiated depending on whether the information source for the manufacturer-proposed utility values was EQ-5D obtained during the main trials. Issues concerning the reliability of the data source, which could arise as a result of investigation of a small number of subjects or use of an unclear protocol [81, 82], are considered resolvable, because manufacturers could design detailed plans for improving their application of EQ-5D measurement in their main clinical trials.

In contrast, issues concerning inappropriate values for the UK population and inappropriate data adjustment are not always resolvable. For example, manufacturer-proposed utility values derived from multi-regional clinical trials have the potential for providing inappropriate values for the UK population because there may be differences between the UK and other countries or between patients included in clinical trials and patients in the UK in real-world settings regarding timings of diagnoses, supportive therapies, and intrinsic characteristics [77, 83]. Several manufacturer-proposed values estimated by mapping other HRQOL measures to the EQ-5D were not accepted by the TAC because they used inappropriate value sets that had not yet been validated. Such cases would be resolved by conducting additional validation studies. Meanwhile, adjustment of utility values based on aging may be difficult because, in many cases, the total evaluation period used for the EQ-5D in clinical trials is shorter than that used in epidemiologic studies referred to in appraisals [84]. In short, manufacturer-proposed utility values sourced from applying the EQ-5D in main trials are valuable in terms of showing reliability; however, their use might cause other issues due to the particular characteristics of clinical trials.

In contrast, there were some characteristics of the appraisals that may lead to their acceptance by the TAC. The present study showed that long-term survival follow-up at the time of manufacturer's submission and frequent EQ-5D measurements during the main trial resulted in lowering the risk of non-acceptance. In this case, the manufacturers can confirm the trend of utility decrement over time using evidence obtained through the main trials, and then consider whether they need an adjustment of utility values based on aging [85]. There may be other characteristics. For example, appraisals for first-line cancer treatment or for cancers associated with good prognoses also have the

potential to reduce the risk for non-acceptance [86, 87]. This could be because the general condition of the patients at the time of treatment initiation is more favorable than other cases, which results in a similarity in patients' general conditions between clinical trial settings and real-world settings in UK. The EQ-5D carried out in these main trials is expected to be considered appropriate for the UK population. These cases were thought to mitigate the unfavorable characteristics of clinical trials.

This study has several limitations. First, we did not consider any medicines or associated indications that were outside the scope of the NICE technology appraisals; therefore, the situation regarding HRQOL evaluations in clinical trials concerning such medicines was not examined in this study. Second, when evaluating the TAC's acceptance of utility values, we considered only the information source of the manufacturer-proposed utility values, not the absolute values of the HRQOL or the quality of the referenced HRQOL studies. Moreover, we did not consider the quality of the main trials or the cost-effectiveness analyses conducted by the manufacturers; this might affect the TAC's acceptance as well. Further studies are needed to confirm the impacts of these variables. Third, the presence of subjectivity in the selection of the main trials and the leading cause for non-acceptance cannot be excluded. However, we attempted to base our selection solely on the description in the manufacturers' evidence submission and the TA guidance to mitigate the subjectivity.

In summary, the present study's findings suggest that manufacturers make efforts to apply the EQ-5D in their main clinical trials with the aim of utilizing the resultant scores for the NICE technology appraisals; however, to obtain TAC acceptance in this regard, it is not sufficient merely to meet the NICE method guide. Manufacturers must consider in advance the possible differences between their clinical trial settings and real-world settings in UK, as well as the prospective quality of the EQ-5D data available from their trials, and then refine plans for EQ-5D measurement in order to obtain convincing evidence.

5. Overall Discussion and Conclusion

5.1. Answers to RQ 1 and Strategies for Overcoming the Concerns

Research 1 attempted to ascertain whether there are medicine characteristics that impact the duration of appraisal completion. The appraisals of OMPs required more time for discussion at NICE after manufacturers' evidence submissions. This can result in restrictions on access to the medicines in question. Additionally, the facts that the main clinical trials for the medicines in the appraisal were not DBRCT, and the comparators of the clinical trials and the cost-effective analysis did not match, were associated with longer discussion duration at NICE. These two factors were found to be more prevalent for OMPs compared to non-OMPs.

In the UK, HTA is conducted to determine reimbursement of medicines; Japan conducts it to partially adjust the drug price set under the existing system. Although the purpose of HTA is different between the two countries, it is possible that more time could be spent in the HTA agency's discussions for appraisals of OMPs in Japan as well. This is because the process of discussion on the validity for manufacturers' evidence submission is equally necessary, regardless of the purpose of HTA.

