NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Final appraisal determination

Everolimus for preventing organ rejection in liver transplantation

This guidance was developed using the single technology appraisal (STA) process

1 Guidance

1.1 Everolimus is not recommended within its marketing authorisation for preventing organ rejection in people having a liver transplant.

1.2 People whose treatment with everolimus was started within the NHS before this guidance was published, should be able to continue everolimus until they and their NHS clinician consider it appropriate to stop.

2 The technology

2.1 Everolimus (Certican, Novartis Pharmaceuticals UK) is an analogue of sirolimus. It is an immunosuppressant that inhibits the mammalian target of rapamycin (mTOR) protein and targets the primary causes of progressive allograft dysfunction (also known as chronic rejection) following an organ transplant. It has a marketing authorisation in the UK for 'the prophylaxis of organ rejection in patients receiving a hepatic transplant. In liver transplantation, everolimus should be used in combination with tacrolimus and corticosteroids'.

2.2 Everolimus is taken orally. The recommended starting dosage is 1.0 mg twice daily. The first dose is taken approximately 4 weeks after the transplant. The summary of product characteristics states that for people who have had a liver transplant, exposure to tacrolimus should be reduced to minimise calcineurin-related renal toxicity. The tacrolimus dose should be reduced starting approximately 3 weeks after initiating administration together with everolimus, based on targeted tacrolimus blood trough levels of 3-5 ng/ml.

2.3 The summary of product characteristics for everolimus lists the most common adverse reactions as infections, anaemia, hyperlipidaemia, new onset of diabetes mellitus, insomnia, anxiety, headache, hypertension, cough, nausea, peripheral oedema and impaired healing. For full details of adverse reactions and contraindications, see the summary of product characteristics.

2.4 The company reported that the acquisition cost for everolimus is £148.50, £297.00 and £445.50 for the 0.25 mg, 0.5 mg and 0.75 mg packs respectively, excluding VAT. Costs may vary in different settings because of negotiated procurement discounts.

3 The company's submission

The Appraisal Committee (section 8) considered evidence submitted by Novartis and a review of this submission by the Evidence Review Group (ERG; section 9).

Overview of clinical evidence

3.1 The company's systematic review identified 1 randomised controlled trial (RCT) of everolimus for preventing organ rejection after a liver transplant that it considered relevant to the decision problem: trial H2304. The company did not identify any non-RCT evidence that was relevant to the decision problem.

3.2 H2304 was a 24-month multicentre, open-label randomised controlled trial that evaluated the efficacy and safety of everolimus in combination with reduced-dose tacrolimus compared with standard-dose tacrolimus. The trial was mostly conducted in the US with a limited number of UK patients. It included 719 people aged 18 to 70 years who had a primary liver transplant and had started an immunosuppressive regimen, containing tacrolimus and corticosteroids, 3–7 days after the transplant. Patients were randomised at 30 days (±5 days) after the transplant, to 1 of the following treatment arms:

- Arm 1: everolimus with tacrolimus elimination, in which tacrolimus was completely withdrawn by the end of month 4 after the transplant. This treatment arm was stopped early due to a higher rate of acute rejection and treatment discontinuations and was excluded from further discussion in the company submission.
- Arm 2: everolimus with reduced-dose tacrolimus, in which everolimus was started at a daily dose of 2.0 mg. The dose was targeted to maintain a whole blood trough level of 3–8 ng/mL. After everolimus whole blood trough levels were confirmed to be in the target range, the dose of tacrolimus was tapered to achieve a target whole blood trough level of 3–5 ng/mL by 3 weeks after randomisation and continuing for the remainder of the study.
- Arm 3: standard-dose tacrolimus alone, in which tacrolimus trough levels were targeted to be maintained at 8–12 ng/mL until month 4 and then tapered to a target whole blood trough level of 6–10 ng/mL for the remainder of the study.

3.3 Prednisolone was taken at a minimum dose of 5 mg per day for at least 6 months. Before randomisation, 70% of people in both

treatment groups were having mycophenolate mofetil but this was discontinued at randomisation according to the protocol. People having azathioprine or sirolimus were excluded from the study. Baseline characteristics appeared to be similar between the treatment arms. The proportion of people in the trial with hepatitis C virus was 31.8% in the everolimus arm and 31.3% in the standard-dose tacrolimus arm. For hepatocellular carcinoma, the proportions were 17.1% and 14.4% respectively.

3.4 The inclusion criterion for baseline estimated glomerular filtration rate (eGFR, a measure of renal function) was ≥30 mL/min/1.73 m^2. Randomised patients had a mean eGFR of 81 mL/min/1.73 m^2 that, according to the company's clinical expert, is higher than the eGFR levels typically observed in patients in clinical practice in the UK (usually in the range of 50–65 mL/min/1.73 m^2 at the time of liver transplant).

3.5 The primary outcome was a composite of treated biopsy proven acute rejection (tBPAR), graft loss or death at 12 months after transplantation (excluding events before randomisation). This was presented as the Kaplan–Meier incidence rate, with the difference being determined at the 97.5% confidence interval. Secondary outcomes included graft loss, death, number of acute graft rejections and change in renal function measured by eGFR. No patient-related outcomes such as health-related quality of life were measured in the trial.

3.6 Statistical analysis in H2304 was designed to show the non-inferiority of everolimus with reduced-dose tacrolimus compared with standard-dose tacrolimus alone for the composite outcome of tBPAR, graft loss, or death at 12 months after transplantation. A pre-determined non-inferiority margin of 12% was used in the analysis for the primary outcome based on a p value of less than

0.001. Non-inferiority was demonstrated if the upper limit of the 97.5% confidence interval for the difference between the 2 groups was below 12%. For the composite outcome of graft loss or death, a non-inferiority margin of 10% was used.

ERG comments on the clinical evidence

3.7 The ERG considered that all studies relevant to the decision problem were included in the company's submission. The ERG noted that the clinical effectiveness of everolimus relied upon evidence drawn from the H2304 trial, which it considered was of good quality. However, it considered that the efficacy endpoints used in the trial might not be the most appropriate ones. Clinical opinion sought by the ERG explained that, although the number of acute rejections is a relevant endpoint, these are common and easily treated and long-term survival is a more appropriate outcome for evaluating the effectiveness of immunosuppressive therapies. The ERG also noted the lack of a health questionnaire to directly capture patients' health-related quality of life.

3.8 The ERG highlighted that the average whole blood trough levels of tacrolimus were higher than those initially planned for all arms of the trial and that the reduced-dose tacrolimus group showed trough levels above 5 ng/mL throughout the 12 months. Clinical advisers to the ERG explained that a standard target blood level for a tacrolimus regimen in the UK is 6–8 ng/mL until month 1, just above 6 ng/mL until month 4 and between 5 and 6 ng/mL until the end of the first year. The ERG therefore considered that the reduced tacrolimus blood trough levels in H2304 were equivalent to the standard target blood trough levels of tacrolimus in UK practice.

3.9 The ERG considered that the company's overall approach to the statistical analysis of H2304 was generally sound. It highlighted that in non-inferiority trials, the choice of the non-inferiority margin is

crucial. However, the ERG commented that it could not find a justification for the non-inferiority margin used because the company did not explain its decision in the submission.

Clinical trial results

3.10 The company submission included results for the intention-to-treat (ITT) population from H2304 for 12 and 24 months follow up. For the primary composite efficacy endpoint of tBPAR, graft loss or death, everolimus with reduced-dose tacrolimus was statistically non-inferior to standard-dose tacrolimus alone because the upper limit of the confidence interval [CI] was below the 12% non-inferiority margin for this outcome and the p value was reported to be <0.001. At 12 months the Kaplan–Meier survival probability was 93.3% compared with 90.3% (97.5% CI for the difference −8.7 to 2.6; p-value for non-inferiority <0.001). At 24 months the Kaplan–Meier survival probability was 89.7% compared with 87.5% (97.5% CI for the difference −8.8 to 4.4).

3.11 For the composite outcome of graft loss or death, everolimus with reduced-dose tacrolimus was statistically non-inferior to standard-dose tacrolimus alone (the upper limit of the confidence interval was below the 10% non-inferiority margin for this outcome; the 12 month Kaplan–Meier survival probabilities are academic-in-confidence and cannot be presented). At 24 months after transplantation the Kaplan–Meier probabilities were 92.7% and 93.8%, respectively (97.5% CI for the difference −4.2 to 6.4).

3.12 There were statistically significantly fewer episodes of rejection at 12 months in the group randomised to everolimus with reduced-dose tacrolimus compared with the standard-dose tacrolimus group. At 12 months, the Kaplan–Meier tBPAR-free probability was 96.3% in the everolimus group compared with 89.3% in the standard-dose tacrolimus group (95% CI for the difference −11.6 to

−2.5; p-value for equivalence =0.003). At 24 months, the Kaplan–Meier probabilities were 93.9% and 86.7% respectively (95% CI for the difference −13.5 to −0.9; p-value for equivalence =0.01).

