Office of Health Economics

Research

Research Paper 16/08

An Analysis of NICE Technology Appraisal Decisions 'Recommended in Line with Clinical Practice'

December 2016

Phill O'Neill, Amanda Cole, Aurea Duran and Nancy Devlin

An Analysis of NICE Technology Appraisal Decisions 'Recommended in Line with Clinical Practice'

Phill O'Neill[1], Amanda Cole[1], Aurea Duran[2] and Nancy Devlin[1]

[1] Office of Health Economics, London

[2] Astellas Pharma Europe Ltd (Pfizer Ltd at time of research)

December 2016

For further information please contact:
Phill O'Neill
poneill@ohe.org
Office of Health Economics
Southside, 105 Victoria Street
London SW1E 6QT
United Kingdom
Tel: +44 207 747 7152

CONTENTS

ABSTRACT

Between January 2007 and June 2016 nearly a quarter (89 of 407) of all National Institute for Health and Care Excellence (NICE)-reported decisions for individual technologies assessed through their technology appraisal (TA) process have been characterised by them as *recommended in line with clinical practice* (RiLCP), to be distinguished from *recommended in line with marketing authorisation*. These RiLCP decisions are counted as 'recommended' in NICE statistics summarising technology appraisal (TA) decision outcomes, and 44% percent (89 of 203) of all 'recommended' decisions in the period 2007-2016 are given this classification.

The objectives of this study are twofold: First, we qualitatively assess documentation associated with RiLCP decisions to create a framework to describe common themes and criteria used to reach those decisions. Second, we assess the level of patient access associated with each RiLCP decision using a previously developed method – the 'M' score. We believe that this is the first systematic assessment of this class of decisions.

Whilst RiLCP decisions appeared often to be grounded in the objective of matching guidance with established clinical practice (for example, referencing relevant previous NICE TAs or clinical guidelines), in 40 out of 89 cases we identified clinical/cost-effectiveness as the prime motivation for limiting the patient population (relative to to marketing authorisation). Using the previously developed 'M' score method for measuring the level of access associated with RiLCP decisions, we estimate that, relative to license, an average of 53 out of 100 patients were recommended as suitable for treatment.

Together, the nature of the evidence used to influence recommendations and the level of access associated with them suggests that, for many of these decisions, it is difficult to distinguish between RiLCP and 'optimised' recommendations. In these instances restrictions are being placed on the use of innovative medicines, and reported in a way that is not readily transparent that this is the case. This may have implications for the interpretation of NICE recommendations and for NICE's reported trends in decision outcomes; statistics that are used to judge whether NICE TA processes are achieving their objectives. These include facilitating the diffusion of innovation into the National Health Service (NHS).

BACKGROUND

In order to summarise the outcome of National Institute for Health and Care Excellence (NICE) technology appraisal (TA) recommendations NICE produce 'NICE Statistics', which are updated on a regular basis (1). Decisions are characterised according to four 'types': 'Recommended', 'Optimised', 'Only in research', and 'Not recommended'. Within 'Recommended' decisions, the drug or treatment is recommended for use either (a) in line with the marketing authorisation from the European Medicines Agency (EMA) or Medicines and Healthcare Products Regulatory Agency (MHRA); or (b) in line with how it is used in clinical practice in the National Health Service (NHS) or (c) both.

This means that, for certain decisions by NICE to 'Recommend' a product, that recommendation does not extend to its full marketing authorisation. An explanation is offered in an FAQ: "*Sometimes experts suggest that the technology is unlikely to be used routinely in clinical practice in the UK, to the extent permitted by the license. In these cases, the recommendation may be classed as 'recommended' because this is in line with clinical practice.*" (1).

Further clarification of how this might differ to a decision to 'optimise' use is offered if we consider the description of Optimised decisions, which are recommendations that have a "***material effect*** *on the use of a drug or technology, and it is recommended for a smaller subset of patients than originally stated by the marketing authorisation.* ***This test of materiality takes into account advice from clinical experts on the anticipated use of the technology in routine clinical practice****.*" (1) (Italics added).

