Prucalopride for the Treatment of Chronic Constipation in Women: Single Technology Appraisal

Response to consultation from Norgine Pharmaceuticals Ltd

Comments on cost-effectiveness assessment

There are a number of serious concerns regarding the evaluation process and the provisional recommendations made by the Appraisal Committee. This report is structured using the central tenents of the NICE process which we do not believe have been adequately demonstrated within this particular appraisal. The three specific issues to be considered are:

1. Completeness
2. Transparency
3. Robustness/Validity

Deficiencies in both the completeness and transparency of key aspects of the evaluation and appraisal process mean that that the validity of the cost-effectiveness results informing the Committee's provisional recommendations cannot be concluded to have been robustly demonstrated. Furthermore, the economic model submitted by the manufacturer appears to have serious limitations since it fails to produce results that are consistent with the trial data itself (i.e. the model appears to have poor internal validity) or those which could have been derived by applying a more logical 'common sense' approach.

If these issues are not adequately addressed before the Committee makes its final recommendations, there would appear to be a strong case for appeal based on: *Ground Two: The Institute has formulated guidance, which cannot reasonably be justified in the light of evidence submitted.*

1. Completeness

The manufacturer submission fails to adequately address the relevant decision problem and does not provide an appropriate basis for cost-effectiveness considerations. There are significant differences in the decision problem addressed by the manufacturer submission and the final scope agree with NICE and this has serious implications on the validity and utility of the draft guidance (Table 1). [1]

Item	Agreed scope	Manufacturer submission	Significance
Population	Women with chronic constipation in whom standard laxative regimens have failed to provide adequate relief, and for whom more invasive procedures, such as direct rectal intervention, are being considered	Women with chronic constipation in whom standard laxative regimens have failed to provide adequate relief	Significantly altered position of prucalopride in therapy that has resulted in the clinically opaque draft guidance.
Comparators	Standard therapy without prucalopride Invasive procedures such as rectal interventions (including enemas, suppositories and manual evacuation) Bowel surgery	Placebo	Comparator used in the submission does not represent clinical practice and as such does not provide a robust measure of cost-effectiveness. See section 1 below for further discussion.
Outcomes	Proportion of patients with ≥3 SCBM per week Number of spontaneous complete bowel movements per week Improvement in symptoms of constipation Adverse effects of treatment Health-related quality of life (HrQoL)	In terms of HrQoL, EQ-5D was not collected in the studies and thus not reported in the studies.	It is unclear why EQ-5D was not included in the trials even though it is mandated by NICE and given that SF-36, PAC-Sym and PAQ-QOL were administered. Unclear why SF-6D not considered in mapping. Resulted in opaque utility mapping which formed the basis of the QALY estimate. See section 2 below for further discussion. Adverse events are not addressed in the economic modelling and may have an impact on the QALY

Table 1: Comparison of final scope and submitted decision problem

The Committee's recommendations are based on an inappropriate comparator and there appears to have been no formal (or informal) consideration by the Appraisal Committee of the most likely ICER of prucalopride against an appropriate standard of care (including an option of continued use of different laxatives). Furthermore, it could be argued that the cost-effectiveness of prucalopride compared to an appropriate standard of care cannot reasonably be inferred from the results presented given the limited data considered by the manufacturer and deficiencies in the model.

Section 4.11 of the ACD states that:

"The Committee was therefore persuaded that the most plausible ICER compared with placebo plus rescue medication was likely to be below £20,000 per QALY gained".

