NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

APPEAL HEARING

Advice on pirfenidone for treating idiopathic pulmonary fibrosis (review of TA282) [ID837]

Decision of the panel

Introduction

1. An Appeal Panel was convened on 2 December 2016 to consider an appeal against NICE's final appraisal determination on advice to the NHS on pirfenidone for treating idiopathic pulmonary fibrosis (review of TA282).

2. The Appeal Panel consisted of:

• Dr Jonathan Fear	Chair
• Dr Mark Chakravarty	Industry representative
• Dr Biba Stanton	NHS representative
• Patrick Storrie	Lay representative
• Tim Irish	Non-executive director, NICE

3. Tim Irish declared that he had previously held shares in a company with an interest in lung imaging. Since April 2015 these investments are managed through a blind trust. This was a personal financial non-specific interest and as such does not preclude involvement in the appeals panel. None of the other appeal panel members had conflicts of interest to declare.

4. The Panel considered an appeal submitted by Roche Products Ltd.

5. Roche Products Ltd was represented by:

• [redacted]	Head of Health Economics & Strategic Pricing
• Kevin Jameson	Group Health Economics Manager
• Dr James Mawby	Country Medical Lead – Rare Diseases
• Victoria Wakefield	Counsel, Brick Court Chambers
• Sarah Ellson	Partner, Fieldfisher LLP

6. In addition, the following individuals involved in the appraisal were present and available to answer questions from the Appeal Panel:

• Dr Amanda Adler	Chair, Technology Appraisal Committee B
• Meindert Boysen	Programme Director - CHTE
• Melinda Goodall	Associate Director - Committee B

• Professor John Cairns	Technology Appraisal Committee B representative
• Dr Nicky Welton	Technology Appraisal Committee B representative
• Sophie Cooper	Technical Analyst - CHTE

7. NICE's legal adviser Stephen Hocking (DACBeachcroft LLP) was also present.

8. Under NICE's appeal procedures members of the public are admitted to appeal hearings and several members of the public were present at this appeal.

9. There are two grounds under which an appeal can be lodged:

   Ground 1: In making the assessment that preceded the recommendation, NICE has:
   a) Failed to act fairly
   b) Exceeded its powers.

   Ground 2: The recommendation is unreasonable in the light of the evidence submitted to NICE.

10. The Vice Chair of NICE (Mr Andy McKeon) in preliminary correspondence had confirmed that:

    - Roche had potentially valid grounds of appeal as follows:
      - Ground 1a – NICE has failed to act fairly.
      - Ground 2 –the recommendation is unreasonable in the light of the evidence submitted to NICE.

11. Idiopathic pulmonary fibrosis is a chronic, progressive lung disease in which scarring (fibrosis) occurs. The cause is unknown, but it is thought to be related to an abnormal immune response. Symptoms may include breathlessness and cough. Over time, people can experience a decline in lung function, reduced quality of life, and death. The median survival for people with idiopathic pulmonary fibrosis in the UK from the time of diagnosis is approximately 3 years. People with mild-to-moderate disease live longer than people with severe disease.

12. The appraisal that is the subject of the current appeal provided advice to the NHS on the use of pirfenidone within its marketing authorisation for treating idiopathic pulmonary fibrosis.

13. Before the Appeal Panel inquired into the detailed complaints the following made a preliminary statement: Victoria Wakefield on behalf of Roche and Dr Amanda Adler on behalf of the Appraisal Committee.

14. Victoria Wakefield for Roche stated that Roche's arguments essentially all related to the same overarching error by the Committee: there was a failure to consider the totality of the data. More specifically:

    a. The Appraisal Committee had failed to answer the relevant question for the appraisal as set out in the scope by focusing on the subgroup of patients with Forced Vital Capacity (FVC) 80-90% predicted rather than all patients within the marketing authorisation.
    b. In defining this subgroup, the Appraisal Committee had arbitrarily and unjustifiably drawn a line through the whole population that did not have a scientific basis but was defined solely by NICE's own previous guidance.
    c. There was no statistically robust data to support the consideration of this subgroup.
    d. This approach to defining a subgroup was in breach of the Institute's policies, in particular, paragraph 5.10 of the Guide to the methods of technology appraisal 2013 (the Methods Guide).

