LSE Research Online

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Karin H. Cerri, Martin Knapp and Jose-Luis Fernandez

Decision making by NICE: examining the influences of evidence, process and context

Article (Published version) (Refereed)

Original citation:
Cerri, Karin H., Knapp, Martin and Fernández, José-Luis (2014) *Decision making by NICE: examining the influences of evidence, process and context.* Health Economics, Policy and Law, 9 (2). pp. 119-141. ISSN 1744-1331

DOI: 10.1017/S1744133113000030

© 2013 Cambridge University Press © CC BY 3.0

This version available at: http://eprints.lse.ac.uk/50406/

Available in LSE Research Online: June 2014

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

http://eprints.lse.ac.uk

Health Economics, Policy and Law

http://journals.cambridge.org/HEP

Additional services for ***Health Economics, Policy and Law:***

Email alerts: Click here
Subscriptions: Click here
Commercial reprints: Click here
Terms of use : Click here

Decision making by NICE: examining the influences of evidence, process and context

Karin H. Cerri, Martin Knapp and Jose-Luis Fernandez

Health Economics, Policy and Law / Volume 9 / Issue 02 / April 2014, pp 119 - 141
DOI: 10.1017/S1744133113000030, Published online: 21 May 2013

Link to this article: http://journals.cambridge.org/abstract_S1744133113000030

How to cite this article:
Karin H. Cerri, Martin Knapp and Jose-Luis Fernandez (2014). Decision making by NICE: examining the influences of evidence, process and context . Health Economics, Policy and Law, 9, pp 119-141 doi:10.1017/S1744133113000030

Request Permissions : Click here

CAMBRIDGE JOURNALS

Downloaded from http://journals.cambridge.org/HEP, IP address: 158.143.197.191 on 27 Jun 2014

Health Economics, Policy and Law (2014), **9**, 119–141 © Cambridge University Press 2013. The online version of this article is published within an Open Access environment subject to the conditions of the Creative Commons Attribution licence http://creativecommons.org/licenses/by/3.0/
doi:10.1017/S1744133113000030

Decision making by NICE: examining the influences of evidence, process and context

KARIN H. CERRI*

Personal Social Services Research Unit, London School of Economics and Political Science, London, UK

MARTIN KNAPP

Professor of Social Policy, Personal Social Services Research Unit, London School of Economics and Political Science, London, UK

JOSÉ-LUIS FERNÁNDEZ‡

Principal Research Fellow, Personal Social Services Research Unit, London School of Economics and Political Science, London, UK

Abstract: The National Institute for Health and Clinical Excellence (NICE) provides guidance to the National Health Service (NHS) in England and Wales on funding and use of new technologies. This study examined the impact of evidence, process and context factors on NICE decisions in 2004–2009. A data set of NICE decisions pertaining to pharmaceutical technologies was created, including 32 variables extracted from published information. A three-category outcome variable was used, defined as the decision to 'recommend', 'restrict' or 'not recommend' a technology. With multinomial logistic regression, the relative contribution of explanatory variables on NICE decisions was assessed. A total of 65 technology appraisals (118 technologies) were analysed. Of the technologies, 27% were recommended, 58% were restricted and 14% were not recommended by NICE for NHS funding. The multinomial model showed significant associations ($p \leq 0.10$) between NICE outcome and four variables: (i) demonstration of statistical superiority of the primary endpoint in clinical trials by the appraised technology; (ii) the incremental cost-effectiveness ratio (ICER); (iii) the number of pharmaceuticals appraised within the same appraisal; and (iv) the appraisal year. Results confirm the value of a comprehensive and multivariate approach to understanding NICE decision making. New factors affecting NICE decision making were identified, including the effect of clinical superiority, and the effect of process and socio-economic factors.

Received 3 October 2012; revised 3 October 2012; accepted 15 January 2013; first published online 21 May 2013

*Correspondence to: Dr Karin H. Cerri, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. Email: k.h.cerri@lse.ac.uk

‡The original version of this article was published with the incorrect author name. A notice detailing this has been published and the error rectified in the online and print PDF and HTML copies.

Introduction

Health Technology Assessment (HTA) decisions represent a key point within the complex decision making process that governs funding and access for pharmaceuticals in many health care systems. HTA decisions are of interest because of their implications for patients, health care providers, funding bodies, technology manufacturers and health policy makers. Successful implementation of HTA requires multi-disciplinary assessment of the range of social, economic, clinical and health care system organisational consequences stemming from the introduction of a new technology relative to the existing standards of care (Henshall *et al.*, 1997; Velasco Garrido and Busse, 2005; HTAI, 2007; EUnetHTA, 2011). Therefore, it is important to understand what factors are in fact driving HTA decisions.

Established in 1999, the National Institute for Health and Clinical Excellence (NICE) is responsible for providing guidance to the National Health Service (NHS) in England and Wales on the funding of new technologies and their use (Chalkidou, 2009). There are four types of guidance that NICE provides including: clinical guidelines, public health guidelines, guidelines on safety and efficacy of interventional procedures and recommendations on 'health technologies' (such as surgical interventions and pharmaceuticals; NICE, 2008b). One of the key rationales for setting up NICE was to help tackle the geographic inequality in access to technology or the phenomenon more frequently referred to as 'postcode prescribing' (Summerhayes and Catchpole, 2006). Since 2002, NICE's recommendations have been mandatory and NHS organisations have had to comply, usually within three months.[1]

Within NICE, the Centre for Health Technology Evaluation focuses on issuing guidance on the appropriate use and funding of technologies. The NICE HTA process involves a panel of clinical, academic and industry stakeholders and the general public (NICE, 2008a). Appraisals by NICE are governed by the use of established standard methodology for the evaluation of clinical and economic characteristics of the technology (NICE, 2008a; Littlejohns *et al.*, 2009). As outlined in its methods guide (NICE, 2008a), a range of clinical criteria are evaluated during a NICE appraisal and the submission of cost-utility evidence is an integral part of the process. In addition to economic and clinical criteria, the patients' perspective and patient evidence, as well as the perspective of carers and other members of the NHS, are taken into consideration via the consultation and stakeholder submission processes. NICE operates a process that emphasises the roles of social values in decision making (Littlejohns *et al.*, 2009), especially during appraisal of clinical effectiveness and cost-effectiveness evidence (NICE, 2008b). Social value principles considered by NICE include a focus on promoting equality and key ethical principles in decision making, including the notion of 'procedural justice', a principle through which NICE is committed to "ensuring

1 However, the new Conservative–Liberal Democrat government elected in May 2010 have indicated that they will change NICE's role and its coverage recommendations may only have advisory status.

that the processes by which healthcare decisions are reached are transparent, and that the reasons for the decisions are explicit" (NICE, 2008b, p. 12).

Given the multidimensional nature of NICE decision making and the high degree of stakeholder involvement, it is hypothesised that decisions are influenced not only by the evidence supporting the technology, but also by the assessment processes used and the context in which they operate. The overall objective was to examine the impact of evidence, process and context factors on decisions made by NICE to recommend, restrict or not recommend new technologies for use in England and Wales.