We understand 'material effect' to mean that the conditions of use specified in an 'optimised' decision will result in *lower* uptake than the uptake that a straightforward 'Recommendation' would have engendered. This therefore implies that a recommendation in line with clinical practice, whilst limited compared with market authorisation, is consistent with how it would be used in current clinical practice. Moreover, it suggests that the changes to clinical practice that would be required in order to support the use of a product as wide as that permitted by its licence are either not feasible or not worthwhile.

Whilst each appraisal is different in terms of the type of product it deals with, the level of evidence available, and the context of use, the terminology that is used to characterise a decision by NICE is important as this effects an understanding of NICE's role in limiting or promoting access to new technologies, based upon their assessment of their effectiveness and cost effectiveness.

Between January 2007[1] and June 2016 NICE report that they have made 89 individual technology appraisal (TA) decisions categorised by them as 'recommended in line with clinical practice' (hereafter RiLCP). There are instances where a number of TA decisions are associated with a single appraisal, for example where NICE assess a medicine in both combination and monotherapy. Using NICE's own count of decision outcomes between

[1] 2007 was chosen as a start date as NICE costing templates, needed for our analysis, began to be systematically produced at this time.

2007 and June 2016, RiLCP represent 22% of all decisions, and represent 44% of the 203 decisions to 'recommended' a product, and hence are significant when assessing trends in NICE outputs. They are reported as 'recommended' in NICE's summary statistics which are used in NICE published statements. This categorisation is not well explained and implications for patient access are not clear. The aims of this paper are to:

(a) Characterise RiLCP decisions, assessing whether and what common characteristics underlie these decisions; and

(b) Consider the level of patient access associated with these decisions. For a lay person or health care professional a description of 'recommended' implies that all patients who could have received the treatment (e.g. in line with its marketing authorisation) have been recommended to do so. We undertake an assessment of the level of patient access associated with a recommendation using a previously developed measure: the 'M-score'.

METHODS

To achieve the two aims, separate analyses were undertaken, one assessing RiLCP decision documentation to discern characteristics of those recommendations and the second applying our measure of patient access to these decisions.

Characterising RiLCP decisions

In order to gain further insight into the rationale behind and implications of a recommendation 'in line with clinical practice', we assessed the deliberations and recommendations of the committee as summarised in TA guidance documentation for technologies that have received this designation.

All products with a RiLCP outcome according to the NICE TA Decision Summary document between January 2007 and June 2016 were evaluated. By considering the wording of the final recommendation, as well as the summaries of evidence and the committee's consideration of that evidence, we identified emerging themes relating to the choice of recommendation. We summarise these emergent themes below, and provide examples in the results section.

Exploring the rationale for recommending a technology 'in line with clinical practice'

Between January 2007 and end of June 2016 NICE made 89 RiLCP decisions. All TAs were included, even if they had been subsequently superseded (analysis conducted both on the early and subsequent versions with the relevant decision at time of publication). By assessing the details of each recommendation contained within the relevant appraisal, two reviewers on the study team identified six main themes relating to the perceived rationale for recommending a technology in line with clinical practice, and categorised each decision according to these themes. The themes were:

- Reference to a previous NICE TA
- Existence of a relevant clinical guideline
- The product fits within an established pathway of care

- Clinical opinion
- Clinical / cost-effectiveness matching
- Non-pharmaceutical
- Other reason.

Most recommendations could be seen to relate to more than one of the categories above, and reviewers selected both a main and a secondary reason for each decision. Reasons 1 to 4 are related, and can be regarded as reflecting clinical practice on a more formal (TA or clinical guideline) or less formal (established practice/clinical opinion) basis. Whilst clinical/cost-effectiveness matching was often identified as a secondary reason behind a RiLCP decision, this also appeared to be the primary reason in many cases. Non-pharmaceuticals were set apart for reasons described below, and other possible rationales were also highlighted. In the results section we describe each reason and provide examples.