However, clearly a comparison with placebo plus rescue medication does not provide an adequate basis for informing cost-effectiveness considerations for prucalopride in the context of current NHS practice. Furthermore, including this as the only comparator in a cost-effectiveness analysis contradicts existing NICE methods guidance which states:

"Relevant comparators are identified, with consideration given specifically to routine and best practice in the NHS (including existing NICE guidance) and to the natural history of the condition without suitable treatment" Ref: Section 2.2.4 of the 'Guide to methods of technology appraisal' [2]

Similar concerns regarding the appropriateness of the comparator appear to have been expressed by NHS representatives during the meeting itself (Section 4.5 of the ACD) and were also highlighted by the Evidence Review Group (ERG) in their independent report:

"The comparator group in the model is inappropriate because a key assumption of the economic model (p130 of the submission) is that 'Placebo data from the prucalopride clinical trial were taken as an approximation for the efficacy of response for patients on laxatives'. If the model is considering laxative treatment as the comparator to prucalopride, the comparator data taken from trials is inappropriate. Placebo rescue therapy was not equivalent to laxative treatment that would be used in clinical practice since it was limited to one laxative and patients had restricted access to that treatment" (ERG report, p45-46).

Despite these concerns, and lack of accordance with the Institute's methods guide, the Committee's conclusions were clearly based on an inappropriate comparator since reference is only made to the ICER against a placebo comparator. While the Committee appear to have discussed the use of placebo controls in undertaking regulatory trials and the difficulties in defining a standard laxative regimen as a comparator, neither of these issues should lead to a conclusion that a comparison with placebo in the context of reimbursement decisions is appropriate or that some attempt to formally compare against a relevant standard of care is not possible. Indeed, one of the advantages of the decision-analytic

framework underpinning the NICE evaluation process is that these difficulties can be explicitly considered using different assumptions and scenarios. These assumptions and scenarios could have been presented to the Committee and discussed with expert clinical input in determining the most likely ICER of prucalopride compared to an appropriate standard of care. In the absence of these ICER estimates, it is not possible to conclude using the current evidence submitted to the Committee that prucalopride represents an efficient use of NHS resources.

2. Transparency

The current NICE methods guide states that:

"Providing an all-embracing definition of what constitutes a high-quality model is not possible but some guidelines are available. In general, all structural assumptions should be fully justified, and data inputs should be clearly documented and justified in the context of a valid review of alternatives" Ref: Section 5.7.3 of the 'Guide to methods of technology appraisal' [2]

There appear to be significant problems with transparency for key assumptions and inputs applied in the model, leading to significant uncertainties regarding the validity of the model and the associated results.

There appears to no systematic approach to the identification and inclusion of studies informing the model, thus increasing the likelihood of bias in the submission. Critically, ten studies are identified as relevant to the appraisal but only three form the basis of the clinical effectiveness assessment. Further, meta-analysis was not performed. This is in clear contravention to the NICE methods guide. [2]

The ERG report clearly identifies the regression equations used to determine the treatment effects (including both the clinical effectiveness and the mapping of patient outcomes to EQ-5D) applied in the model as the key input parameters. However, they also conclude that it is not possible to verify these using the evidence which has been submitted by the manufacturer. This seems to represent a fundamental challenge to demonstrating the robustness of any subsequent results. Furthermore, it could be reasonably argued that rather than simply seeing these equations as an input to the model, the patient level data and coding used for the regressions and quality of life mapping actually comprise key structural elements of the model itself. Indeed, the Excel model provided by the manufacturer really only represents a front-end interface, allowing a user to re-run a pre-determined set of structural assumptions and parameter estimates. Without providing appropriate access to both the patient level data and the coding used to estimate the equations/utility mapping, it is clearly not possible f to provide any adequate assessment of the validity of the model structure, the inputs or the associated results.

The lack of transparency also relates to key parameters and assumptions, including:

1. **Inadequate justification for mapping between PAC-QOL to EQ-5D rather than using the SF-36 data generated in the trials to estimate utility and failure to demonstrate robustness of results to alternative approaches.** The only justification provided by the manufacturer was that SF-36 was 'limited' and would not provide a robust mapping. However, it is unclear in what respect this data was limited, particularly since SF-36 data was collected in the 3 main studies which were also used to estimate response data. Furthermore, we would challenge anyone to be able to adequately describe or critique what has actually been done by the manufacturer in the mapping exercise based on the limited descriptions and data provided. It is peculiar that SF-36 results were used in the mapping process but did not contribute to the derived EQ-5D scores. This may be highly significant given the lack of statistical significance for the SF-36 scores save in the PRU-INT 6 study for the physical component score at week 4.