3.13 Everolimus with reduced-dose tacrolimus was associated with better preservation of renal function compared with standard-dose tacrolimus alone at 12 and 24 months after liver transplantation. The difference in mean eGFR was 8.50 mL/min/1.73 m^2 (p<0.001, 97.5% CI 3.74 to 13.27) at month 12 and 6.66 mL/min/1.73 m^2 (p<0.0001, 97.5% CI 1.90 to 11.42) at month 24.

3.14 The company presented results for a number of pre-defined subgroup analyses including subgroups based on age, gender, race, eGFR, hepatitis C status, and cause of end-stage liver disease. The company reported that the overall pattern of the event rates within subgroups was similar to that observed in the overall population.

ERG comments on the clinical trial results

3.15 The ERG noted that the composite primary outcome of the trial (tBPAR, graft loss and death) combined 2 outcomes in which the treatment effects of everolimus relative to the comparator worked in different directions. The ERG commented that while the results for the outcome of graft loss or death favoured the standard-dose tacrolimus arm of the trial, the addition of tBPAR to the composite endpoint favoured everolimus with reduced-dose tacrolimus. The ERG questioned the appropriateness of the composite endpoint.

3.16 The ERG highlighted that there were no statistically significant differences between the treatment groups in the rates of graft loss or death at 12 or 24 months after transplantation. However, there were statistically significantly fewer episodes of acute graft rejection in the everolimus group compared with the standard-dose

tacrolimus group. The ERG considered that the effectiveness of everolimus was largely dependent on the choice of clinical outcomes and whether these included acute rejection episodes or graft losses.

Adverse effects of treatment

3.17 The company did not identify any trials that reported adverse events for everolimus with reduced-dose tacrolimus, apart from H2304. In H2304, lipid changes occurred more frequently in the everolimus group than in the standard-dose tacrolimus group (at 24 months, 26.9% compared with 11.6%, RR 15.4, 95% CI 8.5 to 22.2). At 24 months, the incidence of new onset diabetes mellitus was higher in the everolimus group (20.8% compared with 16.5%, RR 4.3, 95% CI 2.6 to 11.2). The company did not report any statistically significant differences in the occurrence at 12 months of diarrhoea, headache, hypertension, wound healing, biliary leaks, new onset diabetes mellitus, infections or renal failure. There were fewer cases of renal failure in the everolimus group (15 compared with 21 in the standard-dose tacrolimus group at 12 months) but no tests of significance were reported.

3.18 The company reported that 74.3% of people in H2304 tolerated everolimus with reduced-dose tacrolimus up to month 12. No patients developed severe renal dysfunction (eGFR <30 mL/min/1.73 m^2). The company commented that evidence from the network meta-analysis indicated that the treatments were comparable in terms of safety.

Network meta-analysis

3.19 In the absence of direct trial evidence, the company did a systematic review and network meta-analysis. This estimated the relative effectiveness of everolimus with reduced-dose tacrolimus

for preventing organ rejection in people having a liver transplant in the maintenance phase, compared with:

- mycophenolate mofetil (in combination with standard-dose tacrolimus, reduced-dose tacrolimus or standard-dose ciclosporin)
- azathioprine (in combination with standard-dose tacrolimus or standard-dose ciclosporin)
- standard-dose tacrolimus.

3.20 The company identified 22 RCTs that assessed the efficacy of these treatments, with or without corticosteroids. It reported that there was heterogeneity between studies that provided challenges to building a feasible network, such as: lack of reporting of characteristics that were potential treatment effect modifiers, variation in the definition of the tacrolimus and ciclosporin arms with respect to dosage, variations in definitions of outcomes, and variation in the duration or use of corticosteroid therapy. However, the company considered that the evidence was sufficient to create feasible networks for 13 of the 16 clinical endpoints extracted from the studies. These included overall survival, graft survival, tBPAR, and renal function.

3.21 The company also reported some discrepancies between H2304 and the other trials. For example, more patients in H2304 had diabetes and hypertension at baseline, and the standard-dose tacrolimus group had better overall survival, graft survival and tBPAR-free probabilities than the standard-dose tacrolimus groups of the other trials. The company reported that, because H2304 was the only trial of everolimus with reduced-dose tacrolimus, it was not possible to conduct subgroup analyses excluding this study. H2304 was therefore assumed to be comparable with the rest of the evidence.

3.22 The results of the network meta-analysis were presented as a consistency model, in which direct and indirect evidence were assumed to be consistent for any 'closed loops' in the evidence network. An inconsistency model was also presented if data were available (that is, using direct evidence only). The company reported that all models were based on the NICE Decision Support Unit Document 4 (Inconsistency in networks of evidence based on randomised controlled trials, 2011) and that the parameters of the different models were estimated within a Bayesian framework using a Markov Chain Monte Carlo method as implemented in the WinBUGS/OpenBUGS software package. To assess heterogeneity in the treatment effects for a particular pair-wise comparison caused by the treatment effect modifiers, both fixed effects and random effects were modelled.

3.23 For overall survival at 12 and 24 months after transplantation, the company reported that everolimus with reduced-dose tacrolimus was expected to be comparable to all other treatments. It was ranked 5th and 4th of the interventions at 12 and 24 months, respectively. At 24 months, overall survival was 85.3% (95% credible interval 72.5 to 92.9), compared with:

- mycophenolate mofetil with ciclosporin (overall survival 88.9%, 95% credible interval 43.6 to 98.8)
- mycophenolate mofetil with standard-dose tacrolimus (overall survival 88.4%, 95% credible interval 83.8 to 92.0)
- standard-dose tacrolimus (overall survival 87.4%, 95% credible interval 84.1 to 90.2)
- azathioprine with standard-dose tacrolimus (overall survival 85.8%, 95% credible interval 76.7 to 91.8).

3.24 For graft survival at 12 and 24 months after transplantation, the company reported that everolimus with reduced-dose tacrolimus

was expected to be comparable to all other treatments. It was ranked 5th and 4th of the interventions at 12 and 24 months, respectively. At 24 months, graft survival was 79.7% (95% credible interval 67.3 to 88.4), compared with:

- mycophenolate mofetil with ciclosporin (graft survival 86.0%, 95% credible interval 49.4 to 97.5)
- mycophenolate mofetil with standard-dose tacrolimus (graft survival 85.3%, 95% credible interval 80.0 to 89.6)
- standard-dose tacrolimus (graft survival 82.6%, 95% credible interval 79.1 to 85.8)
- azathioprine with standard-dose tacrolimus (graft survival 80.8%, 95% credible interval 70.6 to 88.2).

3.25 For the outcome of being tBPAR-free at 3, 6 and 12 months after transplantation, everolimus with reduced-dose tacrolimus was ranked as the best therapeutic option of all the interventions. At 12 months, the absolute estimate was 89.5% (95% credible interval 82.3 to 94.4), compared with:

- mycophenolate mofetil with standard-dose tacrolimus (tBPAR-free 83.4%, 95% credible interval 75.8 to 88.9)
- mycophenolate mofetil with reduced-dose tacrolimus (tBPAR-free 80.6%, 95% credible interval 74.3 to 85.9)
- standard-dose tacrolimus (tBPAR-free 76.8%, 95% credible interval 72.0 to 81.2)
- azathioprine with standard-dose tacrolimus (tBPAR-free 75.6%, 95% credible interval 65.3 to 83.7)
- azathioprine with ciclosporin (tBPAR-free 72.3%, 95% credible interval 55.4 to 84.6).

3.26 For renal function at 12 months after transplantation (reported in the studies as eGFR or estimated creatinine clearance),

azathioprine with ciclosporin led to the lowest decline, followed by everolimus with reduced-dose tacrolimus. At 12 months, the absolute estimate for everolimus with reduced-dose tacrolimus was −23.1 (95% credible interval −27.4 to −18.7), compared with:

- azathioprine with ciclosporin (change in eGFR from baseline −14.5, 95% credible interval −24.2 to −4.9).
- mycophenolate mofetil with reduced-dose tacrolimus (change in eGFR from baseline −28.2, 95% credible interval −32.3 to −24.1)
- standard-dose tacrolimus (change in eGFR from baseline −31.6, 95% credible interval −32.3 to −30.9).

3.27 The company did scenario analyses, that involved removing specific trials from the network to assess the impact on the results. However, it did not present any results in its submission.

ERG comments on the network meta-analysis

3.28 The ERG found that many of the trials used in the network meta-analysis had substantially different tacrolimus target whole blood trough levels to those used in UK clinical practice. Some studies maintained tacrolimus blood trough levels above 5 ng/ml in the reduced tacrolimus dose arm, but other studies maintained blood trough levels below 5 ng/ml in the standard-dose tacrolimus arm. Therefore, no consistency was seen across studies with respect to target drug levels.