In Japan's cost-effectiveness evaluation system, two of the eight products for which evaluations were completed by July 2022 were OMPs [88, 89]. However, as both evaluations were conducted immediately after the system was launched, it is difficult at this point to assess whether the fact that the product is an OMPs impacted the time required for the appraisal. In other words, even if there was an impact on the time required for the appraisal, it is difficult to determine whether the impact was due to the fact that it was an appraisal for an OMPs or due to the degree of familiarity with the appraisal itself.

One method of reducing the impact on the time required for appraisals is to reduce the time required for discussions at the HTA agency after the manufacturers' evidence submission. However, this is not a problem that can be solved by the accumulation of cases, as this study found that the accumulation of appraisal experiences in the same disease area was not associated with a shorter or longer FS to FAD period. Even if the accumulation of cases were associated with a reduction of the period, the accumulation of cases itself is difficult in appraisals of OMPs.

However, as a method to reduce the impact on the time required for appraisal, it is possible to improve the timeline required for preparation prior to initiating manufacturer's analyses—that is, selecting targeted medicines for

appraisals and determining the analysis framework. For example, in NICE's appraisal of anti-cancer medicines, a system that does not restrict access to medicines, even if it takes more time for discussion, works well by accelerating the process of targeted medicine selection. In Japan's cost-effectiveness evaluation system, the selection of targeted medicines is made at the time of price listing, after which discussions on the analysis framework are initiated. By devising this process, delays in decision-making could be avoided in appraisals of medicines that are expected to take more time for discussion in HTA agencies.

5.2. Answers to RQ 2 and Strategies for Overcoming the Concerns

Research 2 attempted to ascertain whether manufacturers' evidence submissions that conformed to analytical guidelines biased TAC's judgement for manufacturer-proposed utility values. The research found that irrespective of manufacturers' evidence submissions conforming, or not conforming, to NICE's analytical guidelines, it did not bias TAC's acceptance of the manufacturer-proposed utility values. However, focusing on the cases in which TAC did not accept manufacturer-proposed utility values, the study showed that manufacturers' evidence submissions conformity to analytical guidelines biased the reasons for non-acceptance. In cases where manufacturers' evidence submissions were not in conformity to the guidelines, the main reason for non-acceptance was the quality of the information sources. In contrast, when they did conform, the main reasons were discrepancies from the actual medical practice in the UK and inappropriate data adjustment.

Conformity to analytical guidelines is one of the criteria to ensure that the utility values estimated by manufacturers' evidence submissions remain above a certain level of validity, and it does not guarantee acceptance by TAC. In other words, in the case of information sources that are highly restricted or biased, it is unreasonable to claim its validity on the basis of mere conformity to analytical guidelines. Based on the purpose of the appraisals, manufacturer-proposed utility values should be estimated from the perspective of utilizing the most appropriate information sources from among the numerous options available, and it is desirable to recognize that the analytical guidelines serve as one guiding principle for this process.

In Japan's cost-effectiveness evaluation system, analytical guidelines provide recommendations on information sources for utility values, similar to those in the UK. However, at present, there are few examples of evaluations, and

it is difficult to assess whether the same trend can be seen in Japan for manufacturer-proposed utility values as in the present research. However, a close examination of appraisals for which evaluations had been completed by July 2022 revealed that although manufacturer-proposed utility values met the recommendations of analytical guidelines, there was a case in which such values were not accepted by HTA agencies because of inappropriate handling of the values [90]. It is anticipated that the validity of manufacturer-proposed utility values will continue to be a major issue in appraisals, and the impact on the appraisal timeline will also remain a concern. In such cases, the strategy proposed by this research, which is to thoroughly review the quality of the EQ-5D data obtained from main clinical trials in advance of their initiation, is expected to be useful in facilitating the appraisal process.

Manufacturers' evidence submission to the HTA agency must be completed within 270 days of the selection of the targeted medicines, which makes it difficult to obtain new data after selection. In this case, manufacturers have no other choice but to consider the relative appropriateness of the limited number of options available at that time. In order for manufacturers to estimate manufacturer-proposed utility values that meet the purpose of appraisal based on guideline descriptions, strategic preparation from the clinical development stage is considered necessary.

5.3. Recommendations to Manufacturers

To overcome the two concerns related to Japan's cost-effectiveness evaluation system raised in this study, we propose the following three points to manufacturers that play a part in this system. First, we propose that in-house experts in health economics participate in the planning of main clinical trials that form the basis of regulatory approval applications, to provide suggestions based on their use in HTA. The participation of such experts is expected to facilitate discussion of the frequency of the EQ-5D measurement and the overall study design from a higher-order perspective of ensuring acceptable quality for HTA agencies, rather than merely measuring EQ-5D to meet analytical guideline recommendations.