   In their response to the concerns expressed in the ERG report, the manufacturer claims that a separate analysis was undertaken by mapping the SF-36 results contained in the trial to EQ-5D utility values. However, we can find no reference to the results of this analysis or to any discussion by the committee of these results. In the absence of these results being presented and critiqued by the ERG, it is impossible to conclude that the statements made by the manufacturer are based on empirical evidence or are simply conjecture.

   A key omission from the appraisal is any discussion of the appropriateness of the mapping approach to EQ-5D and the possibility of using the SF-36 results from the trial to directly estimate SF-6D utility data. It is unclear why this approach was not also considered given the inherent uncertainty in the methods used in the main analysis. While the SF-6D is not currently part of the reference case approach, it would appear to be a reasonable alternative scenario to have presented to support the robustness of the results given the methodological uncertainty surrounding existing mapping approach.

   The validity of the mapping approach itself is discussed further in section 3.

2. **Failure to consider alternative approaches and assumptions in deriving QALY estimates.** As noted above, it is clear that the robustness of uncertainty surrounding the mapping process, and related assumptions, applied to estimate QALY gains have not been adequately explored. Furthermore, the extrapolation of these estimates over a longer-term

horizon (12-52 weeks) represents another key assumption which is subject to considerable uncertainty which we do not consider has been adequately assessed.

There appear to be two important assumptions applied in the longer term extrapolation: (i) responders to prucalopride maintain their initial short-term utility gain over the longer term horizon; (ii) the initial short-term utility gain reported for the control group is considered to be a short-term 'placebo' effect, which is assumed to wane over the longer-term horizon (such that at 52 weeks patients in the control group have the same utility as reported at baseline). Neither of these assumptions is supported by appropriate empirical evidence and neither assumption is subjected to sensitivity analysis by the manufacturer. The only empirical evidence in support of assumption (i) comes from long-term open label extensions measured using PAC-QOL. The actual data reported is marked AIC and has been removed from the manufacturer's response. However, the absence of a control population and the lack of comparable data demonstrating that the SF-36 data remains stable, means that this assumption is clearly subject to additional uncertainty. Given the uncertainty surround the extrapolation of QALY gains, alternative scenarios should have been presented by the manufacturer as recommended in the NICE methods guide:

"Alternative scenarios should be considered to compare the implications of different assumptions around extrapolation for the results. For example, the duration of treatment effects scenarios might include when the treatment benefit in the extrapolated phase is: (i) nil; (ii) the same as during the treatment phase and continues as the same level; or (iii) diminishes in the long term" Ref: Section 5.7.3 of the 'Guide to methods of technology appraisal' [2]

As it stands, only the most optimistic scenario for prucalopride seems to have been considered. Furthermore, this optimism is compounded by the separate assumptions made for the long-term quality of life for the comparator group, where the 'waning' of the effect is assumed to reflect a temporary 'placebo' effect. However, it has been demonstrated that estimated cost-effectiveness and associated policy decisions may be sensitive to the assumptions regarding the mechanism underlying placebo responses and, in the absence of other evidence, additional sensitivity analysis should be undertaken.[3] In summary, the assumptions employed by the manufacturer appear overly optimistic towards the incremental QALY gain for prucalopride and no additional sensitivity analyses are presented to the Committee.