3.29 The ERG questioned the validity of the network meta-analysis results for the renal outcomes because it considered that the allocation of the different studies' treatment groups to the reduced-dose and standard-dose tacrolimus categories was inconsistent and misleading. Because the standard-dose tacrolimus connector across the network meta-analysis studies was so heterogeneous,

the ERG considered that the results of the network meta-analysis were not robust.

3.30 The ERG found a significant limitation in the network meta-analysis because the data included in the WinBUGS codes did not relate to the submission data and appeared to have been taken from either a different submission or a theoretical exercise. The ERG could not verify which data were used for the analysis of specific outcomes because of a lack of clarity and transparency in the company submission. The ERG stated that it was unclear which studies had been included in the analysis for the tBPAR outcome. The company highlighted that their submission provided network diagrams and data tables for the acute rejection outcome at various time points. However, the ERG commented that it was still unclear which studies had been included.

3.31 The ERG commented that the company's scenario analyses, that removed specific trials from the network to assess the impact on the results, lacked transparency and were not informative because no results were presented.

Cost effectiveness

3.32 The company did not identify any existing cost-effectiveness analyses of everolimus with reduced-dose tacrolimus that were relevant to the decision problem.

3.33 The company developed a patient-simulation model that evaluated the cost-effectiveness of everolimus with reduced-dose tacrolimus, with or without corticosteroids (that is, people were assumed to have corticosteroids initially and then tapered off completely from 6 months onwards), compared with azathioprine or mycophenolate mofetil with standard-dose tacrolimus for maintenance immunosuppressive therapy. The model considered a hypothetical

cohort of patients who had a liver transplant for any reason. The model included a core hepatic rejection model and a renal sub-model. The core model consisted of 6 health states: 'stable post-transplant, 'acute rejection', 'acute steroid resistant rejection', 'severe chronic rejection (leading to graft loss)', 'mild chronic rejection', and 'hepatic-graft related death'.

3.34 The company included the renal sub-model to demonstrate the 'renal sparing' effect of everolimus with reduced-dose tacrolimus. The sub-model included 5 health states defined by stages of chronic kidney disease (CKD) as measured by eGFR. These ranged from no CKD (eGFR 90+) to CKD stage 5 (eGFR < 15). It also included a renal-related death state. Patients could also leave the model from natural (background) mortality.

3.35 The model had a lifetime time horizon (80 years) and a cycle length of 3 months. The company stated that this reflected expert opinion that most acute rejection occurs 3 months after transplantation. A half-cycle correction was not applied. The model used discount rates of 3.5% for costs and QALYs and an NHS/Personal and Social Services perspective.

ERG comments on the model structure

3.36 The ERG considered that the company did not provide enough evidence to justify their approach of using a patient-level simulation. The company reasoned that the use of a patient-simulation model is appropriate when the patient flow is determined by the time since the last event or by the history of previous events. However, the ERG highlighted that only the severe chronic rejection state was affected by the time since the last event, and that the renal sub-model transition probabilities were not time dependent and also did not depend on history of previous events. Based on assessment of patient heterogeneity, and the patient

baseline characteristics simulated in the economic model, the ERG considered that a patient-simulation model was not needed, and that a cohort state-transition model would have been more appropriate.

3.37 The ERG considered that more emphasis should have been placed on the renal component of the economic model and also that more interaction between the 2 models should have been considered, perhaps within 1 broader model structure. The ERG stated that this was because immunosuppressive therapy after liver transplantation has an impact on renal functioning, and renal functioning has an impact on graft survival.

3.38 The ERG found that the reporting of the model's structure and assumptions lacked clarity and that few justifications were provided for the assumptions used. In particular, the ERG questioned the clinical plausibility of the mild chronic rejection state (an asymptomatic state that patients could only move to 1 year after transplantation). The ERG's clinical expert adviser did not see a valid or justifiable reason for patients only to progress to this state 1 year after transplantation, therefore the relevance of including this health state in the model was not clear.

3.39 The ERG's clinical adviser stated that 3-month cycles were too long to capture all the relevant events and that monthly cycles may have been more appropriate. The ERG also stated that because the cycle length of 3 months was relatively long, a half-cycle correction should have been applied. The ERG highlighted that although the impact of a half-cycle correction on the model outcomes would not be significant, no justification was given by the company as to why this was not applied.

3.40 The ERG considered that the time horizon of 80 years was unnecessarily high given that the average starting age of people in

the model was 54 years. The ERG highlighted that after 40 years (when the average age in the model was 94 years), 100% of patients would have died.

3.41 The ERG expressed concern about the number of simulations (10,000) and the lack of stability in the patient-simulation model. The ERG explored this in its exploratory analyses (see sections 3.66 and 3.67).

Model parameters

3.42 In the core hepatic rejection model, disease progression from the stable post-transplant state to the acute rejection state was determined by the immunosuppressive regimen (the treatment group). The probability of progression was derived from the probability of being free from treated biopsy proven acute rejection (tBPAR) at 3, 6 and 12 months for each treatment group, which was calculated from the network meta-analysis. The transition probabilities from the stable post-transplant state remained constant from the 5th cycle (starting at month 13) onwards. All other transition probabilities for the other health states were assumed to be constant and were therefore independent of the immunosuppressive regimen.

3.43 For the renal sub-model, transition probabilities for the first year were based upon the annual decrease in eGFR from baseline and were dependent on the treatment in year 1. The relative difference between treatments was calculated from the network meta-analysis and used to derive the absolute eGFR decrease for each treatment. After year 1, the company assumed that renal function followed a natural progressive decline meaning that transition probabilities were assumed to be constant. These transition probabilities were not time dependent or treatment-arm dependent and were based on underlying disease progression probabilities.

3.44 Health-related quality of life data were not collected as part of H2304, therefore the company did a systematic literature review to identify health state utility values. The company identified 7 studies, 5 of which were studies measuring EQ-5D in a UK population. All of the studies provided data for patients after transplantation, but none of the studies provided utility data specific to either acute or mild rejection, nor did they report disutility data specific to adverse events. Utility scores for the health states in the hepatic rejection model and for the renal sub-model were based on Ratcliffe et al., 2002 and Neri et al., 2012 respectively, both UK studies using EQ-5D. The study by Neri et al., 2012 assessed the relationship between health utility and renal function in people who had a kidney transplant.

3.45 Patients in the core hepatic model were assumed to have a stable health-related quality of life over time, with most states assuming an asymptomatic state with a utility value of 0.58. Patients in the more severe state of graft loss (severe chronic rejection) experienced a decrease in utility to 0.53. In the renal sub-model, patients' health-related quality of life decreased in line with their symptoms until renal transplantation, from 0.83 in the 'no CKD' health state, to 0.64 for CKD stage 1 to 2, 0.58 for CKD stage 3, 0.49 for CKD stage 4 and 0.28 for CKD stage 5. The company used the 'minimum method', to take into account potential double counting of utility losses in simultaneous health states (for example, hepatic rejection and renal dysfunction). The minimum method assumes that the lowest value is used as the estimate of joint state utility.

3.46 The company obtained estimates of resource use for the hepatic rejection model from the University Hospitals Birmingham Foundation Trust and these were validated by the company's clinical advisers. For the renal sub-model, the company reported

that NICE Guideline CG182 and Kerr et al 2012 were the 2 sources used to obtain resource use data.

3.47 The company estimated the occurrence of adverse events associated with the different treatment regimens based on the network meta-analysis and the summary of product characteristics for each drug product. These included hypertension, diabetes mellitus, infections, tremor and insomnia. The expected cost per treatment regimen was calculated by applying the cost of treating events with the probability of the events happening. The disutility estimates for treatment related adverse events were estimated from the published literature where possible.

ERG comments on the model parameters

3.48 The ERG questioned the validity and applicability of the economic analysis for the NHS context, because blood trough levels of reduced-dose tacrolimus in H2304 were above what would be considered as reduced levels in UK clinical practice (see section 3.28). It also highlighted that because the standard dose of tacrolimus across the network meta-analysis studies was so heterogeneous, the network meta-analysis results that informed the model were likely to lack robustness (see section 3.29).

3.49 The ERG identified structural errors in the formulae that allocated patients to different health states in the model, which it considered could have been avoided if a cohort state-transition model had been used. The ERG could not correct these errors because of the computational burden needed to run the patient-level simulation model, but it considered that the results were likely to be biased in favour of the everolimus treatment regimen.

3.50 The ERG commented that because of the random generation process used to estimate the baseline eGFR levels some patients

started the renal sub-model model with negative levels of eGFR, which is not clinically plausible. The ERG explained that this meant these patients were immediately allocated to the CKD stage 5 category, where patients with a negative eGFR level could stay for long periods of time (for example over 3 years) until they returned to CKD stages 1–2 following a transplant.