Second, we suggest that the preliminary consultation between manufacturers and the HTA agency after targeted medicines selection should disclose and discuss in detail the information sources of the parameters to be used. If possible, manufacturers should present multiple proposals to the HTA agency for discussion and comparison, which is expected to clarify the advantages and challenges of each source of information. This is expected to be useful in

facilitating discussions at the HTA agency after the manufacturers' evidence submission.

Third, we suggest that manufacturers begin preparations for information sources of parameters used in cost-effectiveness analyses earlier to enable in-depth discussions during the preliminary consultation with the HTA agency. Considering that many of the parameters would be based on main clinical trials that will form the basis of regulatory filings, it is preferable to begin the preparation from the planning stage of such clinical trials.

5.4. Recommendations to HTA Bodies

Similarly, we propose the following three points to HTA bodies, including the HTA agencies responsible for the operation of Japan's cost-effectiveness evaluation system. First, a forum should be established for discussions with manufacturers from the drug development stage with a view to HTA. For example, we believe that there is a requirement of a system that enables representatives of HTA agencies to participate in consultation between manufacturers and the Pharmaceuticals and Medical Devices Agency prior to the commencement of main clinical trials that will form the basis of the application. Currently, the administrative body in charge of reviewing new medicines for their approval is different from that in charge of HTA, so these is a high barrier to its implementation. However, it is considered to be an effective method of ensuring that data for highly reliable HTA is obtained from clinical trials. Such methods have already been utilized in the UK and the EU, and are considered helpful when designing the system [91, 92].

Second, we suggest that manufacturers and the HTA agency begin high-level discussions regarding the cost-effectiveness analyses prior to preliminary consultation. Given the current preliminary consultation timeline, there is insufficient time to sort through the available information sources and their priorities. Consequently, in a few cases, they may have to be discussed after the results of manufacturers' analysis have been obtained, and there are concerns about the impact on the appraisal timeline. Commencing discussions of a highly specialized nature earlier is expected to reduce this risk. A framework for discussions between manufacturers and the HTA agency at the pre-selection stage already exists, and it would be desirable to strengthen the HTA agency structure to facilitate this and encourage manufacturers to make further use.

Third, we propose that reports on the results of appraisals be made public, focusing on the discussion process and

reasons for decisions, which are currently scarcely disclosed. Knowledge from publicly available information would be useful to many stakeholders in terms of allowing manufacturers prepare higher quality data and avoiding duplication of discussions among the manufacturers and the HTA agency in appraisals.

5.5. Conclusion

This study addressed two concerns in the cost-effectiveness evaluation system in Japan through two researches. Both concerns are unique; however, their solutions lie in the same direction: manufacturers and HTA bodies should collaborate and begin preparing for cost-effectiveness evaluation from the clinical development stage of each medicine with a view to implementing it in the future. Although it remains vague until drug price calculation whether cost-effectiveness evaluation will be necessary, strategic preparation from a more up-front stage is necessary to make timely and high-quality decisions.

Given that the cost-effectiveness evaluation system in Japan has a short history of operation and its objectives differ from those in other countries, it is assumed that new concerns will continue to emerge. We hope that a system will be established to continuously provide more efficient medicines within limited financial resources through organic collaboration and constructive discussions between manufacturers and HTA bodies. Finally, we believe this study will be significant in this regard.

Reference

1. Organisation for Economic Co-operation and Development [Internet]. OECD statistics. Available at https://stats.oecd.org/. Accessed 2022 June 10.

2. Statistics Bureau of Japan [Internet]. Population estimates. Available at https://www.stat.go.jp/english/data/jinsui/index.html. Accessed 2022 June 10.

3. Ministry of Health, Labour and Welfare [Internet]. Estimates of National Medical Care Expenditure, fiscal year 2019 [in Japanese]. Available at https://www.mhlw.go.jp/toukei/list/37-21.html. Accessed 2022 June 10.

4. Ministry of Health, Labour and Welfare [Internet]. Decomposing trends in the growth of healthcare expenditure [in Japanese]. Available at https://www.mhlw.go.jp/stf/shingi2/0000137679.html. Accessed 2022 June 10.

5. Mihara M, Kasutani K, Okazaki M, et al. Tocilizumab inhibits signal transduction mediated by both mIL-6R and sIL-6R, but not by the receptors of other members of IL-6 cytokine family. Int Immunopharmacol. 2005;5(12):1731-1740.

6. Makuvet FD. Biologics vs. small molecules: Drug costs and patient access. Medicine in Drug discovery. 2021;9:100075.

7. Ministry of Health, Labour and Welfare [Internet]. Japanese National Health price listing for new drug [in Japanese]. Available at https://www.mhlw.go.jp/stf/shingi/0000055527.html. Accessed 2022 June 11.