The general lack of transparency in the manufacturer's model was a key issue identified by the ERG who concluded:

"If the regression results are to be believed, it is possible that prucalopride is cost-effective. However, the lack of transparency in the results from the 10 prucalopride trials and studies feeding into the economic model and the lack of transparency over the EQ-5D mapping means that it is not possible to establish a more accurate estimate of cost effectiveness." (ERG report, p9)

Some of these concerns appear to be noted by the Committee in paragraph 4.1 of the ACD. However, despite these concerns the Committee concluded that "the ERG had shown the manufacturer's cost-effectiveness estimates to be reasonably stable under varied assumptions" (Section 4.10, ACD p21). There even seems to have been some suggestion that the Committee considered that the results may actually have been conservative since the true costs associated with treating chronic constipation were not included. We do not feel this to be a reasonable conclusion for this appraisal. We believe that the conclusions may not be robust and that a more accurate ICER estimate (i.e. with higher internal validity) could easily be in excess of £30,000 per QALY, particularly given the univariate sensitivity analyses undertaken by the ERG.

3. Robustness/Validity

We consider the Committee conclusions potentially perverse in light of:

(i) The significant uncertainty inherent in key assumptions;

(ii) The poor internal validity of the model; and

(iii) The series of optimistic assumptions applied by the manufacturer to the QALY gain estimates.

Indeed, it is evident from the sensitivity analyses undertaken by the ERG that the assumptions for the QALY gain of prucalopride are a key driver of cost-effectiveness and, importantly, that the cost-effectiveness conclusions do not appear robust to the alternative assumptions considered by the ERG (e.g. QALY gain reduced by 50% to 75%). These findings are important since the manufacturer did not adequately justify their own assumptions in relation to a review of reasonable alternatives, nor did they present appropriate sensitivity analysis demonstrating the robustness to several key inputs/assumptions.

It is particularly concerning that the current model results do not appear to match those which could have been derived by applying a more logical 'common sense' approach by the Committee. That is, given the uncertainties noted, how well does the economic model appear to predict the main trial results? The lack of internal validity of the model results is evident in Figure 1 which replicates Figure

8 of the manufacturer submission (also reproduced as Figure 4 in the ERG report) entitled "Costs and outcomes of constipation treatment – decision tree analysis, UK model".

The blue area represents the estimate of the additional QALY gain achieved with prucalopride compared with placebo predicted by the model. We have previously argued that the assumptions employed by the manufacturer appear optimistic towards prucalopride over the longer-term horizon. This is clearly demonstrated from the *predictions* arising from the economic model that appear to have low internal validity in relation to the *actual* outcomes reported in the main trials.

Tables 2-9 of the ERG report consistently demonstrate that the mean change from baseline for measures of response, PAC-SYM, PAC-QOL and SF-36 data for weeks 1-12 with placebo was approximately half that of the group treated with prucalopride. Hence, a common sense 'mapping' approach to utility and QALY gain might reasonably conclude that the utility gain (compared to baseline) for placebo would be approximately half that of prucalopride. However, it is clear from Figure 1 that this common sense approach is not matched by the predictions of the model at 4 or 12 weeks. The resulting low internal validity of these findings does not appear to have been adequately explained. In the absence of a clear explanation it is possible that the QALY gain estimates with prucalopride may have been over-estimated by the manufacturer by as much as 50%.