Method

Hypothesised drivers of HTA decision making

It was hypothesised, based on the literature available, that HTA decisions were driven by the HTA decision-making process, the evidence considered within that process and by the socio-economic and political context in which those decisions were made (Figure 1). Research on the impact of evidence on the HTA decision has shown that the evidence related to the medicine or other technology under review (whether clinical, economic or otherwise) can influence HTA decisions [see for example, (Devlin and Parkin, 2004; Dakin *et al.*, 2006; Clement *et al.*, 2009; Mason and Drummond, 2009)]. For instance, previous studies of NICE decision making have highlighted the significant impact of clinical evidence as well as the incremental cost-effectiveness ratio (ICER) (Devlin and Parkin, 2004; Dakin *et al.*, 2006). The literature has also raised questions about the degree to which NICE decisions may be influenced by the nature of the disease for which the technology is indicated. There has been particular focus on cancer therapies and the degree to which there may be flexibility in the criteria used to appraise technologies, given the severity of the disease and clinical needs. A review of NICE appraisals related to cancer technologies suggested that there was an increase in non-recommendation over time, although not statistically significant

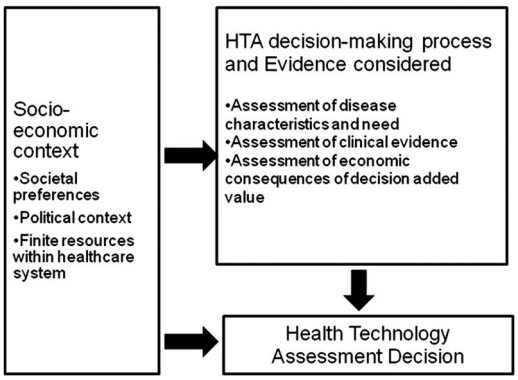

Figure 1. HTA decision making: hypothesised drivers. HTA = health technology assessment

(Mason and Drummond, 2009). In contrast, NICE has also been characterised as being more flexible in its application of a cost-effectiveness threshold to cancer therapies (NICE, 2009). This presents an opportunity within this analysis to further investigate as to what extent disease characteristics influence NICE decision making, when taking into account other confounding factors.

The literature examining the HTA appraisal process provided insights into a number of process-related factors that can potentially influence HTA decisions (e.g. Al *et al.*, 2004; Menon *et al.*, 2005; OECD, 2005; Vuorenkoski *et al.*, 2008). Within NICE decision making, the use of multiple technology appraisals (MTA) or single technology appraisals (STA) process has been hypothesised to affect decision making (Barbieri *et al.*, 2009). MTA and STA are two appraisal processes used by NICE to issue guidance on technologies. The MTA process includes a third-party review and analysis of the clinical and economic evidence, whereas the STA process relies on manufacturer-provided information (Barbieri *et al.*, 2009). However, the lack of an adequate sample size and the use of descriptive rather than multivariate analyses impeded the authors from examining the role of appraisal process on NICE decisions. Therefore, it was identified as an area of interest for this analysis. Finally, reference in the literature was made to the impact of the overall health care and welfare characteristics on HTA decision making, such as health care spending per capita, societal willingness to pay, the structure of the health care system, as well as ethical and social considerations (e.g. Ross, 1995; Owens, 1998; Buxton, 2005; Packer *et al.*, 2006; Bryan *et al.*, 2007). In their model of NICE decision making between 2000 and 2003, Dakin *et al.*, (2006) found a statistically significant effect of the time of appraisal on outcome; therefore, it was of interest to establish whether this effect was observed in later appraisals.

Thus, based on the analytical model, the hypotheses tested in this analysis of NICE decision making were:

- Whether the ICER significantly affects NICE decision making; specifically, whether an increase in the ICER decreases the odds of recommendation.
- Whether technologies for cancer therapies increase the odds of recommendation relative to non-recommendation.
- Whether the use of an MTA process increases the odds of restriction relative to recommendation, compared with the STA process.
- Whether non-recommendations and restrictions are increasing over time relative to recommendations.

In testing these hypotheses, the study also sought to make comparisons with previously published models of NICE decision making (Devlin and Parkin, 2004; Dakin *et al.*, 2006).

Sample

The pharmaceutical technology appraisals performed by NICE formed the basis for the sample included in this analysis. The composition of the sample was determined

through the following inclusion and exclusion criteria. The sample included all pharmaceutical technology appraisals (as opposed to medical devices or other interventions) performed during the period 2004–2009, indicated for an adult population. A five-year time horizon was used to capture a sufficient number of appraisals for analysis. Technology appraisals were excluded from the analysis for any of the following reasons: (i) they focused on a non-adult population (aged <18 years); (ii) the appraised technologies were non-pharmaceutical interventions; (iii) marketing authorisation was withdrawn; or (iv) the full guidance was not available.

Outcome variable

The analysis was designed to reflect as closely as possible the way NICE makes its decisions. To this end, a three-category outcome variable was used where technologies were recommended, restricted or not recommended. A three-category outcome variable was preferred over a binary outcome variable (i.e. covered vs not covered), as it better reflected the multiple coverage options available to NICE. Ranking of the three potential outcomes was not implemented as ranking of outcomes can vary according to the perspective adopted.

NICE guidance (summarised in section 1 of each guidance) indicates whether an intervention should be recommended or not for use in the NHS. A medication was considered as not recommended for use by NICE guidance if the words 'not recommended' were stated in section 1 of the guidance. To distinguish between recommended and restricted interventions, decision rules were developed to help classify which of the recommendations issued from NICE guidance were for routine use, and those that were for restricted use. Specifically, the Raftery classification was utilised to distinguish between restriction and recommendation (Raftery, 2006). Where a recommendation was made for a technology to be used in a population identical to its licensed indication, it was considered to be 'recommended'. Where a recommendation contained one of the following provisos in relation to the technology, it was considered to be 'restricted': (i) it should be used in a sub-population of its licensed indication; (ii) it should be used in a second line or higher line of therapy; (iii) it required monitoring; (iv) it should be used at the lowest acquisition cost; or (v) it required prescription by a specialist (Raftery, 2006).

Explanatory variables

In line with the hypothesised drivers of HTA decision making, 32 variables were defined, including those relating to (i) the clinical and economic characteristics of the technology under appraisal [such as characteristics of randomised controlled trials (RCTs), use of observational data, ICER reported], (ii) the processes used to come to a decision (such as use of STA or MTA) and (iii) the socio-economic context in which these decisions were made [including percentage Gross Domestic Product (GDP) spent on health care, year of appraisal and if disease was identified as a priority]. These are shown in Table 1.

Table 1. NICE data set includes variables, definitions, data extraction rules and data sources

No.	Variable	Definition
1	NoRCT	The number of distinct RCTs that provide data related to the therapeutic indication under evaluation. Excluded: studies that are single arm, that have no randomisation or that are non-interventional. (1)
2	RCTsize	Mean number of patients per RCT. (1)
3	RCTduration	Mean number of weeks that data are collected on patients who entered the RCTs (see variable no. 1). (1)
4	RCTsignificant	Presence of statistically significant superiority of technology vs comparator for primary endpoint(s). If more than one RCT was considered, and the technology showed statistically significant superiority in one trial, but not in another, the results were considered to be 'inconsistent' and classified as such. RCTs designed as 'non-inferiority' studies were classified as not showing any statistically significant superiority (i.e. 'no'). (1)
5	RCTcomparator	Percentage of RCTs where an active comparator was used. (1)
6	ObsStudies	Number of observational studies providing information to support the study pharmaceutical. Observational studies in this circumstance are defined as studies that are non-interventional (i.e. do not explicitly request the patient to take particular medication or the physician to follow a particular protocol). (1)
7	Priority	This variable aims to capture the health policy context in which the decision is made by capturing whether the pharmaceutical in question is linked to a disease area that is prioritised by the Department of Health. Priority disease areas were identified by examining government plans/health documents that highlight national health care system focus. (2)
8	OrphanDesig	This variable captured information on whether or not the technology was recognised by the EMA as an orphan-designated medicine. (3)
9	Disease	The BNF categories were used to classify each technology into the corresponding therapeutic area. (4)
10	Eligiblepop	Reported number of patients eligible for treatment, defined by patient population for which the technology is indicated. (1, 3)
11	AlternativeTx	Captures whether an alternative treatment option was available. An alternative was considered to be available if comparators were clearly defined in the review by the HTA agency. An alternative was considered NOT to be available if it was stated as such in the appraisal, or if 'best supportive care' or 'palliative care' was specified as the comparator. (1)
12	CUA	Presence or absence of a cost-utility analysis. (1)
13	ICER	ICER (Cost per QALY) reported in the HTA dossier for base–case as accepted by the Appraisal Committee. This is defined as the ICER that is related to the recommendation. If more than one ICER is presented due to the recommendation covering more than one population, and then the ICER pertaining to the larger of the populations was reported. In NICE appraisals, if no ICER was reported by the Appraisal Committee, then the ICER reported by the Assessment Group was used. The ICER reported in the manufacturer submission was used if it was not stated in the guidance or the assessment group report. If the technology is reported as dominant or dominated, it was recorded as such in the data extraction sheet. (1)
14	MultipleCEA	Whether more than one cost-utility or cost-effectiveness model was considered during the appraisal. (1)
15	MultipleICERs	If response to variable 14 is 'yes', report range of base–case ICERs presented between the different models reported. The difference between the lowest and highest ICER will be calculated. (1)