8. Ministry of Health, Labour and Welfare [Internet]. Japanese National Health out-of-cycle price revision for drug [in Japanese]. Available at https://www.mhlw.go.jp/stf/shingi2/0000142573.html. Accessed 2022 June 11.

9. Feinberg BA, Lang J, Grzegorczyk J, et al. Implementation of cancer clinical care pathways: a successful model of collaboration between payers and providers. J Oncol Pract. 2012;8(3):e38s-43s.

10. Snowdon AW. A Blueprint for Innovation to Achieve Health System Transformation. Healthc Pap. 2017;16(3):8-11.

11. International Network of Agencies for Health Technology Assessment [Internet]. What is Health Technology Assessment ? Available at https://www.inahta.org/. Accessed 2022 June 11.

12. Beletsi A, Koutrafouri V, Karampli E, Pavi E. Comparing Use of Health Technology Assessment in

Pharmaceutical Policy among Earlier and More Recent Adopters in the European Union. Value Health Reg Issues. 2018;16:81-91.

13. National Institute for Health and Care Excellence [Internet]. Technology appraisal processes. Available at https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/process. Accessed 2022 June 13.

14. Maynard A, Bloor K. The economics of the NHS cancer medicines fund. Appl Health Econ Health Policy. 2011;9(3):137-8.

15. Littlejohns P, Weale A, Kieslich K, et al. Challenges for the new Cancer Medicines Fund. Lancet Oncol. 2016;17(4):416-418.

16. National Health Service [Internet]. Innovative Medicines Fund. Available at https://www.england.nhs.uk/medicines-2/innovative-medicines-fund/. Accessed 2022 June 13.

17. National Institute for Health and Care Excellence [Internet]. Technology appraisal data. Available at https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/data. Accessed 2022 June 13.

18. Department of Health and Social Care [Internet]. Voluntary scheme for branded medicines pricing and access. Available at https://www.gov.uk/government/publications/voluntary-scheme-for-branded-medicines-pricing-and-access. Accessed 2022 June 13.

19. Raimond V, Kaltenbach E, Adam C, et al. Challenges raised by the economic evaluation of CAR-T-cell therapies. The review by the French National Authority for Health. Bull Cancer. 2021;108(12):1162-1169.

20. Schwander B, Banz K, Kaier K, Walzer S. Evaluation of a possible predictor for Federal Joint Committee decisions on early benefit assessments according to the German act on the reform of the market for medicinal products. Health Policy. 2014;117(3):334-344.

21. Tandvards-och lakemedelsformansverket (Dental and Pharmaceutical Benefits Agency) [Internet]. Pricing and reimbursement of medicines. Available at https://www.tlv.se/in-english/medicines/pricing-and-reimbursement-of-medicines.html. Accessed 2022 June 13.

22. College voor Zorgverzekeringen (Health Insurance Board) [Internet]. Advising on and clarifying the

contents of the standard health care benefit package. Available at https://english.zorginstituutnederland.nl/about-us/tasks-of-the-national-health-care-institute/advising-on-and-clarifying-the-contents-of-the-standard-health-care-benefit-package. Accessed 2022 June 13.

23. Canadian Agency for Medicines and Technologies in Health [Internet]. CADTH Reimbursement Reviews. Available at https://www.cadth.ca/cadth-reimbursement-reviews. Accessed 2022 June 13.

24. Department of Health [Internet]. About the PBS. Available at https://www.pbs.gov.au/info/about-the-pbs. Accessed 2022 June 13.

25. Ministry of Health, Labour and Welfare [Internet]. Summary of discussion on formal implementation of cost-effectiveness evaluations in Japan [in Japanese]. Available at https://www.mhlw.go.jp/stf/shingi2/0000172096.html. Accessed 2022 June 14.

26. Hasegawa M, Komoto S, Shiroiwa T, Fukuda T. Formal Implementation of Cost-Effectiveness Evaluations in Japan: A Unique Health Technology Assessment System. Value Health. 2020;23(1):43-51.

27. Kido K, Matsumaru N, Tsukamoto K. Health Technology Assessment in Japan: A Pharmaceutical Industry Perspective. Ther Innov Regul Sci. 2019;53(4):472-480.

28. Ministry of Health, Labour and Welfare [Internet]. Revision of the cost-effectiveness evaluations system in fiscal year 2022 [in Japanese]. Available at https://www.mhlw.go.jp/stf/shingi2/0000212500_00134.html. Accessed 2022 June 14.

29. National Institute of Public Health [Internet]. Guideline for cost-effectiveness evaluation in Japan. Available at https://c2h.niph.go.jp/en/tools/guideline/index.html. Accessed 2022 June 14.

