Company comment	Section	PenTAG Response
Only randomised controlled trials (RCTs) or systematic reviews of RCTs were included as part of the study selection process to assess clinical effectiveness. Sandoz Ltd appreciate that this is a standard and well recognised approach of undertaking systematic reviews in the context of a health technology assessment. It is important to note, however, that as part of the European regulatory submission process, companies producing a generic medicine only need to provide information on the quality of that medicine and demonstrate that the generic formulation produces the same rate and extent of absorption of the active substance in the body as the reference product. As such, RCT data for generic medicines are not typically available as part of the market entry process. By only including RCTs within the analyses, the evidence appraised for immediate-release TAC products is, thus, biased towards the reference product. The majority of the included trials that assess the clinical effectiveness of immediate-release TAC were completed prior to 2010 and the analyses ignore the more recent TAC exposure reports from the UK that are predominantly in abstract format. As described within the original submission by Sandoz Ltd, generic TAC products have a substantial market share in the UK, with recent data indicating the proportions stand at: [redacted] Hence, given the proportion of reference versus generic TAC formulations used in the UK, Sandoz Ltd anticipated the approach taken by the assessment group would have permitted analyses of studies more representative of the recent clinical use and exposure of immediate-release TAC products.	4.1.3.1, p 81	The NICE methods guidance is clear about the types and quality of evidence that should be used to inform its guidance to the NHS
Due to the structure of the cost-effectiveness model adopted within the Assessment Report – in which the drug acquisition costs for immediate-release TAC products were grouped together to produce an average cost based on market share – the analyses conducted fail to interrogate the potential differences in cost-effectiveness between the different immediate-release TAC products. As described within the original submission by Sandoz Ltd, [redacted]	7.3.6.3, p456, table 194	We believe that using eMit database average acquisition costs is closest to the NHS reference case for immediate-release tacrolimus. Furthermore, the NICE methods guidance requires us to use the national list price of drugs. Other discounted prices can only be considered if they are formally agreed with the Department of Health (e.g. Patient Access Schemes). Also, where there are different products (with different prices) in the same class, it is accepted practice to take a

[redacted] Sandoz Ltd would, therefore, like to note that the cost-effectiveness of immediate-release TAC products in the UK can only be expected to improve, when compared with the incremental cost-effectiveness ratio calculations of the Assessment Report, should there be a greater shift towards the use of the lower-priced immediate-release TAC products in the UK.		weighted average. If a formulation of immediate-release tacrolimus is consistently available at a lower cost than that assumed then the cost-effectiveness of immediate-release tacrolimus will further improve.
The assessment assumes all drugs are dosed identically in all trials, yet this may not be the case. Immunosuppression can be increased, and consequently rejection reduced, by increasing the therapeutic drug level. There does not appear to have been consideration of this in the assessment.	General	For the clinical effectiveness section, which reviewed a large number of interventions, outcomes and time points, it was impractical to also investigate variations of drug dose. That said, where high and low dose interventions were reported, only the lower dose (closer to clinical practice) was included. Drug dose was a consideration in the model in terms of costs. Where possible drug dosage was estimated according to time since transplantation and according to concomitant treatments, based on more recent RCTs. No comparisons were made which differed solely in terms of dosage (e.g., we do not compare standard dose ciclosporin to low dose ciclosporin with the same concomitant treatments).
The report considers death and graft survival as distinct entities, censoring death from its graft survival data. The immunosuppressive regimen may contribute to the likelihood of death, and therefore it is inappropriate to analyse the drugs in terms of graft survival data censored for death. NHS Blood and Transplant use the term Transplant Survival to denote the survival of a patient with a functioning graft and it would be helpful to see an analysis using this as an endpoint; it is the endpoint in kidney transplantation that both patient and clinician aim to achieve.	General	Within the model we predict death-censored graft survival, death with functioning graft and death following graft loss. These three processes (plus retransplantation) are sufficient to also derive transplant survival (time to graft loss of any cause including death) and overall survival. To obtain regimen-specific death-censored graft survival we considered the rate of graft loss (excluding DWFG) within 12 months, as well as surrogate predictors for extrapolation. To obtain regimen-specific overall survival we considered the rate of death within 12 months (either with functioning graft or following graft loss) and adjusted the rate of death with functioning graft within the first 12 months to match this. If deaths in RCTs were reliably reported according to whether they were subsequent to graft loss then we could have used an alternative method, but it was frequently not possible to reliably infer this data.