Table 1. *(Continued)*

No.	Variable	Definition
16	ProbabilisticICER	This should be reported as the percentage probability of acceptance at the threshold used by the agency. The probability of medication to be cost-effective at a 30,000 GBP threshold was reported. (1)
17	UnivariateICER	This should be reported as the range of ICERs (min–max) resulting from univariate sensitivity on the base–case. (1)
18	NonCUA	Indicates whether non-cost per QALY economic analyses were submitted and reviewed. (1)
19	BudgetImp	Estimated annual incremental budgetary impact of introducing new medication into the current treatment setting, if the pharmaceutical were to be introduced without any restriction. Pharmaceutical cost only (per year). This is the potential budget impact where a recommendation for use is the total indicated population to be granted. (1)
20	PatientSub	A patient submission was considered to have been included as part of the appraisal process if a submission from a patient group was posted on the webpage pertaining to the guidance. (5)
21	NoDecMakers	Captures the number of decision makers accountable for guidance issued, as reported. (1)
22	CEAprocess	Captures whether cost-effectiveness is a component of the decision-making process or not. If cost-effectiveness analysis is a formal part of the appraisal process, this variable was marked as 'yes'. (1, 6)
23	BudgetImpProcess	Captures whether budget impact considerations are part of decision-making process. (1, 6)
24	Price	Captures whether the price of the technology under appraisal was known during the assessment. (1)
25	MTAvsSTA	Records whether guidance was issued after the appraisal of an individual technology following the STA or MTA process. As the STA process was introduced in 2006, those technologies appraised before this year were automatically considered as having been appraised via an MTA. (1)
26	NoTech	This variable captures the number of technologies appraised simultaneously in the appraisal. (1)
27	Accountability	The HTA agency was examined to assess whether the agency making the funding decisions is also accountable for the pharmaceutical budget or not. (7)
28	Date	Year when decision was issued. (1)
29	PopulationSize	Estimate of population size within remit of the agency performing the evaluation. (8)
30	HealthcareGDP	Percentage of GDP spent on health care during year of decision. (9)
31	HealthExp	Health care budget spent on pharmaceuticals per patient per year, during the same year in which the appraisal was published. (10)
32	Election	This variable captures whether the decision was made within an election year. An election year was defined as a year in which either national government or regional elections took place. (11)

RCT = Randomised Controlled Trials; EMA = European Medicines Agency; BNF = British National Formulary; HTA = health technology assessment; CUA = cost-utility analysis; ICER = incremental cost-effectiveness ratio; NICE = National Institute for Health and Clinical Excellence; MTA = multiple technology appraisals; GDP = gross domestic product; STA = single technology appraisals; TAR = Technical Appraisal Report.
(1) TAR, sections 3–4; (2) Department of Health (2002, 2006a, 2006b, 2007); (3) EMA (2011); (4) BNF, Joint Formulary Committee (2010); (5) Section on NICE appraisal webpage (nice.org.uk) describing the history of the appraisal; (6) NICE (2009); (7) Sorenson *et al.* (2008); (8) National Office for Statistics (2009); (9) OECD (2006); (10) Association of the British Pharmaceutical Industry (ABPI) (2010); (11) BBC (2005).

Data set

A database of information pertaining to NICE decisions and explanatory variables of interest was developed. To create this database, several steps were implemented. Publicly available sources of data containing information on the variables of interest were identified, and a data extraction form was developed to extract information from the different appraisals in a way that was as transparent, reproducible and consistent as possible. Finally, the resulting extracted data were coded and prepared for analysis. The variable definitions and data sources used are shown in Table 1.

Statistical analyses

A multinomial logit regression was used in the analysis to model the probabilities associated with the three types of technology appraisal outcomes (recommended, restricted or not recommended). We selected the 'recommended' outcome as the referent category in the analysis.

The objective of the analysis was to identify, ceteris paribus, the effect of a range of factors potentially associated with NICE appraisal decisions, and to assess which combination of factors best explains the pattern of NICE decisions. Given the wide range of factors considered in the analysis (see Table 1), a process was developed to determine which explanatory variables would appear in the final specification of the model. This process involved the following steps:

- First, bivariate regression models were run to ascertain the degree of correlation between individual explanatory variables and appraisal decisions.
- On the basis of these models, a subset of indicators was selected, which included those variables that showed at least moderate significance levels (indicators with a p-value below 0.25). A preliminary model was estimated including these indicators.
- The model was reduced by removing those variables with significance levels above the 0.10 threshold. To guarantee its stability, this 'base' model was re-estimated by sequentially removing one variable at a time and verifying the stability of the effects on the coefficient and significance level of the remaining estimates.
- The model was subsequently tested through alternative model specifications to examine its robustness and to assess the sensitivity of the results to different assumptions.
- As a final step, the base–case model results were presented to the representatives of NICE to seek feedback on the variables identified within the base–case model, the coefficient and level of significance to assess the validity of the model.

The application of the model-specification process outlined above facilitated the interpretation of the results of the models, while allowing the analysis to explore the impact of the wide range of indicators collected in the study.

Table 2. Description of missing data within NICE data set 2004–2009

Proportion of incomplete entries (%)	8%
Missing entries by outcome variable	Recommended: 5%
	Restricted: 9%
	Not recommended: 9%
List variables with missing values (top three variables)	Univariate uncertainty estimates around ICER
	Probabilistic sensitivity analysis
Missing entries per appraisal (mean and range)	3 (range: 0–9)

NICE = National Institute for Health and Clinical Excellence; ICER = incremental cost-effectiveness ratio.

Although significant effort was made to identify the information relevant to the variables of interest, a limited proportion of data could not be found and was therefore missing (see Table 2). To maximise the sample size, imputation techniques were used to estimate entries for missing values. Missing values were replaced with regression imputation estimates using the 'impute' command in STATA software [Intercooled (IC) Stata version 10.1]. The imputed values obtained were then checked manually to ensure their face validity. In addition, dummy variables were created to identify observations with missing data to test in the regression models whether the lack of data was significantly associated with differences in the outcome variable.