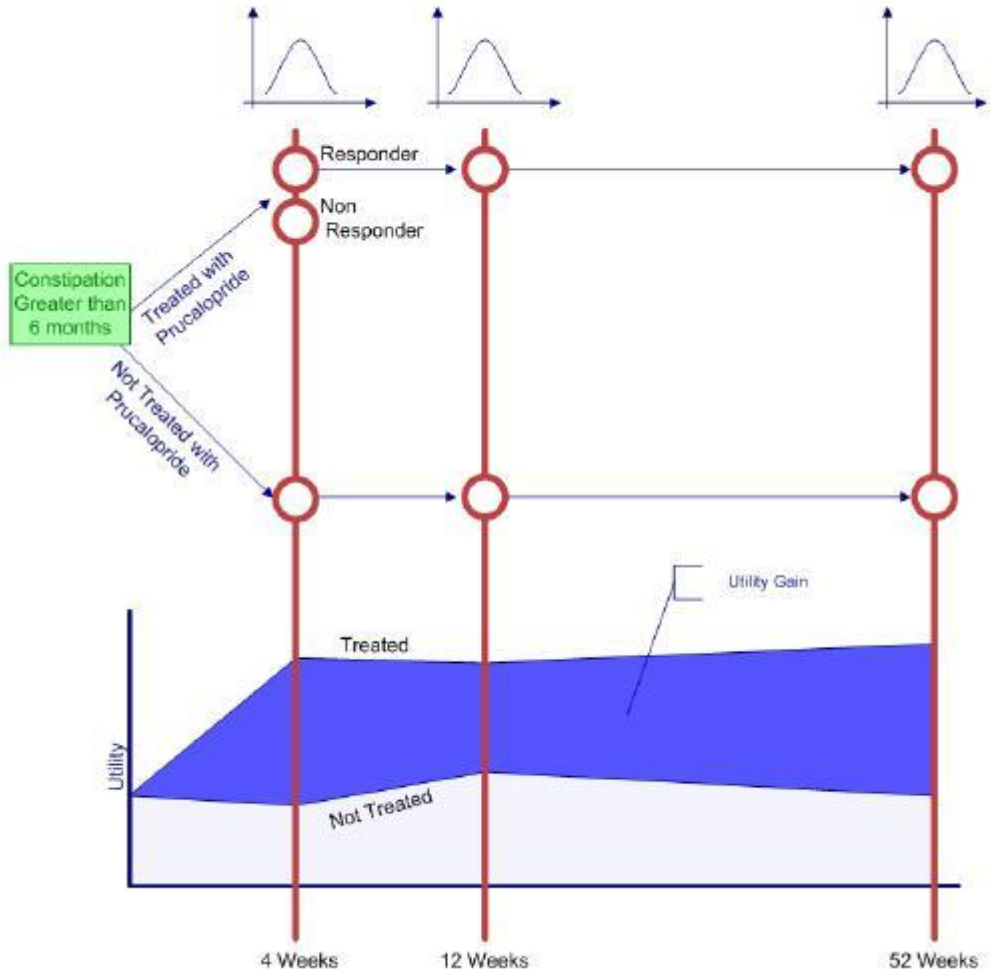

Figure 1: Replication of manufacturer's schematic of model results

Faced with this inconsistency between the trial findings and the model predictions, we do not consider it appropriate to conclude that the subsequent cost-effectiveness estimates are likely to be robust. It naturally follows from the QALY gain calculations, that it is entirely plausible that the cost-effectiveness estimates reported by the manufacturer could also have been over-estimated by as much as 50%. This obviously has clear implications for the robustness of the Committee's provisional decision. Indeed, sensitivity analyses were presented in the ERG report assuming a 50% reduction in QALY gain (Table 15, p 71 of the ERG report) where the ICER estimates were reported to be either above or close to a £30,000 threshold. These specific results do not appear to be mentioned in the ACD.

Faced with such uncertainty, and applying a 'common sense' logic, it seems difficult to accept either the conclusions of the Appraisal Committee, that the most likely estimate of the ICER is likely to be below £20,000, or the ERGs conclusion that while this estimate is likely to be optimistic it is not possible to establish a more accurate estimate of cost effectiveness. It would seem reasonable to

conclude that the most likely estimate of the ICER is much more likely to be closer to a £30,000 threshold (or above) than it is to the £20,000 threshold.