30. Stoykova B, Drummond M, Barbieri M, Kleijnen J. The lag between effectiveness and cost-effectiveness evidence of new drugs. Implications for decision-making in health care. Eur J Health Econ. 2003;4(4):313-318.

31. Akehurst RL, Abadie E, Renaudin N, Sarkozy F. Variation in Health Technology Assessment and Reimbursement Processes in Europe. Value Health. 2017;20(1):67-76.

32. Ford JA, Waugh N, Sharma P, Sculpher M, Walker A. NICE guidance: a comparative study of the introduction of the single technology appraisal process and comparison with guidance from Scottish Medicines Consortium. BMJ Open. 2012;2(1):e000671.

33. Varnava A, Bracchi R, Samuels K, Hughes DA, Routledge PA. New Medicines in Wales: The All Wales Medicines Strategy Group (AWMSG) Appraisal Process and Outcomes. Pharmacoeconomics. 2018;36(5):613-624.

34. Barham L. Single technology appraisals by NICE: are they delivering faster guidance to the NHS?. Pharmacoeconomics. 2008;26(12):1037-1043.

35. Casson SG, Ruiz FJ, Miners A. How long has NICE taken to produce Technology Appraisal guidance? A retrospective study to estimate predictors of time to guidance. BMJ Open. 2013;3(1):e001870.

36. Walton MJ, O'Connor J, Carroll C, Claxton L, Hodgson R. A Review of Issues Affecting the Efficiency of Decision Making in the NICE Single Technology Appraisal Process. Pharmacoecon Open. 2019;3(3):403-410.

37. Brauer CA, Rosen AB, Greenberg D, Neumann PJ. Trends in the measurement of health utilities in published cost-utility analyses. Value Health. 2006;9(4):213-218.

38. Longworth L, Rowen D. NICE DSU Technical Support Document 10: The Use of Mapping Methods to Estimate Health State Utility Values. London: National Institute for Health and Care Excellence; April 2011.

39. Brazier J, Rowen D. NICE DSU Technical Support Document 11: Alternatives to EQ-5D for Generating Health State Utility Values. London: National Institute for Health and Care Excellence; March 2011.

40. Ara R, Wailoo A. NICE DSU Technical Support Document 12: The Use of Health State Utility Values in Decision Models. London: National Institute for Health and Care Excellence; July 2011.

41. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technol Assess. 2014;18(9):1-224.

42. Reed E, Kössler I, Hawthorn J. Quality of life assessments in advanced breast cancer: should there be more consistency?. Eur J Cancer Care. 2012;21(5):565-580.

43. Adlard N, Kinghorn P, Frew E. Is the UK NICE "reference case" influencing the practice of pediatric quality-adjusted life-year measurement within economic evaluations?. Value Health. 2014;17(4):454-461.

44. Tosh JC, Longworth LJ, George E. Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. Value Health. 2011;14(1):102-109.

45. Rose M, Rice S, Craig D. Does Methodological Guidance Produce Consistency? A Review of

Methodological Consistency in Breast Cancer Utility Value Measurement in NICE Single Technology Appraisals. Pharmacoecon Open. 2018;2(2):97-107.

46. Hill H, Rowen D, Pennington B, Wong R, Wailoo A. A Review of the Methods Used to Generate Utility Values in NICE Technology Assessments for Children and Adolescents. Value Health. 2020;23(7):907-917.

47. The National Institute for Health and Care Excellence [Internet]. Single technology appraisal timeline. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/process/sta-timeline. Accessed 2022 July 4.

48. Mayor S. NICE to issue faster guidance on use of drugs by NHS. BMJ. 2005;331(7525):1101.

49. Haycox A. Does 'NICE blight' exist, and if so, why?. Pharmacoeconomics. 2008;26(12):987-989.

50. Kaltenthaler E, Tappenden P, Booth A, Akehurst R. Comparing methods for full versus single technology appraisal: a case study of docetaxel and paclitaxel for early breast cancer. Health Policy. 2008;87(3):389-400.

51. The National Institute for Health and Care Excellence [Internet]. Technology appraisal recommendations. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/data/appraisal-recommendations. Accessed 2022 July 4.

52. Kaltenthaler E, Papaioannou D, Boland A, Dickson R. The National Institute for Health and Clinical Excellence Single Technology Appraisal process: lessons from the first 4 years. Value Health. 2011;14(8):1158-1165.

53. The National Institute for Health and Care Excellence [Internet]. Guide to the processes of technology appraisal. Available from: https://www.nice.org.uk/process/pmg19/chapter/foreword. Accessed 2022 July 4.