A series of sensitivity analyses were performed on the base–case regression model to help evaluate the robustness of the results. The sensitivity analyses included: (i) examining the impact of a binary rather than a three-category outcome variable; (ii) restricting the base–case analysis to complete observations, thus excluding observations with imputed values; and (iii) estimating the model assuming ordinal properties of the outcome variable.

Descriptive statistics were calculated for each extracted variable, stratified by outcome group (recommended, restricted or not recommended). For categorical variables, the χ^2-test was conducted across the three outcomes, at a 0.05 significance level. With regard to non-categorical variables, analysis of variance and Kruskal–Wallis tests were carried out, at a level of significance of 0.05. Statistical analyses were performed using IC Stata (version 10.1 2009).

Results

Of 99 NICE technology appraisals, 65 were included in the analysis, representing 118 technologies. A total of 34 technology appraisals were excluded from the analysis for the following reasons: (i) they focused on a non-adult population ($n = 5$); (ii) they appraised non-pharmaceutical interventions ($n = 23$); (iii) marketing authorisation was withdrawn ($n = 2$); or (iv) full guidance was not available ($n = 4$). The majority of decisions restricted funding for the appraised technology (58%), 27% were recommended, whereas the least common decision was non-recommendation (14% of technologies).

Univariate analysis

Of the 32 explanatory variables examined, descriptive analysis suggested that a subset of 20 variables may play an important role in determining NICE decision making (Table 3). Six variables related to the clinical evidence supporting the technology under evaluation. The first three variables described the nature of the randomised clinical trial data available in terms of the number of trials, sample size and trial duration. Interventions recommended for use or for restricted use were supported by a higher number of RCTs (mean of seven and eight trials, respectively), compared with interventions not recommended for use (mean three trials, $p < 0.05$). The mean size of the patient sample included in RCTs was higher for those interventions recommended by NICE (mean = 1765 patients), compared with those interventions restricted or not recommended by NICE (1044 and 1154 patients, respectively, $p < 0.05$). The mean trial duration across the three outcome groups was 96, 66 and 82 weeks, respectively (ns). Of the recommended interventions, 59% demonstrated statistically significant superiority, as opposed to 29% of restricted and 35% of not recommended interventions ($p < 0.05$). Recommended interventions were more likely to have active comparators within the RCTs (63%) than interventions that were restricted or not recommended (40%, 44% of RCTs with active comparators, respectively, $p < 0.05$). Very little observational data were referred to in NICE appraisals (the mean across appraisals was 0.6 observational studies, not significant across outcomes).

The effect of economic evidence on NICE decision was significant. The majority of NICE decisions were backed by cost-utility evidence (95%). In 65% of decisions, multiple economic models were considered in the appraisal process and the use of multiple models was consistent across outcome categories. For the interventions supported by a cost-utility analysis, the ICER was significantly different ($p < 0.001$) between the recommended interventions (mean ICER of £17,782), compared with the restricted interventions (mean ICER of £24,867) and the interventions not recommended for use (mean ICER of £99,239). The probability of the ICER remaining below £30,000 was 61% for recommended interventions, 41% for restricted interventions and 8% for interventions not recommended for use ($p < 0.01$). With regard to univariate sensitivity analyses, the range of uncertainty was smallest for those interventions that were restricted for use (£19,747–£ 57,146), the widest range of uncertainty was observed for interventions not recommended for use (£92,379–£731,151; $p < 0.05$). Although the ICER is an important variable in NICE decision making, there are exceptions to its usefulness as a predictor of HTA decisions – there are recommended technologies with very high base–case ICERs, as there are technologies with lower base–case ICERs that are not recommended (Figure 2). This suggests the importance of examining a combination of factors to explain NICE decision making.

The effect of the appraisal process was found to be significant. On average, 25% of decisions followed the STA process – and this varied across outcomes: 41% of

Table 3. NICE decisions 2004–2009: mean values for explanatory factors, stratified by decision (lower, upper 95% CI)

	Total ($n = 118$)	Recommended ($n = 32$)	Restricted ($n = 69$)	Not Recommended ($n = 17$)	p-value	
NoRCT	6.7 (5.2, 8.3)	6.8 (2.1, 11.4)	7.6 (5.9, 9.2)	3.4 (2.1, 4.8)	<0.05	3
RCTsize	1249 (807, 1691)	1765 (804, 2727)	1044 (442, 1645)	1154 (464, 1844)	<0.05	3
RCTduration	76.2 (63.5, 88.9)	95.9 (65.9, 125.8)	65.8 (51.5, 80.2)	81.9 (41.5, 122.3)	ns	1
RCTsignificant						
Yes	38% (29%, 47%)	59% (41%, 77%)	29% (18%, 40%)	35% (10%, 61%)	<0.05	2
No	15% (9%, 22%)	6% (−3%, 15%)	17% (8%, 27%)	24% (1%, 46%)	ns	2
Inconsistent	43% (34%, 52%)	31% (14%, 48%)	51% (39%, 63%)	35% (10%, 61%)	ns	2
RCTcomparator	47% (39%, 55%)	63% (46%, 80%)	40% (30%, 50%)	44% (18%, 69%)	<0.10	3
ObsStudies	0.6 (0.1, 1.1)	1.5 (−0.4, 3.4)	0.3 (0.1, 0.5)	0 (0,0)	ns	1
CUA	95% (91%, 99%)	100% (100%, 100%)	94% (89%, 100%)	88% (71%, 105%)	ns	2
ICER	£31,910 (£20,945, £42,874)	£17,782 (£11,066, £24,498)	£24,867 (£21,002, £28,731)	£99,239 (£11,882, £186,597)	<0.01	3
MultipleCEA	65% (56%, 74%)	69% (52%, 86%)	63% (51%, 75%)	67% (40%, 94%)	ns	2
MultipleICERs – low range	£13,318 (£10,511, £16,125)	£8607 (£4,551, £12,664)	£13,417 (£10,119, £16,714)	£21,441 (£9,663, £33,220)	<0.01	3
MultipleICERs – high range	£107,421 (£66,886, £147,956)	£83,666 (−£12,619, £179,951)	£129,690 (£85,202, £174,179)	£83,242 (£33,511, £132,974)	<0.01	3
Univariate ICER – low value	£25,417 (£6,412, £44,422)	£7,881 (£5,234, £10,527)	£19,747 (£4,460, £35,035)	£92,379 (−£58,911, £243,668)	<0.01	1
UnivariateICER –high value	£167,389 (£56,865, £277,913)	£113,286 (£1293, £225,279)	£57,146 (£29,962, £84,329)	£731,151 (−£83,313, £1,545,615)	<0.05	3
ProbabilisticICER	43% (34%, 52%)	61% (45%, 77%)	41% (29%, 53%)	8% (−1%, 17%)	<0.01	3
NonCUA	22% (14%, 30%)	22% (7%, 37%)	28% (17%, 38%)	0% (0%, 0%)	<0.10	2
BudgetImp (million)	£708 (£182, £1235)	£36 (£18, £54)	£828 (£51, £1606)	£1,632 (−£610, £3874)	<0.10	3
Eligiblepop	2,288,409 (1,189,422, 3,387,396)	392,063 (82,017, 702,109)	3,194,243 (1,419,388, 4,969,097)	2,270,955 (−290,109, 4,832,019)	<0.01	3
AlternativeTx	89% (83%, 95%)	88% (75%, 100%)	90% (83%, 97%)	88% (71%, 105%)	ns	2
PatientSub	87% (81%, 93%)	91% (80%, 101%)	84% (75%, 93%)	94% (82%, 107%)	ns	2
NoDecMakers	30 (28, 32)	28 (25, 31)	31 (29, 34)	28 (23, 33)	ns	1
CEAprocess	100% (100%, 100%)	100% (100%, 100%)	100% (100%, 100%)	100% (100%, 100%)	n/a	n/a
BudgetImpProcess	100% (100%, 100%)	100% (100%, 100%)	100% (100%, 100%)	100% (100%, 100%)	n/a	n/a
Price	100% (100%, 100%)	100% (100%, 100%)	100% (100%, 100%)	100% (100%, 100%)	n/a	n/a