The sample of decisions was randomly split and divided among the two OHE reviewers. Information was extracted using a common framework of categorisation and data extraction (providing quotation references where relevant). This qualitative exercise inevitably includes an element of subjectivity, and therefore we do not emphasise the quantitative summary of our categorisations. Rather, we highlight, describe, and provide examples of the common themes, and explore how these relate to NICE's recommendation framework.

Measuring patient access for RiLCP decisions using the M score

O'Neill and Devlin (2010) developed an approach to assess the extent of access associated with NICE TA decisions (2). This measure is labelled the M score, defined as $M = (p/P)*100$, where M is a measure of the level of patient access (0 = no access, 100 = full access), P is the set of patients considered in the guidance as potential candidates for treatment (given the licensed use and the scope of NICE's appraisal), and p is a subset of those patients, for whom NICE did recommend treatment. In the original paper we considered M relative to individual medicines or groups thereof. This was because NICE can and do make different recommendations, associated with different levels of access, for two or more medicines for the same indication. It is the case that from the point of view of access NICE may not have recommended the specific treatment for a subgroup of patients but have recommended an alternative. For example patients with diabetes will not be recommended access to newer treatments if their symptoms are controlled on older medications. By combining the subgroups of patients the level of access for the indication can be measured. For the purposes of this analysis we calculate M for individual medicines. This is because we are assessing each individual medicine in the context of a reported RiLCP decision.

Measure M provides a means to benchmark access. The sources of information for P and p are NICE's costing templates (which are included in the materials produced by NICE to support each TA). The purpose of these costing templates is to assist NHS organisations (e.g. Clinical Commissioning Groups) to anticipate the budgetary impact of NICE's recommendations, and to assist with local NHS resource allocation and planning. In many (but not all) instances, the information provided to support these cost estimates

enables a comparison to be made between the licensed indication under consideration and the actual recommendation made by NICE.

RESULTS

Categories of NICE recommended in line with clinical practice decisions

Table 1 summarises the key emergent themes from the qualitative assessment of RiLCP decisions, along with an example of each. These summarise our own assessment of why the decision appears to have been labelled RiLCP, i.e. why it has not been recommended in line with the full marketing authorisation. We describe the themes in the table in more detail below.

Reference to a previous TA

For several recommendations, reference to another NICE technology appraisal could be regarded as the basis for NICE recommending a product in line with 'clinical practice'; this could either be reference to the recommendation criteria for a previous appraisal for a similar product, or an earlier TA considering the same product (i.e. a TA which the new guidance replaces). An example of the former is Adalimumab (TA 125): "*Furthermore, the Committee agreed that criteria for the use of adalimumab should be identical to the criteria listed in current NICE guidance for the use of etanercept in the treatment of adults with Psoriatic Arthritis (NICE technology appraisal guidance 104)" (3, p.17).* An example of the latter is the recommendation for Clopidogrel and Dipyridamole in TA210 (which replaced TA90), where *"The Committee concluded that the data published after NICE technology appraisal guidance 90 supported the conclusions in that guidance." (4, p.21)*. Whilst the decision determination for TA90 was labelled 'Optimised', the new determination was 'in line with clinical practice'. These reasons for labelling a recommendation 'in line with clinical practice' appear justified and consistent with the definitions outlined in the introduction. It suggests that, because TAs must be implemented, an update to that guidance which is consistent with the previous recommendations is in line with the (newly established) clinical practice, despite the decision overall being optimised compared with the drugs' marketing authorisation. For the appraisal of Afilbercept, described in Table 1, the marketing authorisation permitted wider use for the treatment of (wet) age-related macular degeneration. However the more restrictive criteria of a comparator product for which a TA had been produced was applied, due to most of the trial evidence incorporating these criteria, despite clinical input suggesting wider use might be beneficial (5).