54. Department of Health and Social Care, UK [Internet]. Voluntary scheme for branded medicines pricing and access. Available from: https://www.gov.uk/government/publications/voluntary-scheme-for-branded-medicines-pricing-and-access. Accessed 2022 July 4.

55. Connolly E, O'Donnell H, Lamrock F, Tilson L, Barry M. Health Technology Assessment of Drugs in Ireland: An Analysis of Timelines. Pharmacoecon Open. 2020;4(2):287-296.

56. Sabry-Grant C, Malottki K, Diamantopoulos A. The cancer drugs fund in practice and under the new framework. Pharmacoeconomics. 2019;37(7):953-962.

57. Morrell L, Wordsworth S, Schuh A, Middleton MR, Rees S, Barker RW. Will the reformed Cancer Drugs

Fund address the most common types of uncertainty? An analysis of NICE cancer drug appraisals. BMC Health Serv Res. 2018;18(1):393.

58. The National Institute for Health and Care Excellence [Internet]. Conditions and diseases. Available from: https://www.nice.org.uk/guidance/conditions-and-diseases. Accessed 2022 July 4.

59. The National Institute for Health and Care Excellence [Internet]. Ibrutinib for previously treated chronic lymphocytic leukaemia and untreated chronic lymphocytic leukaemia with 17p deletion or TP53 mutation. Technology appraisal guidance [TA429]. Available from: https://www.nice.org.uk/guidance/ta429. Accessed 2022 July 4.

60. Hettle R, Corbett M, Hinde S, Hodgson R, Jones-Diette J, Woolacott N, et al. The assessment and appraisal of regenerative medicines and cell therapy products: an exploration of methods for review, economic evaluation and appraisal. Health Technol Assess. 2017;21(7):1-204.

61. Aballéa S, Thokagevistk K, Velikanova R, Simoens S, Annemans L, Antonanzas F, et al. Health economic evaluation of gene replacement therapies: methodological issues and recommendations. J Mark Access Health Policy. 2020;8(1):1822666.

62. Walton M, Sharif S, Simmonds M, Claxton L, Hodgson R. Tisagenlecleucel for the treatment of relapsed or refractory B-cell acute lymphoblastic leukaemia in people aged up to 25 Years: An evidence review group perspective of a NICE single technology appraisal. Pharmacoeconomics. 2019;37(10):1209-1217.

63. Kahan BC, Cro S, Doré CJ, Bratton DJ, Rehal S, Maskell NA, et al. Reducing bias in open-label trials where blinded outcome assessment is not feasible: strategies from two randomised trials. Trials. 2014;15:456.

64. Cox E, Wade R, Peron M, Dietz KC, Eastwood A, Palmer S, et al. The clinical and cost effectiveness of inotuzumab ozogamicin for the treatment of adult relapsed or refractory B-cell acute lymphoblastic leukaemia: An evidence review group evaluation of a NICE single technology appraisal. Pharmacoeconomics. 2019;37(9):1081-1091.

65. Thorlund K, Zafari Z, Druyts E, Mills EJ, Sadatsafavi M. The impact of incorporating Bayesian network meta-analysis in cost-effectiveness analysis - a case study of pharmacotherapies for moderate to severe COPD. Cost Eff Resour Alloc. 2014;12(1):8.

66. Rinciog C, Watkins M, Chang S, Maher TM, LeReun C, Esser D, et al. A cost-effectiveness analysis of nintedanib in idiopathic pulmonary fibrosis in the UK. Pharmacoeconomics. 2017;35(4):479-491.

67. Ramaekers BLT, Wolff RF, Pouwels X, Oosterhoff M, Van Giessen A, Worthy G, et al. Ixekizumab for treating moderate-to-severe plaque psoriasis: An evidence review group perspective of a NICE single technology appraisal. Pharmacoeconomics. 2018;36(8):917-927.

68. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. Med Decis Making. 1989;9:142-9.

69. National Institute for Health and Care Excellence [Internet]. Guide to the methods of technology appraisal. Available from: https://www.nice.org.uk/process/pmg9/chapter/foreword. Accessed 2022 July 16.

70. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, present and future. Appl Health Econ Health Policy. 2017;15:127-37.

71. Christiansen ASJ, Møller MLS, Kronborg C, Haugan KJ, Køber L, Højberg S, et al. Comparison of the three-level and the five-level versions of the EQ-5D. Eur J Health Econ. 2021;22:621-8.

72. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85:365-76.

73. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: Development and validation of the general measure. J Clin Oncol. 1993;11:570-9.