Table 3. *(Continued)*

	Total ($n = 118$)	Recommended ($n = 32$)	Restricted ($n = 69$)	Not Recommended ($n = 17$)	p-value	
MTAvsSTA	25% (17%, 32%)	34% (17%, 52%)	16% (7%, 25%)	41% (15%, 67%)	<0.05	2
NoTech	2.8 (2.5, 3.2)	2.0 (1.5, 2.5)	3.4 (2.9, 3.9)	2.1 (1.5, 2.7)	<0.01	3
Accountability	0% (0%, 0%)	0% (0%, 0%)	0% (0%, 0%)	0% (0%, 0%)	n/a	n/a
Date	2007 (2006, 2007)	2006 (2006, 2007)	2007 (2006, 2007)	2007 (2007, 2008)	<0.05	1
PopulationSize (million)	53.90 (53.80, 54.00)	53.8 (53.7, 54)	53.9 (53.8, 54.1)	54.2 (54, 54.4)	<0.05	1
HealthcareGDP	8% (8%, 8%)	8% (8%, 8%)	8% (8%, 8%)	8% (8%, 8%)	<0.05	3
HealthExp	£175 (£173, £176)	£173 (£171, £175)	£174 (£173, £176)	£178 (£175, £180)	<0.05	1
Election	7% (2%, 11%)	6% (−3%, 15%)	7% (1%, 14%)	6% (−7%, 18%)	ns	2
Priority	56% (47%, 65%)	59% (41%, 77%)	57% (45%, 69%)	47% (21%, 74%)	ns	2
OrphanDesig	3% (0%, 5%)	3% (−3%, 9%)	3% (−1%, 7%)	0% (0%, 0%)	ns	2
Disease						
Cardiovascular system	10% (5%, 16%)	13% (0%, 25%)	12% (4%, 19%)	0% (0%, 0%)	ns	2
Central nervous system	15% (9%, 22%)	6% (−3%, 15%)	22% (12%, 32%)	6% (−7%, 18%)	<0.10	2
Ear, nose and oropharynx	0% (0%, 0%)	0% (0%, 0%)	0% (0%, 0%)	0% (0%, 0%)	ns	2
Endocrine system	1% (−1%, 3%)	0% (0%, 0%)	1% (−1%, 4%)	0% (0%, 0%)	ns	2
Eye	2% (−1%, 4%)	3% (−3%, 9%)	0% (0%, 0%)	6% (−7%, 18%)	ns	2
Gastro-intestinal system	2% (−1%, 4%)	0% (0%, 0%)	1% (−1%, 4%)	6% (−7%, 18%)	ns	2
Infections	12% (6%, 18%)	19% (4%, 33%)	7% (1%, 14%)	18% (−3%, 38%)	ns	2
Malignant disease and immunosuppression	31% (22%, 39%)	41% (23%, 59%)	23% (13%, 33%)	41% (15%, 67%)	ns	2
Musculoskeletal and joint diseases	19% (12%, 27%)	16% (2%, 29%)	22% (12%, 32%)	18% (−3%, 38%)	ns	2
Nutrition and blood	3% (0%, 5%)	0% (0%, 0%)	4% (−1%, 9%)	0% (0%, 0%)	ns	2
Obstetrics/gynaecology, & urinary-tract disorders	0% (0%, 0%)	0% (0%, 0%)	0% (0%, 0%)	0% (0%, 0%)	ns	2
Respiratory system	1% (−1%, 3%)	0% (0%, 0%)	1% (−1%, 4%)	0% (0%, 0%)	ns	2
Skin	5% (1%, 9%)	3% (−3%, 9%)	6% (0%, 11%)	6% (−7%, 18%)	ns	2

Test 1. Both ANOVA and Kruskal–Wallis test indicate similar level of statistical significance; Test 2. χ^2-test used as categorical variable; Test 3. Either ANOVA or Kruskal–Wallis test indicates statistical significance.
ns = not significant; n/a = not applicable.

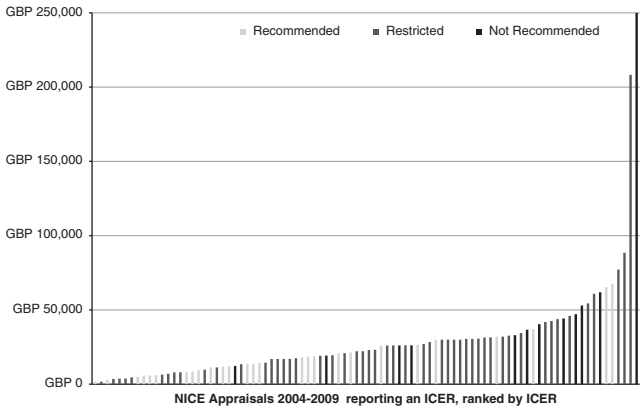

Figure 2. Base-case ICER(GBP) reported in NICE appraisals, by decision category (n = 90)

interventions that were not recommended for NHS funding followed the STA process, as opposed to 16% of restricted interventions and 34% of recommended interventions ($p = 0.031$). The number of technologies reviewed per appraisal was also statistically significantly different between outcome groups: restricted interventions were simultaneously appraised together with an average of 3.4 technologies compared with recommended and not recommended interventions (two technologies appraised simultaneously on average).

A range of socio-economic context variables varied significantly across NICE decision categories, including the year of appraisal, population size, health care expenditure as percentage of GDP and average health care expenditure per patient per year.

Multivariate analysis

The multinomial model showed significant associations ($p = 0.10$) between NICE outcome and four variables: (i) whether statistical superiority of the primary endpoint in the clinical trial was demonstrated by the appraised technology, (ii) the ICER, (iii) the number of pharmaceuticals appraised within the same appraisal and (iv) the year of the appraisal (Table 4). Specifically, demonstration by the technology under appraisal of statistically significant superior efficacy decreased the log odds of a restriction or non-recommendation ($p = 0.006$ and $p = 0.016$). A unit increase in the ICER increased the log odds of moving from a recommended to a restricted decision ($p = 0.009$) vs recommendation and a not recommended decision vs recommendation ($p < 0.001$). The number of technologies reviewed simultaneously within the same appraisal had a statistically significant impact on the decision between recommendation and restriction ($p < 0.05$) but not between recommendation and non-recommendation (ns).

Table 4. Multivariate analysis of NICE decisions 2004–2009: base–case model results ($n = 118$)

	Log odds	p-value	95% CI	
Restricted				
Clinical superiority demonstrated in RCT	−1.568	0.006	−2.679	−0.457
ICER	0.000048	0.009	0.000012	0.000085
Number of technologies appraised simultaneously	0.489	0.005	0.144	0.834
Year of appraisal	0.358	0.072	−0.032	0.748
Constant	−718.973	0.071	−1500.761	62.815
Not recommended				
Clinical superiority demonstrated in RCT	−2.01	0.02	−3.64	−0.38
ICER	0.000087	<0.001	0.000042	0.000131
Number of technologies appraised simultaneously	0.12	0.65	−0.39	0.63
Year of appraisal	0.67	0.03	0.07	1.27
Constant	−1351.14	0.03	−2552.90	−149.38

NICE = National Institute for Health and Clinical Excellence; CI = confidence interval; RCT = Randomised Controlled Trials; ICER = incremental cost-effectiveness ratio.
Recommended technologies are the reference case.
Multinomial logistic regression, pseudo-R^2: 0.26.