The National Library of Medicine Medical Subject Heading (MeSH term) Is “kidney transplantation”, not kidney transplant, renal transplant, or any of the other terms listed in the executive summary or body of the document. Could PenTAG confirm that they did use the MeSH term “kidney transplantation” as one of their search terms for this assessment.	p32	Yes, as detailed in Appendix 1, ‘kidney transplantation’ was one of the search terms
“Only economic evaluations from UK, USA, Canada, Australia, and western Europe were included as these settings may include data generalizable to the UK. I would question how generalizable data for cost effectiveness are for some of these countries. For example, until the advent of the Affordable Care Act in the US the patient did not pay for his/her immunosuppression for the first 12 months, but had to thereafter. Consequently graft survival in the US Was poorer beyond the first year, compared to other healthcare systems like the UK and Australia.	p35	We agree. We have included these countries for completeness, but have discussed the issues around generalisability such as in the example raised by the commentator, which incidentally is relevant for some of the evidence submitted by the industry. We thought it was better to present this evidence with the corresponding caveats rather than entirely exclude these studies, especially in light of the methodological insight offered by some of these studies.
PenTAG have developed their own economic model, and estimated graft survival independently from UK Transplant Registry data presumably supplied by NHS Blood and Transplant (NHSBT). PenTAG do not appear to have consulted NHSBT with the graft survival aspects of the model – NHSBT have a transplant statistical unit that has an international reputation for its expertise and data collection, and has established models for predicting graft life. They are also familiar with the limitations of their own data.	p36	The registry data was indeed supplied by NHSBT, but this data is available to all on request and did not involve any further consultation or discussion. PenTAG developed a model for graft survival which allowed the impact of the use of different regimens to be estimated in two parts: 1. The graft survival to 12 months is adjusted using a regimen-specific odds ratio from network meta-analyses of RCT data 2. The graft survival after 12 months is adjusted using a regimen-specific hazard ratio from a surrogate relationship based on acute rejection, graft function and NODAT within the first 12 months. The economic model is sufficiently flexible that alternative graft survival models can be plugged in if it is believed that better models exist.
“… In all except one of these trials this was primarily due to the fact that patients, donors or organ characteristics were not representative of the current NHS in England (e.g. >90% deceased donors, or ‘suboptimal transplants’, or ‘high risk of rejection population).” I am not sure what is being said here. The donor demographics	4.2.2.6, p92	This paragraph is referring to all included studies, but highlights that only 11 of them were applicable to the current NHS. No studies were excluded from the clinical effectiveness systematic review according to quality assessment or applicability to the NHS. They were included or excluded on the basis of the criteria specified in the protocol.

in the UK comprise 64% deceased donors of which 37% fulfil the UNOS expanded criteria definition and 41% are donors following circulatory death. None of the trials has studied the same population of donor kidneys that are currently being transplanted in the UK *and* included recipients at high---risk of rejection. Nevertheless only 11 studies were considered admissible, while most were not. I would like to see a better justification as to why each of the trials in Appendix 4 were ruled adequate or inadequate in terms of applicability to the NHS.		
Cadaveric transplant. Current terminology refers to transplants from someone who has died as Deceased donor transplants, not “cadaveric transplants” **Calcineurin inhibitor** Ciclosporin is variously referred to throughout the document as ciclosporin (the UK accepted spelling), cyclosporine, and cylsoporine (page 374). It should be standardised. **Regimen** An immunosuppressive protocol involving a number of drugs is usually referred to as a regimen. In this document the words regimen, regime, and regiment are used interchangeably. **1---Haplotype identical** I am not clear why this term is included. It is poorly defined in this list of terms, and is not used anywhere else in the document. **List of abbreviations.** KM: kaplan---meier Should read “Kaplan---Meier”	p25	Thank you for your comments. These issues will be corrected prior to publication in the HTA monograph
Three main service provision steps for the management of a transplant. May I suggest that there are 4, and that the second is preservation of the kidney (see TA165).	p28	Thank you for this comment we will make this addition to the background prior to publication in the HTA monograph.
Ciclosporin and azathioprine are not mentioned here but should	p28	We will indicate these in this section prior to publication of the report.

be since they are frequently used as comparators in the studies described in the assessment report.		
"mycophenolate sodium is an enteric---coated formulation of mycophenolate mofetil (MPS)" Incorrect. Mycophenolate sodium is the sodium salt of mycophenolic acid; mycophenolate mofetil is the mofetil salt of mycophenolate.	p29	Thank you. This will be corrected prior to publication of the report.
"**sirolimus** (Rapamune®[Pfizer]) which is a non---calcineurin inhibiting immunosuppressant and acts as an antiproliferative; **everolimus** (Certican ® [Novartis Pharmaceuticals]) which is a proliferation signal inhibitor and is an analogue of sirolimus." Sirolimus and everolimus have identical actions on proliferation through exactly the same pathway (inhibition of TORC1). I am not clear why there actions are described differently here.	p29	Thank you for your comment. We agree this is misleading and the report will be amended.
Outcome measures - "Graft function – (estimated) glomerular filtration rate (eGFR), which is an estimate of actual glomerula [sic] filtration rate, using a formula involving age, weight, gender, and serum creatinine." The data you analysed include other measures of graft function, such as creatinine clearance. GFR stands for glomerul**ar** filtration rate. There are a number of equations that can be used to estimate it. The most common is the MDRD equation which does not involve weight (Levey et al. Ann Intern Med 1999;130(6):461.) Weight is used by the Cockcroft Gault formula which gives an estimate of creatinine clearance, not GFR, although the two are similar (Cockcroft & Gault. Nephron 1976;16(1):31)	p31	Thank you for your comment. The definition of graft function will be clarified prior to publication of the report.
The rate of mortality following graft loss. This is a subject of much current interest in the transplant community. Did PenTAG use the UK Transplant Registry data to supply the NHS numbers of the recipients to The Renal Registry, or indeed to the Office of National Statistics, in order to ensure an accurate picture of the rate of death before and after graft loss? Or did they assume that the Renal Registry captured all the transplant	p38	We did not find robust published statistics for the rate of mortality following graft loss. We note from a conference abstract by Webb et al. [Webb 2012] that although there is an elevated mortality risk following graft loss for the first three years compared to control patients awaiting transplantation, this effect was not significant after three years.