3.51 The ERG raised some concerns about the utility values used in the model. For example, it noted that the company had assumed that quality of life in the acute rejection and mild chronic rejection health states was the same as in the stable post-transplant state. However, clinical opinion sought by the ERG suggested that patients in the acute rejection and mild chronic rejection health states would require hospitalisation and that this would reduce their quality of life relative to the stable post-transplant state. The ERG also considered that the utility value of 0.83 for the 'no CKD' health state was more likely to represent CKD stage 1 and that the utility value of 0.64 for CKD stages1 to 2 was too low.

3.52 The ERG was generally satisfied with the sources used to obtain unit costs for the hepatic model. However, it highlighted that clinical practice varies across centres for patients having a liver transplant but that the company obtained resource use data from 1 centre only and it was therefore important to validate these against different sources. In general, clinical opinion sought by the ERG disagreed with some of the resource data reported in the submission. The ERG commented that GP visits were unlikely to occur as frequently as described because most of these patients would be managed in secondary care. The ERG also highlighted that the number of tests required in some health states may have been underestimated. For example, in the stable post-transplant state, people may also require a blood test to check immunosuppressive drug trough levels. The ERG also considered

that the cost associated with the mild chronic rejection health state (£640) seemed too high for an asymptomatic condition that does not require treatment and it was not clear why this was higher than the cost associated with the stable post-transplant health state (£73).

3.53 The ERG noted that the company used the most expensive brand price (£1.61 per mg) for tacrolimus (Prograf) in the economic model with no apparent justification. The ERG suggested that a weighted average price of £1.30 per mg (based on the market share information) for tacrolimus would have been more appropriate.

3.54 The ERG was generally satisfied with the estimation of adverse events and the quality of life data used to reflect these in the core hepatic rejection model. However, it did not agree that the costs and utility losses associated with everolimus-related adverse events should have been included for the 3 months in the first model cycle because everolimus therapy starts 1 month after surgery, and therefore the adverse events associated with the drug should have been considered only for 2 months.

Cost-effectiveness results

3.55 Everolimus with reduced-dose tacrolimus was more costly and resulted in more QALYs than the other treatment regimens. The deterministic base-case incremental cost-effectiveness ratio (ICER) estimated by the company was £110,797 per QALY gained compared against mycophenolate mofetil with standard-dose tacrolimus (incremental costs £38,004, incremental QALYs 0.343), and £187,842 per QALY gained compared against azathioprine with standard-dose tacrolimus (incremental costs £35,221, incremental QALYs 0.188). The ICER for the azathioprine treatment regimen compared with the mycophenolate mofetil regimen was £17,895 per QALY gained.

3.56 No deterministic sensitivity analyses were reported in the company's submission. The company ran a probabilistic sensitivity analysis using 1000 simulations for 1000 patients. The company reported that the results of the probabilistic sensitivity analysis were similar to the deterministic base-case results. The probabilistic ICERs for the everolimus treatment regimen were £105,526 and £184,714 compared with the mycophenolate mofetil and azathioprine treatment regimens, respectively.

3.57 The company's probabilistic sensitivity analysis indicated that for all simulations the ICERs for the everolimus treatment regimen compared with the mycophenolate mofetil regimen were higher than £30,000 per QALY gained. Similarly, the majority of simulations (>99%) were above £30,000 per QALY gained for the comparison of the everolimus and azathioprine treatment regimens. The analysis indicated that everolimus with reduced-dose tacrolimus was likely to be the most cost effective therapy only if the maximum acceptable ICER was over £200,000 per QALY gained.

3.58 The company undertook a number of scenario analyses. In scenario 1, the company removed the mild chronic rejection state. This increased the ICER for the everolimus treatment regimen compared with the mycophenolate mofetil regimen to £227,528 per QALY gained. The azathioprine treatment regimen was dominated by the mycophenolate mofetil regimen.

3.59 In scenario 2, the company removed the opportunity for re-transplant. This resulted in an ICER for everolimus in combination with reduced-dose tacrolimus of £121,972 per QALY gained compared with the mycophenolate mofetil treatment regimen and £117,285 per QALY gained compared with the azathioprine treatment regimen.

3.60 In scenario 3, the company removed the renal sub-model. This increased the ICER for everolimus in combination with reduced-dose tacrolimus to £312,279 per QALY gained compared with the mycophenolate mofetil treatment regimen and to £374,832 per QALY gained compared with the azathioprine treatment regimen. The company stated that the large impact on the results reflected the benefit that the everolimus treatment regimen provides for patients through a renal sparing effect.

3.61 In scenario 4, the company reduced baseline eGFR from 81 ml/min/1.73 m^2 to 60 ml/min/1.73 m^2. This resulted in an ICER for everolimus in combination with reduced-dose tacrolimus of £184,372 per QALY gained compared with the mycophenolate mofetil treatment regimen and £179,427 per QALY gained compared with the azathioprine treatment regimen. The company highlighted that the results of the scenario analyses demonstrated that the model was sensitive to changes in baseline eGFR.

ERG comments on the cost-effectiveness results

3.62 The ERG noted a logical error in the model when it analysed data provided by the company on the average number of cycles spent in the different health states of the model. It noted that there was a total of 320 cycles in the model and that in the hepatic rejection model patients spent an average of 41 cycles in the everolimus with reduced-dose tacrolimus group. In the remaining 279 cycles the patients were dead. However, in the renal model, patients only spent 31 cycles alive in the model, meaning that 10 cycles were 'missing' from the renal sub-model. The ERG suggested that this logical error reflected a problem in the model formulae and/or structure.

3.63 The ERG commented that it would be useful to understand what the key drivers of the economic model were, especially considering

the high ICERs presented, but that this was not possible because the company did not undertake deterministic sensitivity analysis.

3.64 The ERG highlighted that the results of the probabilistic sensitivity analysis were based on a small number of simulations (1000) and were unlikely to have generated reliable estimates.

3.65 The ERG queried the results of the company's scenario analyses (see 3.60 to 3.63). The ERG could not find a plausible reason why removing the mild chronic rejection state in the company's scenario 1 analysis increased the ICER for the everolimus treatment regimen compared with the mycophenolate mofetil regimen but decreased it compared with the azathioprine regimen. It also noted that there was no consistency for scenario 2 (removing the re-transplantation option from both models) and scenario 4 (decreasing the baseline eGFR level from 81 mL/ min/ 1.73 m^2 to 60 mL/min/1.73 m^2), with the ICERs increasing compared with the mycophenolate mofetil treatment regimen and decreasing compared with the azathioprine treatment regimen.

ERG exploratory analyses

3.66 The ERG ran 2 iterations of the company's base-case model using the same number of simulations and the same assumptions**,** to test the model's stability with regard to the ICER results. The ERG found considerable variation in the ICERs reported, especially for the everolimus treatment regimen compared with the mycophenolate mofetil regimen, with the ICERs ranging from £110,797 to £120,651 per QALY gained (nearly a 9% change). The ERG concluded that there was instability in the base-case ICERs.

3.67 The ERG tried to determine the cause of instability in the model results. Therefore it "fixed" the baseline characteristics of patients (by taking their mean values) in the simulation model instead of

allowing these values to vary in each simulation according to a distribution. The reason for this analysis was to understand if the variation in results was generated by the simulated patient characteristics or if it was attributable to other problems in the model. The results of this exercise generated ICERs for the everolimus treatment regimen compared with the azathioprine regimen that ranged from dominant to £797,558 per QALY gained. For the everolimus treatment regimen compared with the mycophenolate mofetil regimen, the ICERs ranged from £431,348 to £582,668 per QALY gained. The ERG therefore considered that the instability of the model could not be solved by fixing patient baseline characteristics as mean estimates.

3.68 The ERG lacked confidence in the ICERs presented by the company but it considered that it would not be helpful to run any additional analyses with different input values because of the instability of the results. The ERG stated that it could not therefore make any predictions regarding the true cost effectiveness of everolimus.

Additional evidence following consultation on the Appraisal Consultation Document

3.69 No comments from individual patients or professional groups were received during consultation. The company submitted additional evidence relating to the generalisability of the H2304 trial. It acknowledged that there is limited published evidence showing that tacrolimus trough levels of 5 ng/mL or below are achieved in clinical practice, following a liver transplant. The company presented a summary of target tacrolimus trough levels, and those achieved, from clinical trials identified in a 2012 systematic review. These data showed that in the 3 studies that included reduced-dose tacrolimus in combination with mycophenolate mofetil, none of the

studies achieved mean tacrolimus trough levels below 5 ng/ml. The company reported that no studies comparing reduced-dose tacrolimus with azathioprine were identified in the systematic review. In addition, the company submitted evidence showing that most patients in H2304 who had everolimus with reduced-dose tacrolimus had a mean tacrolimus-trough level within or below the target range of 3– 5 ng/ml by month 12.