For every additional year, this increased the odds of a restriction (vs recommendation; p = 0.072) and the odds of a non-recommendation (vs recommendation; p = 0.028). When NICE decisions were regressed with these four variables, the resulting pseudo-R^2 was 0.26, suggesting that the four variables explained ~26% of the variability in NICE decisions.

Sensitivity analyses

A series of sensitivity analyses were performed on the base–case regression model to help evaluate the robustness of the results. The first sensitivity analysis assessed whether the impact of the explanatory variables on decision making varied if a binary outcome variable was utilised instead of the base–case three-category outcome variable. The logistic regression results using a binary outcome variable (covered vs not covered) no longer showed a predictive value for demonstration of clinical superiority (p = 0.657). However, in other aspects, the results are similar to the base–case analysis in that they confirm the impact of the ICER, number of pharmaceuticals appraised and the year of appraisal on NICE decisions.

In the second sensitivity analysis, the regression analysis was run for a subset of complete observations ($n = 98/118$). This sensitivity analysis was implemented with the knowledge that removing incomplete observations from the analysis could bias the analysis. The pseudo-R^2 for this model was 0.32,

suggesting that this set of variables explains ~32% of the variability observed in NICE decision making, as opposed to 26% in the base–case model. In this sensitivity analysis, the impact of the ICER and the year of appraisal remained similar to that observed in the base–case analysis: however, the impact of the demonstration of statistical superiority, as well as the number of technologies appraised simultaneously, was weaker in this sensitivity analysis as a statistically significant effect on the odds of non-recommendation was no longer observed.

In the third sensitivity analysis, ordinality of the outcome variable was assumed and ordinal logistic regression was used. The results of this analysis show very similar results to the base–case analyses.

Discussion

This study aimed to contribute to the literature on NICE decision making in several ways. First, it examined NICE decision making beyond its initial years: published multivariate analyses of NICE decision making have relied on data from 2000 to 2003, during the years following NICE's inception (e.g. Dakin *et al.*, 2006). The results of this new analysis provide a complementary view on the characteristics of technologies reviewed by NICE beyond 2003, the factors that have continued to have an impact on NICE decision making and opportunity to examine change in NICE decision making outcomes over time. Second, this study aimed to complement extant studies of NICE decision making by considering new variables not previously examined. The value of this approach was twofold: it allows for the examination of impact of new variables on NICE decision making (e.g. impact of therapeutic indication or use of STA vs MTA processes); and it allows for the control of confounding factors in the multivariate analysis.

Finally, by increasing the number of appraisals included in the analysis, this study has greater statistical power when examining factors potentially having an impact on NICE decision making. For example, Dakin *et al.* (2006) reported that regression analysis including cost per QALY data was not feasible because of the relatively small sample size. Therefore, the impact of the cost-effectiveness ratio was assessed combining both cost per QALY and cost per life year cost-effectiveness ratios and was not found to have a statistically significant impact on the regression analysis.

This study of NICE decisions suggests that a combination of clinical, economic and process factors explains decisions made by NICE to recommend, restrict or not recommend pharmaceutical technologies for use. The internal validity of the results obtained in this analysis was examined in two ways: first, by comparing the results with published analyses of NICE decision making, and second, by sharing the base–case model results for review with a leading member of NICE.

Pattern of NICE decision making

The pattern of NICE decisions observed in this analysis is not dissimilar to that reported in a study by Kanavos *et al.* (2010), which examined NICE decisions in 2007–2009: that analysis revealed that, of the technologies appraised, 19% were recommended, 63% were restricted and 18% were not recommended. Clement *et al.* (2009) examined NICE decisions between 2001 and 2008 and reported 87% of technologies as listed (recommended or restricted), leaving 13% of technologies as not recommended. Analysis of NICE decision making during the year 2000–2003 by Dakin *et al.* (2006) reported that 21% of technologies were recommended for routine use, 66% were for restricted use and 13% were not recommended. Devlin and Parkin (2004) examined NICE outcomes using a binary outcome variable, and during the period 2000–2002 found that 71% of appraisals recommended use of the technology and 29% of appraisals did not. Therefore, the pattern observed within the data set used in this new study seems similar to that observed in previous studies, suggesting that the method of classification and data extraction used was robust.

Impact of clinical evidence and disease characteristics on NICE decision making

It was hypothesised that clinical variables, which were found to be significant explanatory variables in previous NICE analyses, would maintain their significance in this analysis. In the analysis of NICE decision making presented here, demonstration by the technology under appraisal of statistically significant superiority in its primary endpoint increased the odds of recommendation. Technologies recommended by NICE demonstrated statistically significant superior efficacy over the comparator in 59% of appraisals, compared with 29–35% of technologies that were restricted or not recommended. This result can be seen to reflect the role of evidence-based medicine in HTA decisions and the fact that NICE defines the value of the compound in terms of the ability of the technology to demonstrate, with greater certainty, its incremental clinical value through superiority-designed trials that provide stronger data to support a funding decision than technologies not able to provide evidence of superior efficacy. Comparing this result with previous studies, it is noteworthy that Devlin and Parkin (2004) did not measure the demonstration of clinical superiority, whereas Dakin *et al.* (2006) measured this variable but did not find it to have a statistically significant impact, although the sample size was smaller than that used here, and their study examined appraisals made in 2000–2003.

It was hypothesised that technologies for diseases characterised by high unmet medical need may increase the log odds of recommendation relative to non-recommendation – with particular focus on cancer therapies. Through the model-specification process, the effect of cancer therapies on the log odds of non-recommendation was not observed and thus the final base–case model does

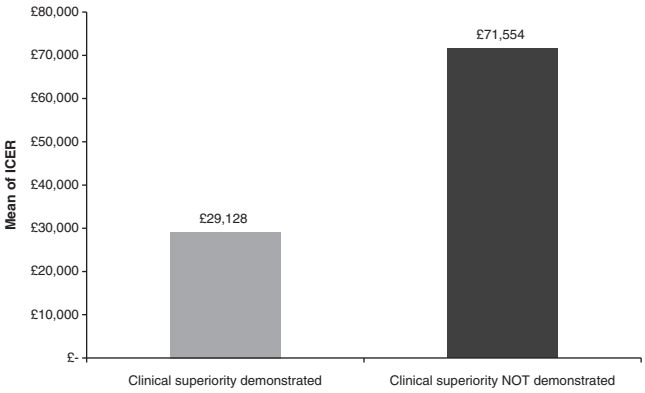

Figure 3. NICE decision-making: mean ICER stratified by demonstration of clinical superiority in RCT (n = 90)

not include this particular variable. The hypothesis was also not supported by univariate analyses: there was no statistically significant difference detected in the proportion of cancer therapies across NICE decision categories.