15. Dr Amanda Adler for NICE, explained that:

    a. The scope of the appraisal was to examine the use of pirfenidone within its marketing authorisation for mild to moderate pulmonary fibrosis as well as considering subgroups by disease severity defined by FVC (such as above or below 80%). She stated that the Appraisal Committee had looked at the totality of data for patients within the marketing authorisation but had decided that it was more appropriate to consider subgroups separately because looking at the whole group might mask important differences in cost-effectiveness within the group. An average value for cost effectiveness was not the right approach in this case.
    b. The Appraisal Committee felt that examining the 80-90% subgroup was relevant because current practice is to treat patients with FVC 50-80% predicted. A previous committee had found it reasonable to use an 80% FVC threshold. The appellant itself had used an 80% threshold. The 80% value was arbitrary, but reflected current NHS practice. 90% was chosen because there was very little data for patients with FVC> 90% (and the new data included no one with FVC>90%).
    c. The Committee chose not to reverse the recommendation for the 50-80% group - although there was uncertainty surrounding it - because it felt that this would be unfair. The Committee therefore chose to concentrate on a subgroup to whom treatment could potentially be extended.
    d. The model had not assumed any difference in the clinical effectiveness of pirfenidone between the subgroups. However, it was judged that the greater cost of treating patients with milder disease could result in important differences in cost-effectiveness between these subgroups.
    e. The methods guide encourages the consideration of subgroups if there are potential differences in clinical or cost-effectiveness between them.

Appeal Ground 1a: In making the assessment that preceded the recommendation, NICE has failed to act fairly: *In failing to consider the totality of data in respect of 'adults with mild to moderate idiopathic fibrosis' (which is both the full licensed indication and the relevant population as identified in the final scope), and in particular determining that 'the subgroup with a FVC between 80% and 90% predicted was the relevant population for decision making (para 4.5 FAD) the Committee acted contrary to policy and procedures (in particular paragraphs 3.2.2, 5.1.4, 5.10 and 6.2.18 of the Methods guide) with inadequate reasons and unfairly.*

16. The appellant raised its concerns about this aspect of the appraisal under both ground 1(a) and ground 2.1. As there was no convenient way of disentangling the issues under each ground, the appeal hearing considered them together, and this letter will do likewise.

17. The Panel questioned the Committee as to what it considered the relevant population to be. Dr Amanda Adler replied that they took the population as defined in the scope, but noted that the scope said they were to look at subgroups if the data permitted. They then chose not to withdraw treatment from the existing treated population, (50-80% FVC) and published an Appraisal Consultation Document concluding that pirfenidone was not a good use of NHS resources in the mild patient population (>80%). The core question was whether pirfenidone was a good use of resources when compared to best supportive care in that population.

18. Dr Amanda Adler confirmed that the Committee had reconsidered the prior decision in the 50-80% FVC subgroup.

19. The Committee were asked about their view on the clinical effectiveness of pirfenidone in the different subgroups, with reference to their statement in the FAD (4.10) both that it was "not clear whether pirfenidone was more, less or equally effective" and that "the committee agreed that...it was more likely to be less effective". Sophie Cooper stated that the economic model used took the same hazard ratios for the 50-80% subgroup and for >80%: the same clinical effectiveness was assumed in all subgroups. Dr Amanda Adler stated she did not feel there was an inconsistency. Professor Cairns said there was uncertainty in all subgroups, but that for the >80% group it was more likely that treatment was less effective.

20. The Appraisal Committee was asked by the Panel whether it was the case that there was no substantial difference between the ICERs for the whole population (50-90% predicted FVC) versus the 50-80% predicted FVC subgroup. Dr Amanda Adler replied that the data for the FVC 80-90% group showed ICERs higher than the NICE threshold.