Table 1 Summary of criteria for RiLCP decisions identified

Criterion	Example
1. Reference to previous NICE TA Explicitly refers to a previous NICE TA of the same or a related product	TA294 Afilbercept for wet age-related macular degeneration. Afilbercet recommended for use in patients according to same criteria as previous NICE TA for competitor product, until both can be appraised in a multiple technology appraisal. This was despite suggestion by clinical experts that Afilbercept may provide benefits to a wider patient population.
2. Reference to relevant clinical guideline Recommendations are matched to existing clinical guidelines including NICE guidelines	TA138 Inhaled corticosteroids (ICS) for the treatment of chronic asthma. BTS/SIGN guideline is referred to describe the step-wise approach to treatment that ICS use should conform with.
3. Fits within an established pathway of care Recommendation aligned with a pre-existing patient management pathway (with no particular 'guideline' for reference)	TA223 Treatment of intermittent claudication in people with peripheral arterial disease, where Naftidrofuryl *"represents one part of a wider programme of management"*.
4. 'Clinical opinion' Guidance is explicitly influenced by input from clinical experts	TA315 Canagliflozin in combination therapy for treating type 2 diabetes. Only recommended where sulfonylurea was not appropriate given extensive experience of clinicians with the current treatment.
5. Clinical/Cost-effectiveness evidence matching Guidance recommendation is shaped by the clinical and/or cost-effectiveness evidence for the technology	TA181 Pemetrexed for the first-line treatment of non-small-cell lung cancer. Recommended for specific histologically-determined subgroup based on clinical effectiveness evidence. Requires a *change* in clinical practice to implement.
6. Non-pharmaceutical	*All* non-medicinal products in our sample that were recommended were done so 'in line with clinical practice'.

Existence of a relevant clinical guideline

Similar to linking to a previous TA, referencing a relevant clinical guideline was found to be a relatively common reason behind a recommendation in line with clinical practice. Often, clinical guidelines were used to specify eligibility for treatment. For example, in TA143 where the committee discussed the criteria for starting therapy, and was "*in agreement in general with the criteria set out in the BSR guidelines for prescribing TNF-α inhibitors in adults with ankylosing spondylitis in terms of diagnostic criteria (modified*

New York criteria), confirmation of sustained active spinal disease, and failure of conventional treatment to control symptoms " (6, p.33).

The product fits within an established pathway of care

The perceived reasons behind the few appraisals that were assigned to this category were very similar to those described for reason 2, the difference being that there was no particular guideline to refer to. Mainly this was because the therapy was to be introduced into a clinical area with a pre-existing pathway of care, for example for TA114 where the committee noted that buprenorphine and methadone for the management of opioid dependence should be part of a programme of supportive care (7).

Clinical opinion

As described, this fourth (less formal) variant we identified of representing clinical practice was where the primary reason for an 'in line with clinical practice' decision appeared to be clinical opinion. There are instances where guidance documents made explicit mention of this input to determine the nature of the final guidance or to influence clinical or cost-effectiveness considerations. An example is TA278 which considered the use of omalizumab for the treatment of asthma: *"The Committee heard from the clinical specialists and patient experts that, in current UK clinical practice, the population for which omalizumab would be considered was smaller than that covered by the marketing authorisation. Clinicians currently optimise a person's asthma treatment before considering omalizumab; for those whose asthma remains poorly controlled, and affects their quality of life, omalizumab is considered as an add-on treatment. […] The Committee concluded that only people with the most severe persistent allergic asthma despite optimised treatment would currently be offered omalizumab"* (8, p.37).

Clinical/cost-effectiveness matching

Those recommendations that appeared to be restricted (compared with marketing authorisation) according to clinical/cost-*effectiveness* criteria rather than clinical *practice* were labelled under this category. This perceived motivation for an 'in line with clinical practice' recommendation appears to be the most difficult to differentiate from 'optimised' decisions, in particular where its implementation required a *change* in clinical practice, and where the specification of the recommendation was materially different from the marketing authorisation.