3.70 The company submitted an updated model. This increased the number of simulations in the base-case and scenario analyses from 10,000 to 40,000, which the company reported increased the stability of the cost-effectiveness results. It included 11 input changes, including:

- amending the renal-efficacy data in the model. Instead of taking the estimated 12-month decrease in renal function from the arm of the network meta-analysis that looked at standard-dose tacrolimus, it was updated to take it from the arm that looked at mycophenolate mofetil with reduced-dose tacrolimus. The company acknowledged that, in clinical practice, the studies that informed the reduced-dose tacrolimus arm would be considered to use a standard dose because trough levels were consistently higher than 7.5 ng/ml
- using the average brand price for tacrolimus as calculated by ERG, rather than the Prograf brand price
- applying adverse event costs for everolimus for 2 months instead of 3 months in the first cycle, because treatment starts 30 days after a transplant
- shortening the time horizon of the model from 80 years to 40 years

- recalculating renal-progression rates using the correct rate–probability conversion equation to generate the correct risk of progression to subsequent CKD stages per 3-monthly cycle.
- correcting a number of errors relating to the calculation of transition probabilities
- correcting the estimate of baseline eGFR levels so that no patients had negative levels
- correcting an incorrect formula in the renal sub-model that led to 10 'missing' cycles identified by the ERG

3.71 The company presented an updated base-case analysis that included the 11 amendments to the economic model. This increased the base-case ICER for everolimus with reduced-dose tacrolimus when compared with using mycophenolate mofetil with standard-dose tacrolimus from £110,797 to £176,604 per QALY gained. It also decreased the base-case ICER for everolimus with reduced-dose tacrolimus compared with using azathioprine with standard-dose tacrolimus from £187,842 to £104,782 per QALY gained. The company reported that a second run of the updated model resulted in smaller variations in the ICERs than in the original base-case results and that this was because of the increase in simulations from 10,000 to 40,000. The ICERs were slightly higher for everolimus compared with the mycophenolate mofetil regimen than for everolimus compared with the azathioprine regimen. The company explained that this is because mycophenolate mofetil with standard-dose tacrolimus is associated with a slightly lower per-cycle rate of acute rejection (0.6%) when compared with everolimus (2.5%) and azathioprine (1.6%) at 13 months follow-up and beyond. The company also reported that the mycophenolate mofetil regimen had an adverse-event-related disutility score (−0.011) that was more favourable than azathioprine (−0.015), but less favourable than everolimus (−0.009).

3.72 The company's updated model included a number of new scenario analyses, including:

- reducing the baseline eGFR from 81 ml/min/1.73 m^2 to 60 ml/min/1.73 m^2, to be more reflective of the patient population in the UK. The ICER for everolimus with reduced-dose tacrolimus was £197,404 per QALY gained compared with the mycophenolate mofetil treatment regimen and £107,618 per QALY gained compared with the azathioprine regimen.
- removing the mild–chronic rejection state from the model. The ICER for everolimus with reduced-dose tacrolimus was £172,893 per QALY gained compared with the mycophenolate mofetil treatment regimen and £95,794 per QALY gained compared with the azathioprine regimen.
- amending the utility values for acute rejection, acute steroid-resistant rejection and mild–chronic rejection from 0.58 to 0.56 so that they were lower than the stable post-transplant health state. The ICER for everolimus with reduced-dose tacrolimus was £171,116 per QALY gained compared with the mycophenolate mofetil treatment regimen and £103,858 per QALY gained compared with the azathioprine regimen.

3.73 The company also repeated a scenario analysis that the ERG had done to test the stability of the original model. The company fixed the baseline characteristics, which resulted in an ICER of £320,637 per QALY gained for everolimus with reduced-dose tacrolimus compared against mycophenolate mofetil with standard-dose tacrolimus. The ICER per QALY gained was £161,462 for everolimus with reduced-dose tacrolimus, compared against azathioprine with standard-dose tacrolimus. The company highlighted that this demonstrated that its updated model was more stable and produced smaller variations in the ICERs when the model was rerun.

3.74 Using the original model, the company explored the impact of reclassifying tacrolimus dosing on the acute-rejection efficacy inputs in the model by re-running the network meta-analysis. Studies with treatment arms that included tacrolimus at trough levels less than 5 ng/ml were reclassified as 'reduced tacrolimus', while studies with treatment arms at trough levels of more than 5 ng/ml were reclassified as 'standard tacrolimus'. This reclassification of studies in the network meta-analysis changed the probability of acute rejection for the 3 treatment regimens and resulted in ICERs for everolimus with reduced-dose tacrolimus of £101,893 per QALY gained compared with the mycophenolate mofetil treatment regimen and £73,827 per QALY gained compared with the azathioprine regimen.

3.75 Full details of all the evidence are in the Committee papers.

4 Consideration of the evidence

The Appraisal Committee reviewed the data available on the clinical and cost effectiveness of everolimus, having considered evidence on the prevention of organ rejection after a liver transplant, and the value placed on the benefits of everolimus by people with the condition, those who represent them, and clinical experts. It also took into account the effective use of NHS resources.

4.1 The Committee discussed the impact of liver transplantation on patients. It heard from the patient experts that for the majority of people who have had a liver transplant their main concerns are about rejection of the transplanted liver and managing their long-term condition, and that all patients put great value on the prolonged life that a liver transplant can give them. A patient expert also stressed the importance of a choice of treatments being

available so that treatment could be tailored to best suit each individual. The Committee agreed that although a liver transplant can offer substantial benefits in terms of survival, managing their long-term condition and the concern about possible rejection can impact on a person's quality of life. It concluded that treatments that reduce the chance of organ rejection and minimise the adverse effects of long-term drug therapies are highly valued by patients.

4.2 The Committee discussed the treatments used to prevent rejection in people who have had a liver transplant. It heard from one of the clinical experts that local protocols consist of induction therapy with monoclonal or polyclonal antibodies followed by maintenance therapy with a calcineurin inhibitor in combination with an anti-proliferative agent (most commonly azathioprine, but increasingly mycophenolate mofetil), and corticosteroids for 3 months. The clinical expert stated that tacrolimus is the standard calcineurin inhibitor used in liver units as part of maintenance treatment, and that clinicians aim to reduce the initial dose over the first 12 months to a target blood trough level of 3–5 ng/ml. This is desirable because it is recognised that there is both a short-term and long-term association of liver transplantation with reduced renal function. Some of this reduction occurs at the time of the liver transplant, and some later, particularly in the first year after a transplant. Although the causes of reduced renal function are multifactorial, the nephrotoxic effect of tacrolimus is considered to be a significant contributory factor. The Committee agreed that an early reduction in the dose of tacrolimus after transplantation is therefore desirable to minimise the detrimental impact on renal function.

4.3 The Committee explored the potential benefits of everolimus in combination with a reduced dose of tacrolimus. The Committee heard from the clinical expert that in current clinical practice the desired reduction in tacrolimus levels may not be rapidly or

consistently achieved. The clinical expert referred to an audit of UK liver transplant units in 2013, which showed that the average tacrolimus blood trough level was 8 ng/ml at 3–6 months after transplantation, and 7 ng/ml at 12 months. The Committee also heard that although there was increasing clinical awareness of the desirability of reducing tacrolimus levels, the audit had demonstrated substantial heterogeneity in practice, with blood trough levels ranging from 5 ng/ml to 12 ng/ml. The Committee understood from the clinical expert that an advantage of treatment with everolimus was that it allowed earlier reduction in the tacrolimus dose than with azathioprine or mycophenolate mofetil, beginning 3 weeks after transplantation; helping to preserve renal function in the critical first year after the transplant. It might, therefore, be of particular benefit to patients with, or at risk of developing, renal dysfunction. However, the Committee also heard from another clinical expert that there was already a trend towards reducing tacrolimus blood trough levels and that there was uncertainty in the clinical community about the additional value of everolimus. It also heard that there are currently no nationally agreed protocols for the use of immunosuppression after transplant, or an agreed way of identifying those who might benefit most from everolimus. The Committee additionally heard from one of the clinical experts that everolimus has anti-tumour properties that could, theoretically, be of value in preventing the recurrence of hepatocellular carcinoma, if this was the reason for the transplant. The Committee concluded that everolimus with tacrolimus may allow earlier reduction in the dose of tacrolimus (to blood trough levels lower than those currently achieved in clinical practice) with the intention of better preserving renal function. It also concluded that there appeared to be a lack of consensus in the clinical community about the clinical advantages of everolimus.