21. Professor John Cairns was asked by the Panel whether with ICERs for the whole population within, but at the top of, the NICE threshold (and with considerable uncertainty) the Appraisal Committee had felt it was within their remit to look for groups which might be less cost-effective. He replied that this was correct. He added that the Committee were aware that in the NICE appraisal of nintedanib, nintedanib had shown extended dominance over pirfenidone, although the

Committee appreciated that different manufacturers took different approaches to economic modelling.

22. Dr Amanda Adler added that it was not unusual to look for subgroups even if treatment in the whole population appeared cost effective. If the Committee had been starting anew the approach might have been different, but existing guidance recommended treatment for the 50-80% group. The Committee felt steered to look at mild disease. She felt that for an extension to guidance you would look at the extension group separately.

23. Professor John Cairns was asked to consider the modelling results that appeared to show that the cost effectiveness for the 50-80% group was essentially the same as for the 50-90% group. He explained that that was a standard result: the larger group masks the cost ineffectiveness of a smaller subgroup. The Committee believed that use would be cost-ineffective in the 80-90% group. Dr Amanda Adler added that the Committee knew that all the ICERs were likely to be underestimates because of its concern about the modelling of the stopping rule.

24. Meindert Boysen added that if the Committee knew that the ICERs for 50-80% and 50-90% were essentially the same, and that the ICER for the >80% group was above £30,000, then it was reasonable to recommend treating only patients with moderate disease. If those analyses did not exist would the Committee have called for them? - he felt probably not. If treatment was cost effective overall it would not be reasonable to "dig" for subgroups in which a treatment was not cost effective.

25. Sophie Cooper explained that the reason that cost effectiveness differed between the moderate and mild populations, despite an assumed equivalent clinical effect, was that the groups differed in how long they remained progression free, how long they were on treatment, and so on. The mild group received more benefit but incurred more cost and so received less favourable ICERs. Dr Amanda Adler added that there were far more data in the 50-80% group than the 80-90% group, and so it was more possible to be certain of the results for the 50-80% group.

26. The Appeal Panel appreciates that the Committee had an unusual task before it. It is also very wary of being seen to lay down general rules. That is not its role, and it cannot anticipate every twist and turn of a future appraisal. However, it has concluded that the Committee's approach was erroneous on the facts of this case, and it must give its reasons to guide the Committee on any reconsideration there may be.

27. In every appraisal, the starting point to define the question put to the Committee is the same: the scope. In this case the scope was in conventional form. The Committee were charged to " *appraise the clinical and cost effectiveness of pirfenidone within its marketing authorisation*". Only under "other considerations" did the scope record: " *If evidence allows, subgroup analysis by disease severity, defined by FVC (such as above and below or (sic) 80% FVC) and/or diffusing capacity for carbon monoxide, will be considered*".

28. Unless a scope specifies otherwise, the Appeal Panel considers that there is a soft presumption that the starting point for any Committee should be consideration of the whole patient group as one, with a view to making one recommendation for that group. Where different recommendations are to be made for different groups of patients, the reason for departing from one recommendation should be clear and adequate. The Panel does not suggest that this should be a particularly high hurdle to surmount.

29. In this case the Appeal Panel felt that the documentation and the evidence it heard in the hearing did not show that this Committee had adequately considered the possibility of a single recommendation before considering subgroups. The reasons are:

    a. Nowhere in the FAD is there a discussion of the whole patient population position. Although FADs are not to be read as if they are legal documents, and it is possible to supplement the reasoning in a FAD from other Committee papers, on such a central issue some discussion is needed in the FAD if the Panel is to be convinced that adequate consideration was given in Committee.
    b. The reference in the FAD committee meeting slides to a group of patients with FVC of >50% being the result of “combining subgroups" (rather than being the licensed population) suggested subgroups were being taken as the default position, and that it was a departure from that position that would need to be justified.
    c. The reference in the appeal hearing to considering an "extension to guidance" also suggested subgroups were being taken as the default.
    d. The reference in the appeal hearing to be "charged with looking at subgroups" also suggested that the Committee misunderstood its remit and may have considered that a recommendation based on subgroups was a requirement, rather than a question for its judgement.