It should be noted that for the majority of examples of matching the recommendation to the availability of evidence, matching was based on restricting access to patients whose characteristics reflected those in clinical trials (i.e. sub-groups were excluded because of no evidence rather than uncertain evidence). An example of such a recommendation is TA169 for Sunitinib, where "*The Committee noted that the sunitinib trial was conducted only with participants that had a good ECOG performance status of 0 or 1. Therefore the Committee concluded sunitinib is a clinically effective first-line treatment for advanced and/or metastatic RCC for patients with an ECOG performance status of 0 or 1*" (9, p.17).

An example which is particularly difficult to distinguish from an optimised decision, as its implementation would actually involve a *change* in clinical practice, is TA181 for

pemetrexed. In contrast to the marketing authorisation, the drug is only recommended for patients with confirmed adenocarcinoma or large-cell carcinoma. According to clinical trial data, the drug was most effective in these histological subtypes, and had not been shown to be more effective than standard practice, and hence more cost effective, for patients with non-squamous NSCLC with unspecified histology. However, determining this histology was not currently part of routine clinical practice but NICE recommended that this should be the case (10, p.17).

Non-pharmaceutical

This reason was used to categorise TAs that considered a non-pharmaceutical intervention. These were set apart as it can be noted that *all* non-pharmaceuticals (devices, procedures, behavioural therapies etc.) that were '*Recommended'* were done so '*in line with clinical practice'*. This is likely to be because a recommendation in line with market authorisation simply does not make sense for this category of interventions.

'Other'

Whilst none of the 'other' reasons identified were considered to be a primary reason for the categorisation of RiLCP, some other interesting potential motivations emerged. These included manufacturer expectation (where the evidence supplied by the manufacturer only related a sub-group of patients relative to license) and logistics (relating to concerns of logistical factors required locally to deliver treatment). There is one decision where off-label use is assessed because it has been established in clinical practice.

During extraction it became apparent that there were many decisions where more than one of the above factors played a role in the decision. It is also the case that some of the criteria above are clear in some cases but subject to interpretation in others as there is some overlap between categories. The table below is indicative of the various criteria identified as underlying RiLCP decisions. It reports 'primary' decisions, those that the extractor felt was the most important and 'secondary' decisions. These are not intended to be definitive summary statistics but used to demonstrate that a wide variety of factors appear to underline these decisions.