Clinical effectiveness

4.4 The Committee considered the evidence presented by the company on the clinical effectiveness of everolimus in combination with reduced-dose tacrolimus. It understood that the main source of evidence was the H2304 randomised controlled trial and that everolimus was statistically non-inferior to standard-dose tacrolimus given as monotherapy for the primary composite outcome of treated biopsy proven acute rejection (tBPAR), graft loss or death at both 12 and 24 months. The Committee heard from the clinical expert that very early episodes of acute rejection are rarely problematic and are usually successfully treated, but preventing episodes of acute rejection later was important because they may progress to steroid-resistant rejection and rapid graft-loss. Any treatment that reduced episodes of acute rejection would be clinically beneficial. The Committee also heard from a clinical expert that mild chronic rejection may sometimes be associated with poor compliance with medication. Although initially asymptomatic, it can have a slowly progressive course to graft loss. The Committee also noted that everolimus with reduced-dose tacrolimus was associated with a smaller reduction in renal function compared with standard-dose tacrolimus at 12 and 24 months after transplantation as measured by change in estimated glomerular filtration rate (eGFR). The Committee concluded that the results of H2304 suggested that everolimus with reduced-dose tacrolimus was non-inferior to standard-dose tacrolimus alone for the composite hepatic outcome in the trial, and was an effective treatment for reducing the decline in renal function when compared with standard-dose tacrolimus.

4.5 The Committee considered the generalisability of H2304 to clinical practice in England. It noted that only a small number of patients were recruited to the trial from UK centres and that the direct

relevance of the trial to standard NHS practice in England was limited because the comparator, standard-dose tacrolimus as monotherapy, is not used. The Committee heard from the clinical expert that people in the trial also had better baseline renal function than is typically seen for patients in England (81 mL/min/1.73 m^2 in H2304 compared with approximately 60 mL/min/1.73 m^2 in clinical practice). In addition, the Committee noted the ERG's comments that a reduced dose of tacrolimus was considered standard care in the UK and that the blood trough levels for reduced-dose tacrolimus in the everolimus arm of H2304 (at around 5–6 ng/ml at 6 and 12 months) would be higher than expected in UK clinical practice. However, the Committee recalled the comments from the clinical expert that although the aim in clinical practice was to reduce the blood trough levels to below 5 ng/ml, an audit had shown that this did not seem to be consistently achieved (see section 4.2). The Committee appreciated that the audit provided only a single assessment of tacrolimus blood trough levels achieved in clinical practice in 2013, but that nevertheless it represented the best available evidence on current UK practice. It also noted that the company had stated in its response to the appraisal consultation document that most people in the everolimus with reduced-dose-tacrolimus arm of H2304 achieved tacrolimus blood trough levels below 5 ng/ml. Taking this into account, the Committee concluded that patients in the everolimus with reduced-dose tacrolimus arm of H2304 probably had lower tacrolimus blood trough levels than generally achieved in current clinical practice, although the magnitude of the difference remained uncertain. The Committee agreed that H2304 was well conducted but noted the high drop-out rates, the better renal function of participants than in clinical practice, and the limited long-term follow-up. It concluded that, although H2304 showed that everolimus with reduced-dose tacrolimus was an effective treatment for preventing organ rejection

and preserving renal function compared with standard-dose tacrolimus, there was uncertainty about how any benefit demonstrated in the trial would translate into clinical practice.

4.6 The Committee considered the network meta-analyses presented by the company to estimate the relative effectiveness of everolimus compared with the comparators specified in the scope. It was aware that the results of the network meta-analyses suggested that everolimus was expected to be comparable to other treatments for the outcomes of overall survival and graft survival at 12 and 24 months after transplantation, and that everolimus was ranked as the best treatment for reducing the incidence of tBPAR at 3, 6 and 12 months. The Committee noted the ERG's comments that there was inconsistency across studies with respect to tacrolimus blood trough levels. It also noted the ERG's concerns that the data used for the tBPAR outcome could not be verified because of a lack of clarity and transparency in the company submission (see section 3.30). It was also aware that in response to the appraisal consultation document, the company had carried out an exploratory analysis of the network meta-analysis, reclassifying some studies with respect to standard or reduced-dose tacrolimus, but had used this to explore hepatic outcomes only. The Committee concluded that there was considerable uncertainty in the results of the network meta-analyses because the dose of tacrolimus was so heterogeneous between the included studies and the company's approach lacked transparency because it was unclear which studies had been included for the analysis of specific outcomes.

4.7 The Committee considered the adverse events associated with everolimus in combination with reduced-dose tacrolimus from H2304. It noted that lipid changes occurred more frequently in the everolimus group than in the standard-dose tacrolimus group and that there was a higher incidence of new onset diabetes mellitus

and biliary leaks in the everolimus group at 24 months after transplantation. The Committee noted the clinical expert's comments that in practice the side effects of everolimus were significant but manageable and that with patient engagement they would be better tolerated. It also noted the comments that treatment with everolimus reduced the known risks associated with tacrolimus. The Committee concluded that the side effects of treatment with everolimus were manageable for patients and that treatment with everolimus could reduce the dose and therefore the risks associated with tacrolimus.

Cost effectiveness

4.8 The Committee considered the company's economic model and the critique and exploratory analyses performed by the ERG. It noted that the company presented a patient-simulation model consisting of a core hepatic rejection model and a renal sub-model. The Committee accepted the ERG's concerns about the structure and complexity of the model and agreed that a patient-simulation model may not have been the most appropriate design. It also questioned why the company had presented 2 discrete models instead of an integrated model, because hepatic and renal function are not entirely independent. It also accepted the ERG's comments that the 3-month cycles in the model were too long to capture all the relevant events and that monthly cycles may have been more appropriate, and that the time horizon of 80 years was unnecessarily long. The Committee noted that following consultation on the appraisal consultation document, the company had submitted an updated model that incorporated 11 amendments (see section 3.70) including a shorter time horizon of 40 years. However, the Committee considered that the other 10 amendments to the model addressed only issues related to model inputs and did not address specific concerns raised by the ERG about structural

issues. The Committee concluded that the choice of 2 separate models and the way in which they had been constructed was not necessarily the most appropriate approach to the economic evaluation and that these concerns applied to both the original and updated models supplied by the company.

4.9 The Committee discussed the original and updated model inputs, noting that the company had derived the efficacy estimates from the network meta-analysis. The Committee agreed that there was considerable uncertainty in the results of the network meta-analysis, related to the lack of clarity and transparency in the company's submission and inconsistency across studies with respect to tacrolimus blood trough levels (see section 4.6). The Committee noted that in its updated base-case analysis the company had amended the renal-efficacy inputs for the comparator arms, using the reduced-dose-tacrolimus arm of the network meta-analysis rather than the standard-dose arm as in the original submission (see section 3.70) because the company acknowledged that the studies informing the reduced-dose-tacrolimus arm would be considered standard-dose in clinical practice (that is, trough levels were consistently more than 7.5 ng/ml). However, the Committee was aware that these changes had not been validated by the ERG. In an exploratory analysis the company had constructed an updated network meta-analysis by reclassifying studies based on tacrolimus trough levels. The resulting data on hepatic outcomes had been used for the exploratory analysis, but were not incorporated into the updated base case. The Committee concluded that the changes made by the company in the new base case did not address the lack of clarity and transparency in the company's original submission, and did not address some fundamental concerns about the reliability of the model.

4.10 The Committee considered the utility values used in the core hepatic-rejection model and the renal sub-model. It noted the concerns of the ERG that the utility values used by the company for each of the health states in the core hepatic-rejection model did not differ (utility value of 0.58) apart from graft loss (0.53) and death (0.00). The Committee heard from the clinical expert that the utility for people in states other than the stable post-transplant state should be lower because of the deteriorating physical health, the need for more medical intervention and the associated anxiety. It noted that the company had submitted a scenario analysis following consultation on the appraisal consultation document that altered the assumption for the utility associated with the health states of acute rejection, acute steroid-resistant rejection and mild–chronic rejection (that is, it reduced the utility from 0.58 to 0.56 for each of these health states). However, the Committee did not accept that the evidence for the new values was robust and therefore their inclusion did not improve the validity or plausibility of the ICER. For the renal sub-model, the Committee questioned the plausibility of the reduction in utility from 0.83 to 0.64 for patients moving from having no kidney disease to having chronic kidney disease stages 1 or 2 because many people will be asymptomatic and not be aware of the reduced renal function at these earlier stages. It noted that the utility values for the stages of chronic kidney disease were derived from a study of patients who had a renal transplant. However, it accepted the comments from the company that the best available literature sources were used for each of the health states and that, because the lower of the renal and hepatic model utility values was used to estimate joint state utility, no patient in the model would ever be assigned a utility score of 0.83.

4.11 The Committee discussed other limitations of the model. It heard from the clinical expert that non-immunological conditions such as cardiovascular disease were major causes of death for patients in the years following a liver transplant. However, the Committee considered that it was unclear how the company had accounted for this in the model. The Committee also noted that the ERG had identified a logical error in the original model which meant that cycles were 'missing' from the renal sub-model and that this reflected a problem in the model structure or formulae. It noted that the company reported that it had corrected this error in its updated model but was also aware that the ERG had not validated this. The Committee concluded that, despite the company's amendments to the model following the consultation on the appraisal consultation document, there remained significant limitations and a substantial lack of clarity associated with the model as described in sections 4.8 to 4.10.