30. As far as the reasonableness of considering subgroups is concerned, the Panel tended to agree with Meindert Boysen that in a case where it appeared that use of a product was acceptably cost effective in a whole population, it would not normally be reasonable to look for subgroups within that population where use was cost ineffective. However, it would go too far to make that a general rule. Hypothetically if a committee was aware that there existed an identifiable subgroup defined for a proper purpose and in a logical way and in which use was clearly not cost effective, then it might be difficult to say that taking account of that subgroup was unreasonable.

31. In this case the Panel was not yet persuaded that it was reasonable to divide this patient population into subgroups. This is because the modelled difference in cost effectiveness between the group 50-80% and 50-90% was small, and the Committee's conclusions on the 80-90% group were tentative at best. However, the Panel points out that it has already found above that the Committee's reasoning for and consideration of the use of subgroups is unfair, and as a result the Panel may not be fully sighted of the Committee's reasons. While the Panel finds the use of subgroups to be unreasonable on the evidence presented, it

does not rule out that in future a more fully reasoned approach to using these or other subgroups might be reasonable. That would be a matter to be considered at the time.

32. Turning to other considerations under this appeal ground, the Panel was concerned to learn that the Committee had taken account of the results of an economic model presented in another appraisal. The Panel does understand that committees conduct themselves administratively and not judicially. They are expected to bring their wide experience of relevant matters to an appraisal, and there is no need for them to spell out any of this background knowledge. However, they must conduct an appraisal on the evidence before them in that appraisal. If they rely on the economic modelling of the product being appraised that was presented in another appraisal to explain their doubts about the results of the modelling in the appraisal in front of them, then as a bare minimum they must say that this is in their minds before consultation. The Panel would have found that the failure to do so in this case rendered the appraisal unfair of itself.

33. Subject to its findings about the fairness and reasonability of using subgroups at all in this appraisal, the Appeal Panel considered that 80% predicted FVC was an acceptable threshold for defining subgroups that might have different effectiveness or cost-effectiveness. This threshold reflects clinical practice. The Panel was not persuaded that it was significant that that practice might be shaped by past NICE guidance; the Committee is entitled to take clinical practice as it finds it. Further the Panel accepted that the 80% threshold was "arbitrary", in the sense that there was no underlying clinical or scientific rationale for that instead of than some other value, rather, it seemed that if a line had to be drawn, then a line at 80% FVC was "as good as any". The Panel did not feel that such an approach was per se unfair or unreasonable; no doubt many thresholds for many biological markers are in clinical use on much the same basis.

34. The Appeal Panel did not consider it unfair to consider these subgroups on the grounds that there was insufficient data to analyse them with a high degree of certainty because an assumption of no difference in clinical effectiveness for the two subgroups was used in the model. Furthermore, the Panel considers that if a Committee has fairly and reasonably decided that consideration of subgroups is needed, then the fact that data is restricted for one or more subgroups would rarely if ever invalidate the decision to look at subgroups in the first place. Inevitably the data available to analyse a subgroup will be less than the data available to analyse the whole patient population, and this may have an impact on certainty.

35. The Appeal Panel did not agree with the Appellant on the effect of the NICE methods guide on the definition of subgroups. It believes that the intent of the methods guide advice on the definition of subgroups is to avoid data dredging, and the guide's advice on acceptable and unacceptable definitions for subgroups has to be seen in that light. The essential mischief that must be avoided is creating subgroups for the purpose of producing a particular outcome. A subgroup with a plausible biological basis will be one way to achieve this, but there will be others permitted by the guide. In this case, there is nothing suspect about a subgroup using an 80% cut off because the reason for that cut off can be

found in current clinical practice, nor was a 90% cut off suspect as that could be seen to derive from the trial data. The Panel concluded that the methods guide does allow NICE to use a treatment threshold to define clinical subgroups.