Without providing appropriate access to both the patient level data and the coding used to estimate the equations/utility mapping, it only possible to speculate on what could be causing the low internal validity of the economic model. The most likely explanation lies with the mapping function used to map PAC-QOL to EQ-5D (using SF-36 data as the link between PAC-QOL and EQ-5D). Not only is the mapping process subject to significant methodological uncertainty, it also appears subject to potential bias. Indeed, the same publication cited by the manufacturer to support the mapping process also concludes that:

"Our results suggest that approaches mapping the SF-36 onto the EQ-5D are robust across setting and medical condition but overpredict for more severe EQ-5D states. Our results raise doubts over the suitability of mapping for patient datasets which have a proportion of subjects with poorer health or where dimensions are not represented in the target measure. Potential policy implications are that mapping the SF-36 onto the EQ-5D can be useful, but may not be suitable for all populations. [4]

The risk to validity of this methodology is highlighted by the ERG group who note that the utility value derived from the mapping algorithm for the severe chronic constipation groups of 0.585 differs from the distributions of baseline score provided by the manufacturer in their clarification document. These data are highly skewed towards better quality of life and may not represent individuals with severe chronic constipation (Figure 2).

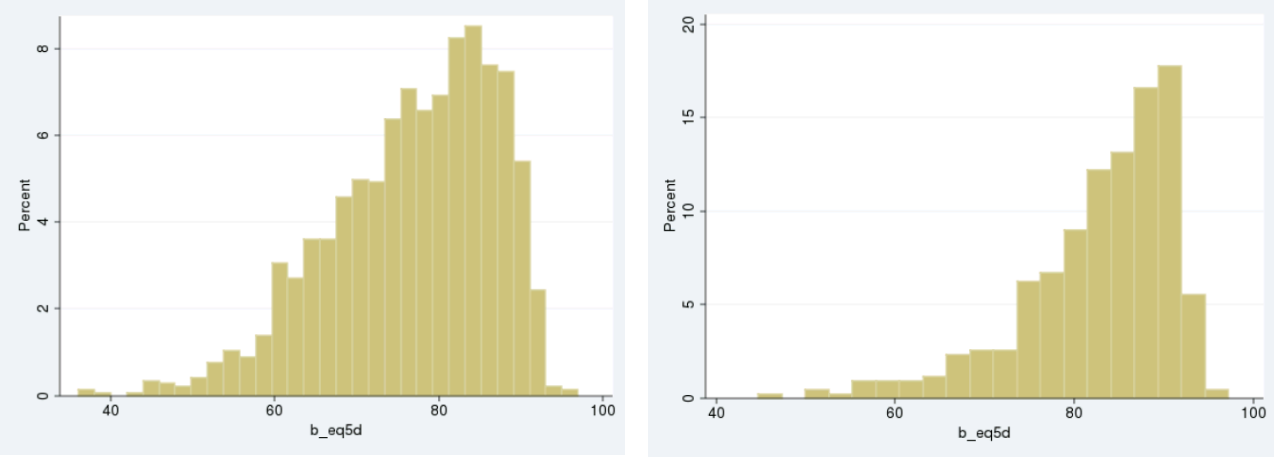

Figure 2: Distribution of baseline EQ-5D (18-65 years old and elderly population)

Given the poor quality of life of the population under consideration, the use of the current mapping approach appears questionable and should be subject to additional investigation before any final recommendations are made. Furthermore, results should also be made available to the Committee using the SF-36 trial data to generate SF-6D utility values.

References

[1] Health Technology Appraisal. Prucalopride for the treatment of chronic constipation in women. Final scope. [Internet] National Institute for Health and Clinical Excellence [Issued February 2010; cited August 2010] Available from: http://www.nice.org.uk/nicemedia/live/12263/47335/47335.pdf

[2] Updated guide to the methods of technology appraisal. National Institute for Health and Clinical Excellence [Issued June 2010; cited August 2010] Available from: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

[3] Hawkins N, Scott DA. Cost-Effectiveness Analysis: Discount the Placebo at Your Peril. Med Decis Making. 2010 Mar 12. [Epub ahead of print]

[4] Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? Health Qual Life Outcomes. 2009 Mar 31; 7:27.