4.12 The Committee discussed the cost-effectiveness results for everolimus with reduced-dose tacrolimus compared with the azathioprine and mycophenolate mofetil treatment regimens (both in combination with standard-dose tacrolimus). It considered that the base-case ICERs from the original and updated models presented by the company were very high: £187,800 and £104,800 per QALY gained respectively when compared with the azathioprine treatment regimen, and £110,800 and £176,600 per QALY gained respectively when compared with the mycophenolate mofetil regimen. The Committee queried the reason for the reversal in the results relative to the comparator in the original, compared with the updated model. It heard from the company that this reflected technical errors in the original model that had been corrected in the updated analysis. The Committee concluded that the company's base-case ICERs from both the original and

updated models were substantially outside the range that would normally be considered a cost-effective use of NHS resources (£20,000–30,000 per QALY gained).

4.13 The Committee discussed the company's scenario analyses. It considered that the most relevant scenario analysis was the one in which the company changed the assumption of baseline eGFR from 81 ml/min/1.73 m^2 in the base case (the mean from the H2304 trial), to 60 ml/min/1.73 m^2, as this was more reflective of the baseline eGFRs seen in clinical practice (see section 4.4). The Committee noted that, when using the original model, this scenario resulted in ICERs for everolimus with reduced-dose tacrolimus of £179,400 per QALY gained compared with the azathioprine treatment regimen and £184,400 per QALY gained compared with the mycophenolate mofetil regimen. When this scenario was used in the updated model, the ICERs increased from the base-case estimate of £104,800 to £107,600 per QALY gained, when compared with the azathioprine regimen, and from £176,600 to £197,400 per QALY gained when compared with the mycophenolate mofetil regimen. It concluded that the ICERs for this scenario in both the original and updated models remained substantially outside the range that would normally be considered a cost-effective use of NHS resources.

4.14 The Committee discussed the way in which the mild–chronic-rejection health state had been implemented in the model. It was aware that both the original and the updated model assumed that if people entered the mild–chronic-rejection health state they never recovered and remained in that state until death. The Committee noted the concerns of the ERG about the clinical validity of how this health state was implemented in the model and that it had a substantial impact on the cost-effectiveness calculations in a scenario analysis where this health state was removed. The

Committee took into account the clinical expert's comment that mild–chronic rejection could have a slowly progressive course to graft loss and that it is associated with an increased number of interventions, visits to hospital and anxiety in patients about deterioration in their health. It concluded that the mild–chronic-rejection health state was clinically important, and would be associated with increased care costs so it was not necessary to exclude it from the model.

4.15 The Committee discussed the robustness of the ICERs presented by the company. It noted that the ERG lacked confidence in the original model results and had concerns regarding the stability of the model because it found considerable variation in the base-case ICERs when the model was re-run without any changes to the model inputs. The Committee acknowledged that the updated model results submitted by the company suggested greater stability (see section 3.73), although these had not been validated by the ERG. The Committee remained cautious about the robustness of the ICERs. However, it was unable to identify any factors that it believed would lower the company's estimates of cost-effectiveness. It concluded that the ICERs for everolimus with reduced-dose tacrolimus were unlikely to be lower than the company's estimates of £184,000 per QALY gained compared with the mycophenolate mofetil treatment regimen and £107,600 per QALY gained compared with the azathioprine treatment regimen. Therefore everolimus with reduced-dose tacrolimus did not represent a cost-effective use of NHS resources and could not be recommended for preventing organ rejection after liver transplantation.

4.16 The Committee considered whether everolimus with reduced-dose tacrolimus was innovative. It noted the company's comments about the need to reduce the complications of treatment with calcineurin

inhibitors, including nephrotoxicity, and the view of a clinical expert that there is currently no treatment other than everolimus that safely enables clinicians to rapidly reduce tacrolimus blood trough levels, thereby potentially preserving renal function. The Committee agreed that everolimus was innovative in its potential to preserve renal function but it could not identify any substantial health benefits that had not been captured in the QALY estimates. It concluded that everolimus had not been shown to be cost effective and could not be recommended for use in the NHS.

4.17 The Committee considered whether it should take into account the consequences of the Pharmaceutical Price Regulation Scheme (PPRS) 2014, and in particular the PPRS Payment Mechanism, when appraising everolimus. The Committee noted NICE's position statement in this regard, and accepted the conclusion 'that the 2014 PPRS Payment Mechanism should not, as a matter of course, be regarded as a relevant consideration in its assessment of the cost effectiveness of branded medicines'. The Committee heard nothing to suggest that there is any basis for taking a different view with regard to the relevance of the PPRS to this appraisal of everolimus. It therefore concluded that the PPRS Payment Mechanism was irrelevant for the consideration of the cost effectiveness of everolimus.

Summary of Appraisal Committee's key conclusions

TAXXX	**Appraisal title: Everolimus for organ rejection in liver transplantation**	**Section**
Key conclusion		
Everolimus is not recommended within its marketing authorisation for preventing organ rejection in people having a liver transplant.		1.1,

The Committee concluded that everolimus with reduced-dose tacrolimus may allow earlier reduction in the dose of tacrolimus (to blood trough levels lower than those currently achieved in clinical practice) with the intention of better preserving renal function. It also concluded that there appeared to be a lack of consensus in the clinical community about the clinical advantages of the benefit associated with everolimus. The Committee concluded that the company's base-case ICERs for everolimus with reduced-dose tacrolimus compared against any relevant comparator from both the original and updated models were substantially outside the range that would normally be considered a cost-effective use of NHS resources (£20,000–30,000 per QALY gained). In addition, the Committee was cautious about the robustness of the ICERs because of the way in which the model had been constructed and because of concerns about the model inputs including: • the considerable uncertainty in the efficacy estimates from the network meta-analysis • the utility estimates for some of the health states were not based on robust evidence	4.3 4.12 4.8 – 4.15
Current practice	

Clinical need of patients, including the availability of alternative treatments	The Committee agreed that although a liver transplant can offer substantial benefits in terms of survival, managing their long-term condition and the concern about possible rejection can impact on a person's quality of life. It concluded that treatments that reduce the chance of organ rejection and minimise the adverse effects of long-term drug therapies are highly valued by patients.	4.1
The technology		
Proposed benefits of the technology How innovative is the technology in its potential to make a significant and substantial impact on health-related benefits?	The Committee concluded that everolimus with reduced-dose tacrolimus may allow earlier reduction in the dose of tacrolimus (to blood trough levels lower than those currently achieved in clinical practice) with the intention of better preserving renal function. It also concluded that there appeared to be a lack of consensus in the clinical community about the magnitude of the benefit associated with everolimus. The Committee agreed that everolimus was innovative in its potential to preserve renal function but it could not identify any substantial health benefits that had not been captured in the QALY estimate in the modelling.	4.2, 4.16

What is the position of the treatment in the pathway of care for the condition?	Everolimus has a marketing authorisation in the UK for 'the prophylaxis of organ rejection in patients receiving a hepatic transplant. In liver transplantation, everolimus should be used in combination with tacrolimus and corticosteroids'. Everolimus is taken orally.	2.1
Adverse reactions	The Committee concluded that the side effects of treatment with everolimus were manageable for patients and that treatment with everolimus could reduce the dose and therefore the risks associated with tacrolimus.	4.7
Evidence for clinical effectiveness		
Availability, nature and quality of evidence	The Committee understood that the main source of evidence was the H2304 randomised controlled trial. It agreed that H2304 was well conducted but noted the high drop-out rates, the better renal function of participants than in clinical practice, and the limited long-term follow-up and it was uncertain how these factors would affect outcomes.	4.4, 4.5

Relevance to general clinical practice in the NHS	The Committee considered that the relevance of trial H2304 to standard NHS practice in England was limited because it recruited a small number of patients from the UK, the comparator (standard dose tacrolimus as monotherapy) is not used in England, and people in the trial had better baseline renal function than is typically seen for patients in England.	4.5
Uncertainties generated by the evidence	There was uncertainty about how any benefit demonstrated in trial H2304 would translate into clinical practice. The Committee concluded that there was considerable uncertainty in the results of the network meta-analyses because the dose of tacrolimus was so heterogeneous between the included studies and the company's approach lacked transparency because it was unclear which studies had been included for the analysis of specific outcomes.	4.5, 4.6
Are there any clinically relevant subgroups for which there is evidence of differential effectiveness?	None were identified by the Committee.	