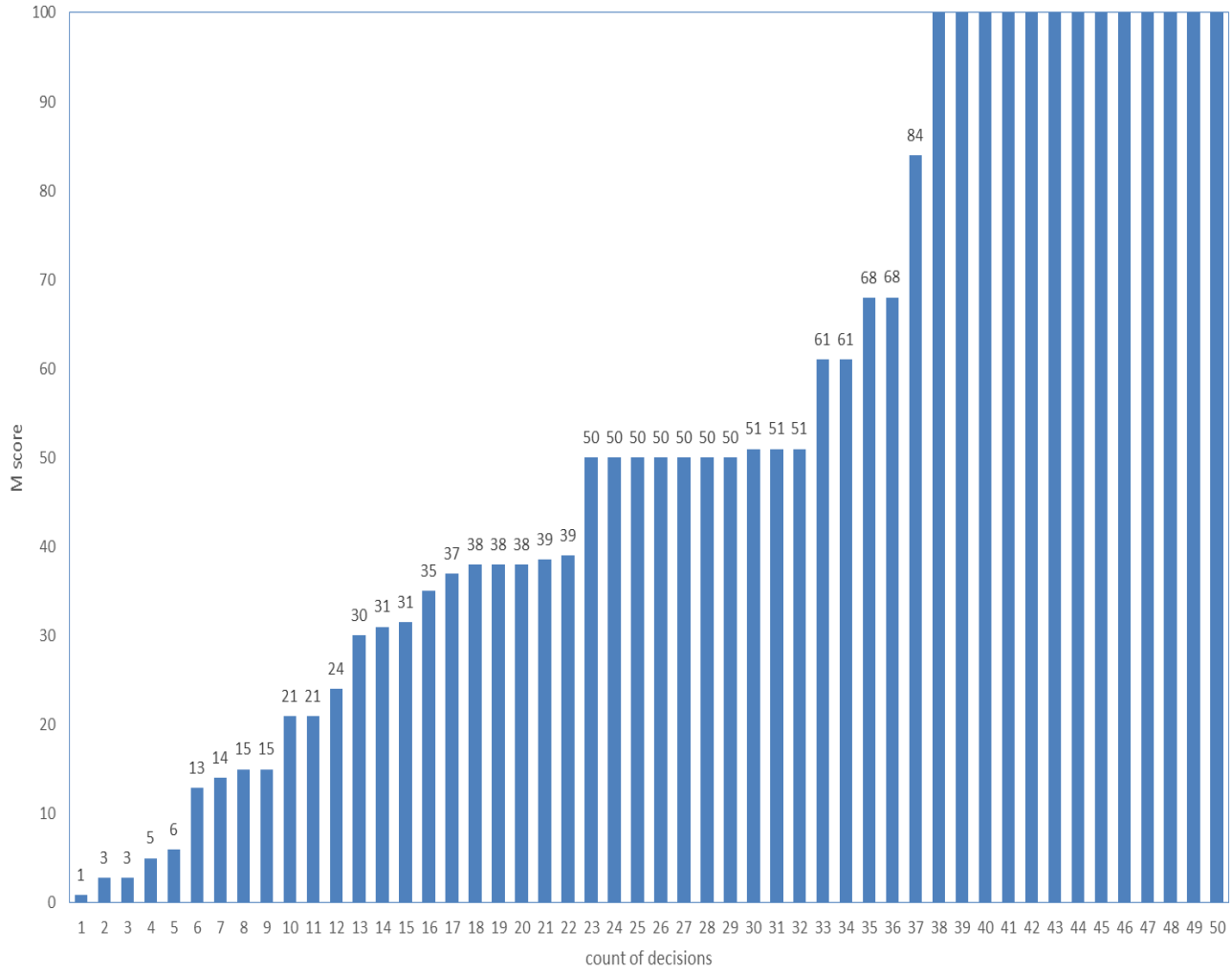
The 'Recommended in Line with Marketing Authorisation' categorisation provides clear criteria and intuitively covers what an unrestricted 'recommended' decision constitutes. For RiLCP this is less clear, and it was sometimes hard to distinguish these from optimised decisions. For many RiLCP decisions, recommendations relate to a clear subgroup of the population under consideration in the appraisal. Reference to previous TAs or clinical guidelines may appropriately reflect the 'clinical practice' that the new treatment is recommended in line with. However where decisions to limit the eligible population appeared to be based solely on considerations of clinical or cost-effectiveness, it is more difficult to understand how this can be distinguished from an optimised decision.

Table 2. 'Primary' and 'Secondary' factor underlying RiLCP decisions – January 2007-June 2016

Criterion	Primary factor	Secondary factor
1. Reference to previous NICE TA	8	8
2. Reference to relevant clinical guideline	5	3
3. Fits within an established pathway of care	13	7
4. 'Clinical opinion'	14	8
5. Clinical/Cost-effectiveness evidence matching	40	21
5a, of which cost effectiveness	33	
5b, of which clinical	7	
6. Non-pharmaceutical	*7*	4
7. Other/not specified	2	7
Total	89	58

To further explore the implications of this categorisation we can use measure M to analyse the degree of access associated with RiLCP decisions. Figure 1 plots M for all RiLCP decisions where an M score could be determined (50 out of 89) between January 2007 and the end of June 2016. The average M score was 53 (i.e. relative to the product's license, an average of 53 out of 100 patients were recommended as suitable for treatment).