74. National Institute for Health and Care Excellence [Internet]. Trastuzumab emtansine for adjuvant treatment of HER2-positive early breast cancer. Technology appraisal guidance [TA632]. Available from: https://www.nice.org.uk/guidance/ta632. Accessed 2022 July 16.

75. Tappenden P, Chilcott JB. Avoiding and identifying errors and other threats to the credibility of health economic models. Pharmacoeconomics. 2014;32:967-979.

76. National Institute for Health and Care Excellence [Internet]. Necitumumab for untreated advanced or metastatic squamous non-small-cell lung cancer. Technology appraisal guidance [TA411]. Available from: https://www.nice.org.uk/guidance/ta411. Accessed 2022 July 16.

77. National Institute for Health and Care Excellence [Internet]. Cabozantinib for previously treated advanced renal cell carcinoma. Technology appraisal guidance [TA463]. Available from: https://www.nice.org.uk/guidance/ta463. Accessed 2022 July 16.

78. National Institute for Health and Care Excellence [Internet]. Ribociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer. Technology appraisal guidance [TA496]. Available from: https://www.nice.org.uk/guidance/ta496. Accessed 2022 July 16.

79. National Institute for Health and Care Excellence [Internet]. Darolutamide with androgen deprivation therapy for treating hormone-relapsed non-metastatic prostate cancer. Technology appraisal guidance [TA660]. Available from: https://www.nice.org.uk/guidance/ta660. Accessed 2022 July 16.

80. National Institute for Health and Care Excellence [Internet]. Reviewing our methods for health technology evaluation: Consultation. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/chte-methods-consultation. Accessed 2022 July 16.

81. National Institute for Health and Care Excellence [Internet]. Aflibercept in combination with irinotecan and fluorouracil-based therapy for treating metastatic colorectal cancer that has progressed following prior oxaliplatin-based chemotherapy. Technology appraisal guidance [TA307]. Available from: https://www.nice.org.uk/guidance/ta307. Accessed 2022 July 16.

82. National Institute for Health and Care Excellence [Internet]. Polatuzumab vedotin with rituximab and bendamustine for treating relapsed or refractory diffuse large B-cell lymphoma. Technology appraisal guidance [TA649]. Available from: https://www.nice.org.uk/guidance/ta649. Accessed 2022 July 16.

83. National Institute for Health and Care Excellence [Internet]. Larotrectinib for treating NTRK fusion-positive solid tumours. Technology appraisal guidance [TA630]. Available from: https://www.nice.org.uk/guidance/ta630. Accessed 2022 July 16.

84. National Institute for Health and Care Excellence [Internet]. Neratinib for extended adjuvant treatment of hormone receptor-positive, HER2-positive early stage breast cancer after adjuvant trastuzumab. Technology appraisal guidance [TA612]. Available from: https://www.nice.org.uk/guidance/ta612. Accessed 2022 July 16.

85. National Institute for Health and Care Excellence [Internet]. Atezolizumab with nab-paclitaxel for untreated PD-L1-positive, locally advanced or metastatic, triple-negative breast cancer. Technology appraisal guidance [TA639]. Available from: https://www.nice.org.uk/guidance/ta639. Accessed 2022 July 16.

86. National Institute for Health and Care Excellence [Internet]. Bortezomib for previously untreated mantle cell lymphoma. Technology appraisal guidance [TA370]. Available from: https://www.nice.org.uk/guidance/ta370. Accessed 2022 July 16.

87. National Institute for Health and Care Excellence [Internet]. Enzalutamide for hormone-relapsed non-metastatic prostate cancer. Technology appraisal guidance [TA580]. Available from: https://www.nice.org.uk/guidance/ta580. Accessed 2022 July 16.

88. National Institute of Public Health [Internet]. [C2H1902] Tisagenlecleucel (Kymriah). Available at https://c2h.niph.go.jp/en/results/C2H1902.html. Accessed 2022 July 30.

89. National Institute of Public Health [Internet]. [C2H1903] Ravulizumab (Ultomiris). Available at https://c2h.niph.go.jp/en/results/C2H1903.html. Accessed 2022 July 30.

90. National Institute of Public Health [Internet]. [C2H1906] Ivabradine (Coralan). Available at https://c2h.niph.go.jp/en/results/C2H1906.html. Accessed 2022 July 30.

91. European Medicines Agency [Internet]. Parallel joint scientific consultation with regulators and health technology assessment bodies. Available at: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance/parallel-joint-scientific-consultation-regulators-health-technology-assessment-bodies. Accessed 2022 July 30.

92. National Institute for Health and Care Excellence [Internet]. Scientific advice. Available at https://www.nice.org.the UK/about/what-we-do/life-sciences/scientific-advice. Accessed 2022 July 30.