Economic evidence and its impact on NICE decision making

Consistent with previously published analysis of NICE decision making (Devlin and Parkin, 2004; Dakin *et al.*, 2006), the ICER had a significant impact. An increase in the ICER increased the odds of a restriction vs recommendation or non-recommendation vs recommendation, and this effect was highly statistically significant. The effect of the ICER was observed consistently throughout all of the sensitivity analyses performed. Thus, in addition to the strength of the clinical data, the incremental costs and benefits associated with the technology, and the resulting ICER, play a significant role in decisions by NICE. This interpretation of the multivariate model results is reinforced by a descriptive assessment of the mean ICER reported for technologies that demonstrated clinical superiority compared with those technologies that did not, suggesting that the ICER is a product of the clinical value demonstrated (Figure 3). It is recognised that ICERs represent a combination of inputs including clinical and economic variables as well as disease and patient characteristics. According to inputs used within an economic analysis, and the range of uncertainty accompanying them, this can lead to differing ICERs. It was on the basis of this knowledge that, in addition to the base–case ICER, the analysis included data on the range of base–case ICERs considered within the NICE review submitted by stakeholders, and sensitivity analysis/uncertainty estimates around the base–case ICER. However, these variablse were not found to have a significant impact on NICE decision making, when adjusting for other factors.

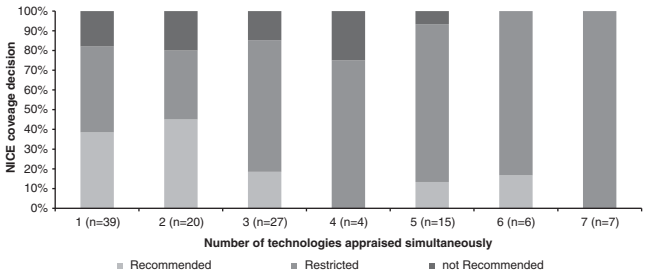

Figure 4. NICE decision-outcome stratified by number of technologies appraised (N = 118)

The number of technologies appraised simultaneously affects NICE decision making

The multivariate analyses also suggest that there are process factors, beyond evidence considerations, which help to explain NICE decisions. An increase in the number of technologies reviewed simultaneously within the same appraisal increased the odds of a restriction relative to a recommendation. It was hypothesised that this may reflect the fact that NICE assessment processes differ according to the number of technologies under appraisal. However, when the model specification was altered to include a variable capturing the use of MTA or STA processes, the effect of this variable on NICE decisions was not statistically significant. It was suggested that the increased odds of restriction associated with higher number of technologies appraised simultaneously may reflect an approach in which a 'winner' is picked among the technologies, with the remainder recommended for restricted use or non-recommendation. Descriptive analysis was performed of NICE decision outcome, stratified by the number of technologies appraised (Figure 4): it suggests that the proportion of restrictions increases with the number of technologies appraised, and that non-recommendation appears to occur most frequently in appraisals with fewer technologies appraised simultaneously. None of the previous multivariate analyses of NICE decision making examined the role of process factors, and thus comparison with earlier findings is not possible.

Year of appraisal affects NICE decision making

In terms of the socio-economic context of NICE decision making, it was hypothesised that non-recommendations and restrictions are increasing over time relative to recommendations. The analyses show that the year of appraisal had a significant impact on NICE decisions – moving from 2004 to 2009 increased the odds of a restriction vs recommendation and non-recommendation vs recommendation (Figure 5). The year of appraisal may reflect multiple socio-economic factors, including the political climate, a change in key NICE staff,

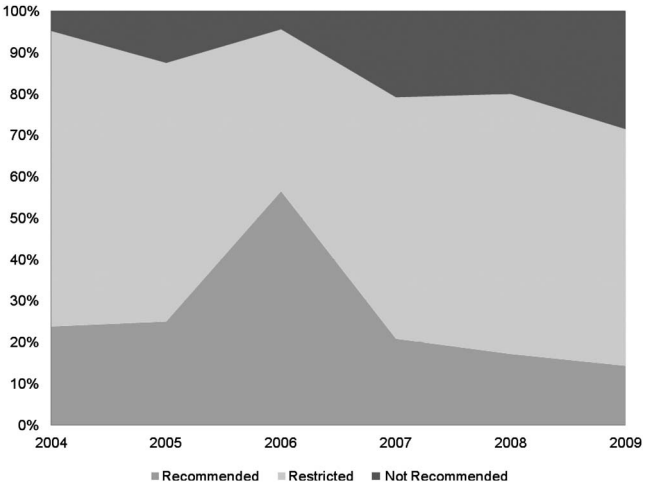

Figure 5. NICE decisions between 2004–2009 (n = 118)

a change in societal preferences or the overall economic context. It should be noted that this analysis was performed before the implementation of changes in NICE appraisal methods, including criteria for consideration in the appraisal of rare diseases and cancer therapies. However, the analysis does cover the period where other key changes took place in the appraisal process, including the introduction of the Single-Technology Appraisal process as well as the 2008 update to the methods guide. By way of comparison, Dakin *et al.* (2006) included this variable in their analyses and also found a statistically significant effect of the time of appraisal on outcome: an increase in the time of appraisal (i.e. as the appraisals came closer to the present), increased the odds of a non-recommendation. However, these authors did not find a statistically significant effect of appraisal date on the odds of a restriction relative to a recommendation.

Limitations

Owing to limited resources, it was necessary to restrict the guidance included for analysis and number of factors taking into account NICE decision making. Although there are four types of guidance that NICE provides (NICE, 2008b), this analysis focused on recommendations on pharmaceutical technologies. This restriction on pharmaceutical technologies was done not only to increase the homogeneity of the sample for analysis in terms of homogeneity of the technologies included, but also to increase homogeneity in the approach to decision making, which can vary by type of guidance.

The variables included in this analysis represent factors for which indicators could be defined and coded, and for which evidence could be collected retrospectively. Although not representing the complete set of factors taken into

account by NICE, they do represent the largest set of variables collected on NICE decisions and represent a range of factors hypothesised to explain NICE decision making. As the results show, the variables included in the analysis explain a proportion of the variance of NICE decision making, highlighting that there is some unexplained variance linked to variables that have not been captured in this analysis, or to the fact that some influences on decisions are simply random.

The database constructed for these analyses, incorporating information on appraisals performed by NICE from 2004 to 2009, was dependent on publicly available information. Thus, it is possible that subtle concepts or rationales discussed by the appraisal committee orally were not captured in the documentation of the appraisal. In addition, the dependence on public information meant that, in the situation where the information available was incomplete, it was not possible to ascertain whether this was because the information was never considered in the appraisal or whether it was considered but not recorded in the documentation. Incomplete observations were taken into account in the regression models by creating dummy variables to examine whether the presence or absence of information on that variable had any explanatory value. None of these dummy variables appeared to have a significant effect on the odds of NICE decisions. A sensitivity analysis was also performed in which incomplete observations were removed from the sample for analysis. The results of this sensitivity analysis were similar to those observed in the base–case analysis, suggesting that the factors identified in the base–case model are robust to alternative model specifications.

There was heterogeneity in the means through which technologies were restricted within NICE decisions. The notion of restriction within NICE decisions ranged from major restrictions, including restriction for use within a subset of the licensed indication, to minor restrictions such as the need for monitoring along with the use of the technologies. The notion of major and minor restrictions was suggested in research on NICE guidance by Raftery (2006), in which various sub-types of restrictions were presented. O'Neill and Devlin also highlighted the variation in the degree of restriction related to NICE decision making (O'Neill and Devlin, 2010). It is a limitation of this analysis to have such heterogeneity in the degree of restriction within a single category. However, in its actual decision making, NICE has more than two coverage options at its disposal and thus the use of a third coverage category within the analysis reflected more closely real-life decision making. To test the impact of using a binary outcome variable on the base–case model results, sensitivity analysis was performed using a binary outcome variable, which confirmed the role of the factors identified in the base–case analysis.