Estimate of the size of the clinical effectiveness including strength of supporting evidence	The Committee concluded that everolimus with reduced-dose tacrolimus was non-inferior to standard dose tacrolimus therapy for the composite hepatic outcome in the trial, and was an effective treatment for reducing the decline in renal function when compared with standard dose tacrolimus.	4.4
Evidence for cost effectiveness		
Availability and nature of evidence	The Committee concluded that the choice of 2 separate models and the way in which they had been constructed was not necessarily the most appropriate approach to the economic evaluation. The Committee agreed that the considerable uncertainty in the results of the network meta-analysis, related to the lack of clarity and transparency in the company's submission and inconsistency across studies with respect to tacrolimus trough levels, significantly undermined the reliability of the model. The Committee concluded that the changes made by the company in the new base case did not address the lack of clarity and transparency in the company's original submission, and did not address some fundamental concerns about the reliability of the model.	4.8, 4.9

Uncertainties around and plausibility of assumptions and inputs in the economic model	The Committee was cautious about the robustness of the ICERs because of uncertainty around: • efficacy estimates from the network meta-analysis • the utility estimates for some of the health states were not based on robust evidence	4.8 – 4.15
Incorporation of health-related quality-of-life benefits and utility values Have any potential significant and substantial health-related benefits been identified that were not included in the economic model, and how have they been considered?	The Committee acknowledged the comments from the company that the best available literature sources were used for each of the health states but it did not accept that the evidence for some of the values used was robust. The Committee agreed that everolimus was innovative in its potential to preserve renal function but it could not identify any substantial health benefits that had not been captured in the QALY estimates in the modelling.	4.10, 4.16
Are there specific groups of people for whom the technology is particularly cost effective?	None were identified by the Committee.	

What are the key drivers of cost effectiveness?	No key drivers were identified by the Committee. The ERG commented that it was not possible to identify the key drivers because the company did not undertake deterministic sensitivity analysis.	3.64
Most likely cost-effectiveness estimate (given as an ICER)	The Committee concluded that the ICERs for everolimus with reduced-dose tacrolimus were unlikely to be lower than the company's estimates of £184,000 per QALY gained compared with the mycophenolate mofetil treatment regimen and £107,600 per QALY gained compared with the azathioprine treatment regimen.	4.15
Additional factors taken into account		
Patient access schemes (PPRS)	The Committee concluded that the PPRS Payment Mechanism was irrelevant for the consideration of the cost effectiveness of everolimus.	4.17
End-of-life considerations	Not applicable.	
Equalities considerations and social value judgements	None identified by the Committee.	

5 Implementation

5.1 The Welsh Assembly Minister for Health and Social Services has issued directions to the NHS in Wales on implementing NICE technology appraisal guidance. When a NICE technology appraisal recommends the use of a drug or treatment, or other technology, the NHS in Wales must usually provide funding and resources for it within 3 months of the guidance published.

5.2 NICE has developed tools [link to www.nice.org.uk/guidance/TAXXX] to help organisations put this guidance into practice (listed below). [NICE to amend list as needed at time of publication]

- Slides highlighting key messages for local discussion.
- Costing template and report to estimate the national and local savings and costs associated with implementation.
- Implementation advice on how to put the guidance into practice and national initiatives that support this locally.
- A costing statement explaining the resource impact of this guidance.
- Audit support for monitoring local practice

6 Related NICE guidance

Details are correct at the time of consultation and will be removed when the final guidance is published. Further information is available on the NICE website.

- Organ donation for transplantation: improving donor identification and consent rates for deceased organ donation. NICE clinical guideline 135 (2011).

- Alcohol-use disorders: Diagnosis and clinical management of alcohol-related physical complications. NICE clinical guideline 100 (2010).
- Living-donor liver transplantation. NICE interventional procedure guidance 194 (2006).

7 Review of guidance

7.1 The guidance on this technology will be considered for review 3 years after publication of the guidance. NICE welcomes comment on this proposed date. The Guidance Executive will decide whether the technology should be reviewed based on information gathered by NICE, and in consultation with consultees and commentators.

Dr Jane Adam
Chair, Appraisal Committee
May 2015

8 Appraisal Committee members, guideline representatives and NICE project team

Appraisal Committee members

The Appraisal Committees are standing advisory committees of NICE. Members are appointed for a 3-year term. A list of the Committee members who took part in the discussions for this appraisal appears below. There are 4 Appraisal Committees, each with a chair and vice chair. Each Appraisal Committee meets once a month, except in December when there are no meetings. Each Committee considers its own list of technologies, and ongoing topics are not moved between Committees.

Committee members are asked to declare any interests in the technology to be appraised. If it is considered there is a conflict of interest, the member is excluded from participating further in that appraisal.

The minutes of each Appraisal Committee meeting, which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Dr Jane Adam (Chair)
Consultant Radiologist, Department of Diagnostic Radiology, St George's Hospital, London

Professor Iain Squire (Vice-Chair)
Consultant Physician, University Hospitals of Leicester

Dr Graham Ash
Consultant in General Adult Psychiatry, Lancashire Care NHS Foundation Trust

Dr Jeremy Braybrooke
Consultant Medical Oncologist, University Hospitals Bristol NHS Foundation Trust

Dr Gerardine Bryant

General Practitioner, Swadlincote, Derbyshire

Professor Aileen Clarke

Professor of Public Health & Health Services Research, University of Warwick

Dr Andrew England

Senior Lecturer, Directorate of Radiography, University of Salford

Mr Adrian Griffin

Vice President, HTA & International Policy, Johnson & Johnson

Dr Ian Lewin

Honorary Consultant Physician and Endocrinologist, North Devon District Hospital

Ms Pamela Rees

Lay Member

Dr Paul Robinson

Medical Director, Merck Sharp & Dohme

Ms Ellen Rule

Director of Transformation and Service Redesign, Gloucestershire Clinical Commissioning Group

Dr Brian Shine

Consultant Chemical Pathologist, John Radcliffe Hospital, Oxford

Dr Peter Sims

General Practitioner, Devon

Dr Eldon Spackman

Research Fellow, Centre for Health Economics, University of York

Mr David Thomson

Lay member

Dr John Watkins

Clinical Senior Lecturer, Cardiff University; Consultant in Public Health Medicine, National Public Health Service Wales

Professor Olivia Wu

Professor of Health Technology Assessment, University of Glasgow

NICE project team

Each technology appraisal is assigned to a team consisting of 1 or more health technology analysts (who act as technical leads for the appraisal), a technical adviser and a project manager.

Christian Griffiths

Technical Lead

Zoe Charles

Technical Adviser

Bijal Joshi

Project Manager

9 Sources of evidence considered by the Committee

A. The Evidence Review Group (ERG) report for this appraisal was prepared by Peninsula Technology Assessment Group (PenTAG):

- Bacelar M, Nakum M, Durand A, et al., Everolimus (Certican®) for preventing organ rejection in liver transplantation: A critique of the submission from Novartis, November 2014

B. The following organisations accepted the invitation to participate in this appraisal as consultees and commentators. They were invited to comment on the draft scope, the ERG report and the appraisal consultation document (ACD). Organisations listed in I were also invited to make written submissions. Organisations listed in II and III had the opportunity to make written submissions. Organisations listed in I, II and III also have the opportunity to appeal against the final appraisal determination.

I. Company:

- Novartis Pharmaceuticals (everolimus)

II. Professional/expert and patient/carer groups:

- British Liver Trust
- British Society of Gastroenterology
- ESPRIT
- Liver4Life
- Royal College of Nursing
- Royal College of Pathologists
- Royal College of Physicians
- United Kingdom Clinical Pharmacy Association

II. Other consultees:

- Department of Health
- NHS England
- Welsh Government

IV. Commentator organisations (did not provide written evidence and without the right of appeal):

- Astellas Pharma (immediate-release tacrolimus, prolonged-release tacrolimus)

- Department of Health, Social Services and Public Safety for Northern Ireland
- Foundation for Liver Research
- Healthcare Improvement Scotland
- National Institute for Health Research Health Technology Assessment Programme
- Peninsula Technology Assessment Group (PenTAG)
- Roche Products (mycophenolate mofetil)
- Sandoz (azathioprine, mycophenolate mofetil, immediate-release tacrolimus)

C. The following individuals were selected from clinical expert and patient expert nominations from the consultees and commentators. They gave their expert personal view on everolimus by attending the initial Committee discussion and providing a written statement to the Committee. They were also invited to comment on the ACD.

- Professor Derek Manas, Consultant Hepatobiliary and Transplant Surgeon, nominated by organisation representing Novartis Pharmaceuticals – clinical expert
- Mr Andrew Langford, Chief Executive of British Liver Trust, nominated by organisation representing British Liver Trust – patient expert
- Mr Richard Hall, Co-Founder of Liver4Life, nominated by organisation representing Liver4Life – patient expert

D. Representatives from the following company attended Committee meetings. They contributed only when asked by the Committee chair to clarify specific issues and comment on factual accuracy.

- Novartis