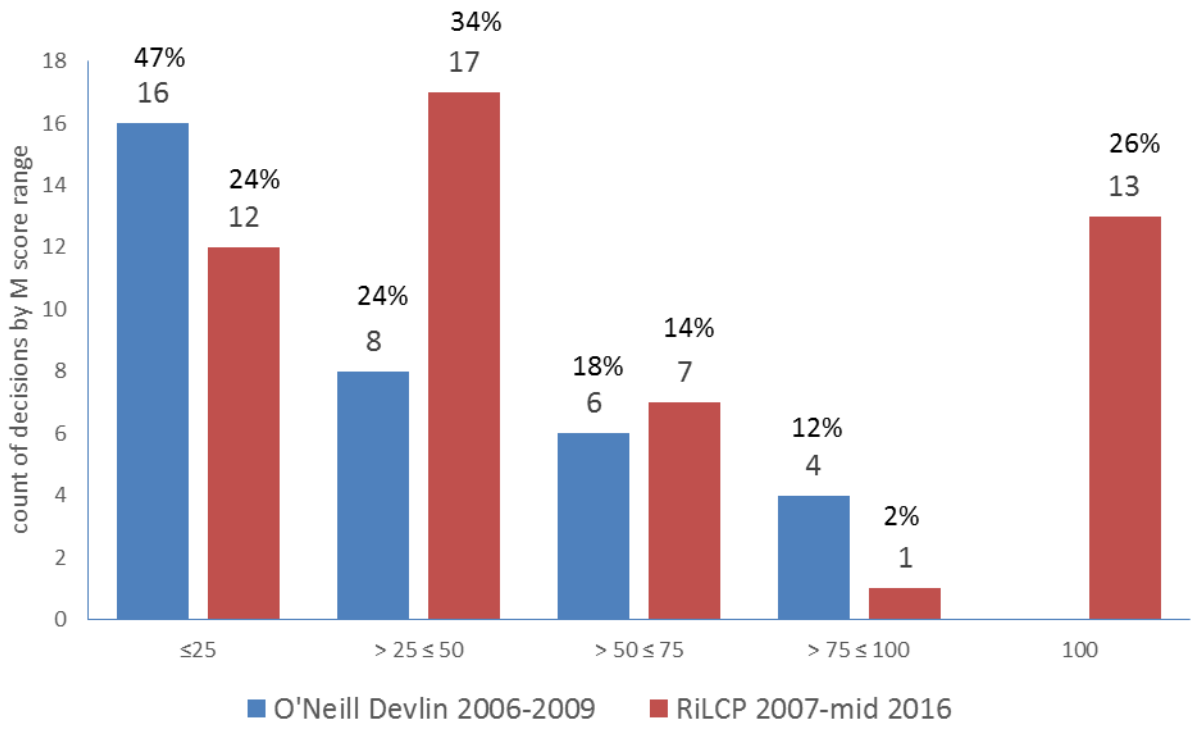
Figure 1. M scores for RiLCP recommended decisions where an M score could be determined



Whilst a score of 53 represented the average for those which could be calculated (n=50), if we conservatively assume that those for which we could not assess the M score (n=39) were associated with full access, then the average M score would be 74.

In O'Neill and Devlin (2010), M scores were calculated for 'optimised' decisions between 2006 and the end of 2009 (n=34 calculated from a sample of 69). One approach for presenting the results was to group individual decisions into ranges of M scores. Repeating this approach for the sample of RiLCP decisions where an M score could be calculated (n=50). The 2010 paper by definition excluded medicines with an M score of 100 as only optimised decisions were assessed. With this proviso comparing the two demonstrates that RiLCP decisions with a score below 100 have a similar distribution of M scores to the M scores for optimised decisions, as found in that study (see figures 4a and 4b below). Both are weighted towards M scores of below 50. This provides further support for the case that many of these are similar to optimised decisions, in terms of the level of patient access that they engender.