Conclusion

The objective of this analysis was to examine the factors that influence decisions made by NICE to recommend, restrict or not recommend pharmaceutical

technologies for use by the NHS in England and Wales. The analysis provided a rich source of data from which to examine the role of each factor on NICE decisions, and more importantly the contribution of each factor while adjusting for the effect of confounding variables. The results suggest that the variability in decisions observed can be explained by a combination of clinical, economic, process and socio-economic factors. The analysis showed that the proportion of restrictions and non-recommendations issued by NICE are increasing over time relative to recommendations. The analysis also confirmed that the demonstration of clinical and economic value is central to NICE decisions. While the NICE appraisal process was also shown to have an impact on decision making, the anticipated effect of the use of STA or MTA processes was not observed; rather, an effect was found for the number of technologies appraised simultaneously. New factors not previously reported to have an effect on NICE decision making were identified, including the effect of clinical superiority on NICE decision making, the effect of the ICER on the likelihood of both restriction and recommendation and that NICE decision making was sensitive to process variables as well as socio-economic factors. The results confirm the value of a comprehensive and multivariate approach to understanding NICE decision making.

Acknowledgements

The authors would like to thank Michael Drummond and Alistair McGuire for their valuable comments on the methods used and interpretation of analyses that were performed.
Disclosure statement: No funding was received for this study. Karin Cerri was an employee of Bristol-Myers Squibb during the time this research was conducted.

References

Al, M. J., T. Feenstra and W. B. Brouwer (2004), 'Decision makers' views on health care objectives and budget constraints: results from a pilot study', *Health Policy*, **70**(1): 33–48.

Association of the British Pharmaceutical Industry (ABPI) (2010), 'Facts & Statistics from the Pharmaceutical Industry – Medicines and the NHS'. www.abpi.org.uk/statistics/section.asp?sect=4#15 [29 January 2010].

Barbieri, M., N. Hawkins and M. Sculpher (2009), 'Who does the numbers? The role of third-party technology assessment to inform health systems' decision-making about the funding of health technologies', *Value Health*, **12**(2): 193–201.

BBC (2005), '2005: Historic third term for Labor', http://news.bbc.co.uk/2/hi/uk_news/politics/vote_2005/6994476.stm [22 December 2010].

Bryan, S., I. Williams and S. McIver (2007), 'Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals', *Health Economics*, **16**(2): 179–193.

Buxton, M. J. (2005), 'How much are health-care systems prepared to pay to produce a QALY?', *European Journal of Health Economics*, **6**: 285–287.

Chalkidou, K. (2009), 'Comparative effectiveness review within the U.K.'s National Institute for Health and Clinical Excellence', *The Commonwealth Fund Issue Brief*, **59**: 1–12.

Clement, F. M., A. Harris, J. J. Li, K. Yong, K. M. Lee and B. J. Manns (2009), 'Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada', *JAMA*, **302**(13): 1437–1443.

Dakin, H. A., N. J. Devlin and I. A. O. Odeyemi (2006), '"Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making', *Health Policy*, **77**: 352–367.

Department of Health (2002), Improvement, Expansion and Reform: The Next Three Years Priorities and Planning Framework 2003–2006 London: Department of Health (Crown Copyright).

Department of Health (2006a), *The NHS in England: The Operating Framework for 2006/7*, London: Crown Copyright.

Department of Health (2006b), *The NHS in England: The Operating Framework for 2007/08*, London: Crown Copyright.

Department of Health (2007), *The NHS in England: The Operating Framework for 2008/09*, London: Crown Copyright.

Devlin, N. and D. Parkin (2004), 'Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis', *Health Economics*, **13**(5): 437–452.

EUnetHTA (2011), 'HTA definition' 2011, http://www.eunethta.net/Public/AboutEUnetHTA/HTA/ [2 February 2011].

European Medicines Agency (2011), 'European Public Assessment Reports', www.ema.europa.eu/ema [30 March 2011].

Health Technology Assessment International (HTAI) (2009), http://www.inahta.org/upload/HTA_resources/AboutHTA_Policy_brief_on_HTA_Introduction_to_Objectives_Role_of_Evidence_Structure_in_Europe.pdf [2 June 2009].

Henshall, C., W. Oortwijn, A. Stevens, A. Granados and D. Banta (1997), 'Priority setting for health technology assessment – theoretical considerations and practical approaches – a paper produced by the priority setting subgroup of the EUR ASSESS project', *International Journal of Technology Assessment in Health Care*, **13**(2): 144–185.

Joint Formulary Committee (2010), *British National Formulary, 60 ed*, London: British Medical Association and Royal Pharmaceutical Society.

Kanavos, P., E. Nicod, S. van den Aardweg and S. Pomedli (2010), 'The impact of health technology assessments: an international comparison', *Euro Observer* **12**(4): 1–7.

Littlejohns, P., S. Garner, N. Doyle, F. Macbeth, D. Barnett and C. Longson (2009), '10 years of NICE: still growing and still controversial', *Lancet Oncology*, **10**(4): 417–424.

Mason, A. R. and M. F. Drummond (2009), 'Public funding of new cancer drugs: Is NICE getting nastier?', *European Journal of Cancer*, **45**(7): 1188–1192.

Menon, D., T. Stafinski and G. Stuart (2005), 'Access to drugs for cancer: Does where you live matter?', *Canadian Journal of Public Health*, **96**(6): 454–458.

National Office for Statistics (2009), Key demographic and health indicators, 1976 onwards: Population Trends. ONS 1976-2008 data.

NICE (2008a), *Guide to the methods of technology appraisal*, NICE. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf [15 May 2009].

NICE (2008b), *Social Value Judgements – Principles for the Development of NICE guidance*. Second edition, NICE http://www.nice.org.uk/media/C18/30/SVJ2PUBLICATION2008.pdf [15 May 2009].

NICE (2009), *Threshold Workshop Report*. NICE.

O'Neill, P. and N. J. Devlin (2010), 'An analysis of NICE's 'restricted' (or 'optimized') decisions', *Pharmacoeconomics*, **28**(11): 987–993.

OECD ed. (2005), *Health Technologies and Decision Making*, The OECD Health Project, OECD Publishing. Printed in France.

OECD (2006), *OECD Health Data*, OECD. http://www.oecd.org/statisticsdata/0,3381,en_2649_37407_1_119656_1_1_37407,00.html [30 June 2008].

Owens, D. K. (1998), 'Interpretation of cost-effectiveness analyses', *JGIM*, **13**: 716–717.

Packer, C., S. Simpson and A. Stevens, EuroScan: the European Information Network on New and Changing Health Technologies. (2006), 'International diffusion of new health technologies: a ten-country analysis of six health technologies', *International Journal of Technology Assessment in Health Care*, **22**(4): 419–428.

Raftery, J. (2006), 'Review of NICE's recommendations, 1999–2005', *BMJ*, **332**(7552): 1266–1268.

Ross, J. (1995), 'The use of economic evaluation in health care: Australian decision makers' perceptions', *Health Policy*, **31**(2): 103–110.

Sorenson, C., M. Drummond and P. Kanavos (2008). *Ensuring value for Money in Health Care: the role of HTA in the European Union*. Cornwall: World Health Organisation, on behalf of the European Observatory on Health Systems and Policies.

Summerhayes, M. and P. Catchpole (2006), 'Has NICE been nice to cancer?', *European Journal of Cancer*, **42**(17): 2881–2886.

Velasco Garrido, M. and R. Busse (2005), *Health Technology Assessment – An Introduction on Objectives, Role of Evidence, and Structure in Europe*, Brussels: European Observatory on Health Systems and Policies.

Vuorenkoski, L., H. Toiviainen and E. Hemminki (2008), 'Decision-making in priority setting for medicines – a review of empirical studies', *Health Policy*, **86**(1): 1–9.