Acknowledgements

First of all, I would like to express my deepest gratitude to Professor Mamoru Narukawa for his countless valuable guidance and suggestions throughout my research. For this dissertation, I would like to thank the members of the review committee: Professor Takeshi Hirayama, Professor Mitsuo Tanabe and Professor Yuji Yoshiyama for their time, interest and insightful comments. I am also indebted to Ms. Yuko Minami for providing kind administrative supports. I would like to thank my laboratory colleagues for their warmth and advice throughout of my research.

I would like to show my greatest appreciation to my supervisors at Daiichi Sankyo Company, Limited, Dr. Kosaku Fujiwara and Mr. Sumiaki Fujitani who made my doctoral research possible.

Lastly, I would also like to express my gratitude to my family for their moral support and warm encouragement.

Appendix

List of Appendix Tables

Appendix Table 1 Multivariable Analysis of Appraised Medicine Characteristics Associated with Key Periods (Excluding One Outlier)

Characteristics (no. of appraisals)	MA to FAD		VAL to FS*		FS to FAD	
	Unstandardized coefficients month, (95%CI)	P value	Unstandardized coefficients month, (95%CI)	P value	Unstandardized coefficients month, (95%CI)	P value
Completion year						
2016-2018 (73)	Reference		Reference		Reference	
2019-2020 (42)	0.332 (-1.116, 1.780)	0.650	-0.705 (-3.164, 1.754)	0.571	1.221 (-0.477, 2.920)	0.157
Application type						
Initial (68)	Reference		Reference		Reference	
Extension (47)	-0.604 (-2.068, 0.859)	0.415	-5.922 (-8.479, -3.364)	< 0.001	-0.053 (-1.770, 1.664)	0.952
Previous appraisal						
1 appraisal added	-0.017 (-0.074, 0.041)	0.566	-0.007 (-0.103, 0.090)	0.894	-0.033 (-0.101, 0.035)	0.338
Cancer medicine						
No (44)	Reference		Reference		Reference	
Yes (71)	-0.031 (-1.666, 1.603)	0.970	-3.830 (-6.673, -0.988)	0.009	2.321 (0.403, 4.238)	0.018
OMP						
No (91)	Reference		Reference		Reference	
Yes (24)	2.508 (0.690, 4.327)	0.007	-0.858 (-3.880, 2.164)	0.574	2.761 (0.628, 4.894)	0.012
AA						
No (110)	Reference		Reference		Reference	
Yes (5)	-0.482 (-3.877, 2.913)	0.779	-7.449 (-13.093, -1.806)	0.010	-1.604 (-5.587, 2.379)	0.426

* 108 appraisals were available, because there were some missing entries in the European Medicines Agency's validation date. Among the 108 appraisals, the 66 were completed from 2016 to 2018, the 68 were initial application, the 67 were cancer medicines, the 24 were OMPs, and the 5 were granted accelerated assessment.
Abbreviation: AA, accelerated assessment; CI, confidence interval; FAD, final appraisal determination; FS, final scope; MA, marketing authorization; OMP, orphan medicinal product; VAL, validation of marketing authorization application

Appendix Table 2　Multivariable Analysis of Factors Influencing the FS to FAD Periods (Including ICER Gap)

Factors (no. of appraisals)	Multivariable analysis (N=84)	
	Unstandardized coefficients month, (95%CI)	P value
Factors regarding cost-effectiveness analyses		
No. of comparators in the FS		
≤ 2 (49)	Reference	
> 2 (67)	-1.659 (-3.438, 0.120)	0.067
ICER gap between the manufacture and the ERG		
≤ 20,000 pound/QALY (56)	Reference	
> 20,000 pound/QALY (37)	-1.491 (-0.264, 3.246)	0.095
Factors regarding clinical trials included in cost-effectiveness analyses		
Time to approval		
≤ 300 days (53)	Reference	
> 300 days (53)	0.937 (-0.823, 2.697)	0.292
Double-blinded randomized control trial		
No (53)	Reference	
Yes (63)	-1.878 (-3.799, 0.042)	0.055
Comparators		
Not specified in the FS (66)	Reference	
Specified in the FS (50)	-1.931 (-3.953, 0.092)	0.061

Abbreviation: ERG, evidence review group; FAD, final appraisal determination; FS, final scope; ICER; incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

Appendix Table 3　Time from Validation of MA Application by the EMA to Approval in Case of OMPs

Time to approval, n (%) *	Median (day)	Range (day)
OMP, 20 (29)	333	204-666
Non-OMP, 49 (71)	344	154-1604

* Only appraisals for initial application were included.
Abbreviation: MA; marketing authorization, OMP: orphan medicinal product.