Figure 2. Comparison of M score ranges for optimised decisions in O'Neill and Devlin (2010) and RiLCP decisions (count and share)



DISCUSSION

The labelling of recommendations from NICE TA decisions are used to assess trends in decision outcomes. Amendments to NICE's HTA process and methodologies, such as End of Life considerations, are made in the context of these statistics.

In many cases the use of RiLCP is a legitimate approach as it enables the committee to frame the evidence in a UK context. For example where a cancer medicine is indicated for patients 'highly likely' to go into remission but this has not been clearly defined or found in the clinical trial evidence available. This is recognised in NICE's 'Guide to the methods of technology appraisal' stating that: "*Healthcare professionals and commissioners of health services provide a view of the technology in relation to current clinical practice. This puts into context the evidence derived from pre- and post-licensing studies, which often relates to efficacy and safety under clinical trial conditions rather than effectiveness in routine clinical practice*" (11, paragraph 4.4.2.)

For the period under review, NICE have described 89 out of 203 recommended decisions as RiLCP. The use of measure M provides context for considerations about access. An overall M score of 53 suggests that, as a group, these are different to recommended in line with marketing authorisation decisions. But there is significant variation in access for medicines in this group, including 13 where access was equivalent to a recommended in line with marketing authorisation decision. We believe that further clarity is required. For example where M is less than 100 a more appropriate designation could be 'Optimised in Line with Clinical Practice'.

Headline statistics reported by NICE do not make transparent that, in nearly half of RiLCP cases in the nine year period under the analysis, factors other than clinical practice appear influential in the decision to offer this designation. We found that the most common factor was cost-effectiveness evidence, and in 40 out of 89 cases either clinical or cost-effectiveness evidence. When this is the case, and where fewer patients are receiving treatment relative to marketing authorisation, then these might arguably be more appropriately labelled as optimised. By labelling these decisions as 'Recommended' the implication is that there are no restrictions placed on access, but this is not the case. NICE TA processes are intended to guide patients' access to innovation, based on their effectiveness and cost effectiveness -indeed the White Paper outlining how NICE would work was subtitled 'Faster Access to Modern Treatment' (12). Statistics supplied by NICE provide evidential context for the interpretation of NICE processes used to assess their performance, and where access to innovation has been restricted this should be transparent.

REFERENCES

1. NICE. 2016. NICE Statistics: Technology appraisal decisions. Available at: http://www.nice.org.uk/News/NICE-statistics [Accessed 30/07/2015, 13/05/2016 and 20/08/2016]

2. O'Neill P and Devlin N. 2010. An analysis of NICE's 'restricted' or 'optimised' decisions. *PharmacoEconomics,* 28: 987-93.

3. NICE. 2007. Final appraisal determination: Adalimumab for the treatment of psoriatic arthritis.

4. NICE. 2010. NICE Technology Appraisal Guidance 210: Clopidogrel and modified-release dipyridamole for the prevention of occlusive vascular events.

5. NICE. 2013. NICE technology appraisal guidance 294: Aflibercept solution for injection for treating wet age-related macular degeneration.

6. NICE. 2008. NICE technology appraisal guidance 143: Adalimumab, etanercept and infliximab for ankylosing spondylitis.

7. NICE. 2007. NICE technology appraisal guidance 114: Methadone and buprenorphine for the management of opioid dependence.

8. NICE. 2013. NICE technology appraisal guidance 278: Omalizumab for treating severe persistent allergic asthma (review of technology appraisal guidance 133 and 201).

9. NICE. 2009. NICE technology appraisal guidance 169: Sunitinib for the first-line treatment of advanced and/or metastatic renal cell carcinoma.

10. NICE. 2009. NICE technology appraisal guidance 181: Pemetrexed for the first-line treatment of non-small-cell lung cancer.

11. NICE. 2013. Guide to the methods of technology appraisal. London, 2013.

12. Department of Health. 1999. Faster access to modern treatment; how NICE appraisal will work. Leeds: NHS Executive.