36. In conclusion, the Appeal Panel considered the Committee acted unfairly because they did not demonstrate consideration of the whole population as defined in the scope as one population before considering the use of subgroups, and their use of subgroups (on the reasoning currently presented) was unreasonable.

37. The Appeal Panel therefore upheld the appeal on this point under both grounds 1(a) and 2.1

Appeal Ground 1b: In making the assessment that preceded the recommendation, NICE has exceeded its powers.

38. There was no appeal under this ground.

Appeal Ground 2: The recommendation is unreasonable in the light of the evidence submitted to NICE.

Appeal point Ground 2.1: *Failing to consider the totality of data in respect of adults with mild to moderate idiopathic fibrosis (which is both the full licensed population and the relevant population in the Final Scope) and in particular determining that the 'sub group of people with an FVC between 80% and 90% predicted was the relevant population for decision making' (para 4.5 of the FAD), was perverse.*

39. Considered and upheld above.

Appeal point Ground 2.3: *The Committee's assessment of clinical effectiveness was perverse.*

40. Victoria Wakefield for Roche stated that:
    a. there was no robust evidence of a difference in the clinical effectiveness of pirfenidone for patients with FVC 50-80% versus >80% predicted.
    b. the FAD is internally inconsistent in stating both that "it was not clear whether pirfenidone was more, less or equally effective in the group with FVC above 80% predicted" but also that it was "more likely to be less effective" in this group.

41. Dr Amanda Adler for NICE explained that:
    a. the statement that pirfenidone was "more likely to be less effective" in this group was based on a pre-specified analysis from the CAPACITY trials suggesting better outcomes in the placebo group amongst patients with FVC >80% predicted but she acknowledged that this result was not statistically significant.
    b. the Appraisal Committee agreed that there was no robust evidence of a difference in clinical effectiveness between the subgroups.

c. the model had assumed equivalent clinical effectiveness for the subgroups, so any differences in cost-effectiveness were driven by the higher cost of treatment in patients with FVC >80% predicted.
d. numbers were small in the >80% population: a statement that there was no difference in clinical effect between the two groups could be a statement that there was in fact no difference, or that there were insufficient patients to establish a difference. The Committee lacked confidence in a claim that there was an equal effect. The FAD would have been better worded to refer to patient numbers being too small to demonstrate a difference.

42. The Appeal Panel noted above that a FAD must not be read as it if is a legal document. However, it is an important document that records the final recommendation and reasons of a Committee. In this case the FAD presents a confused picture as to what the Committee considered the efficacy of pirfenidone to be in the two subgroups, and its reasons for its conclusions. The Panel concluded that the apparent inconsistency between different statements in the FAD was unreasonable.

43. The Appeal Panel therefore upheld the appeal on this point.

Conclusion and effect of the Appeal Panel's decision

44. The Appeal Panel therefore upholds the appeal on the grounds that NICE has failed to act fairly and that the recommendation is unreasonable in the light of the evidence submitted to NICE.

45. The Appeal Panel suggests that the appraisal is remitted to the appraisal committee who must take all reasonable steps to demonstrate consideration of the effectiveness and cost-effectiveness of pirfenidone in the whole population as set out in the scope. Subgroups defined by predicted FVC could be considered if the treatment is not judged cost-effective in the whole population. The Appraisal Committee's assessment of the clinical effectiveness of pirfenidone in any subgroups should be clearly documented, including any uncertainty in the available evidence.

46. There is no possibility of further appeal against this decision of the Appeal Panel. However, this decision and NICE's decision to issue the final guidance may be challenged by applying to the High Court for permission to apply for a judicial review. Any such application must be made within three months of NICE publishing the final guidance.