patient data.		Due to the Markov memoryless property of the PenTAG model it was not possible to assume a temporary increase in mortality following graft loss. We therefore assumed that mortality following graft loss would be same for age-matched prevalent dialysis patients (taken from UK Renal Registry annual report). [Webb 2012] Webb L, Casula A, Tomson C, Ben-Schlomo Y. FP715 Survival after renal transplant failure: a UK Renal Registry analysis. Nephrol Dial Transplant 2012; 27(suppl 2):ii312
"Infection prophylaxis costs were estimated based on the kidney transplant protocol of a UK hospital". The UK hospital quoted was the Royal Devon and Exeter Hospital, which is not a transplant unit but rather it follows up patients transplanted elsewhere. Prophylaxis associated with particular induction regimens may be missed. In addition the prophylaxis and monitoring of CMV is very variable and may differ where rATG is used routinely.	p40	Infection prophylaxis costs are small in comparison to other costs incurred, e.g., maintenance immunosuppression and monitoring. It is our impression that infection prophylaxis is the same for all regimens, except where rATG is used. In the economic model it was assumed that there would be greater use of CMV prophylaxis when rATG is included, adding approximately £900 to total discounted costs. Based on the deterministic analyses, the total discounted cost of rATG regimens would need to reduce by £8,000 to £9,600 for rATG to be cost-effective versus BAS at £20,000 to £30,000 per QALY, suggesting that any changes to infection prophylaxis costs will not lead to rATG becoming cost-effective in this range.
"Since the previous NICE appraisal, the main development in economic evaluation modelling of immunosuppressive regimens is the use of renal function as a surrogate outcome in addition to acute rejection for extrapolating trial efficacy outcomes to long term graft and patient survival" The use of renal function as a surrogate of graft survival was suggested during the previous appraisal. The main economic difference relevant to this appraisal is the widespread availability of generic formulations of most of the immunosuppressants and consequent reduction in their cost.	p44	This is a comment specifically about research methodology and practice, rather than the new economic or market conditions. Although the previous appraisal mentioned the possible use of renal function as a surrogate, the evidence required to model its role in future graft survival and its current costs and health related quality of life implications only became available after the appraisal. Nonetheless, we do agree with the commentator that immunosuppressive drug therapy prices are a key driver of cost-effectiveness.
Limitations Key Strengths. Donor kidney quality has not been looked at. In the UK now the	1.6.6 p48	Thank you for your comment. We agree this is a limitation.

average age of a DBD donor is 52, and a DCD donor is 54. Donor age is a strong predictor of graft quality in all studies where donor factors affecting long term graft survival have been studied, including those by Port et al, Rao et al, and the UKKDRI published in conjunction with NHSBT.		
Limitations: Anaemia The assessment assumed that anaemia would not vary between regimens. Anaemia, and the requirement for exogenous erythropoietin, is much more common in patients on sirolimus (and probably everolimus), and in patients with failing grafts. It is uncommon in patients who are not on mTOR inhibitors and who have good graft function.	1.6.2.1 p49	This is an acknowledged limitation of the economic model. With regards mTOR-I use affecting anaemia, this would only worsen the cost-effectiveness of these agents, which are currently not predicted to be cost-effective at £20,000 to £30,000 per QALY in any case. With regards graft function, this suggests that we may have overcosted for anaemia in the short-term and undercosted in the long-term (since graft function generally declines gradually). We consider that this is unlikely to have biased the results since we identified no high-quality evidence for differing long-term eGFR.
"However, it should be noted that the total number of transplants for the year 2013/14 given in Figure 2 as 2,929 appears to be 14 transplants more than the total number of tranplants[sic] performed around the UK given (2,464 for England, 91 for NI, 112 for Wales and 242 for Scotland) from the same resource.11" This criticism of the NHSBT report is unnecessary. The explanation is simple. PenTAG have arrived at their figure by adding up activity in the NHS hospitals in figure 5.2 of the NHSBT 2013/14 annual kidney report (http://www.odt.nhs.uk/pdf/organ_specific_report_kidney_2014.pdf), and assumed this to be the total UK adult kidney transplant activity. In addition to this 14 live donor kidney transplants were carried out in private hospitals, 5 at the London Clinic, 5 at the London Cromwell, and 4 at London Bridge. The comment directed at NHSBT should be removed. I would have expected that the Assessment Group would have made every effort to ensure that data queries were resolved rather than accept data that they presume to be incorrect, or to infer criticism of another organisation without first checking with them.	2. Background. Page 56.	Thank you for this clarification, we will amend the report accordingly prior to publication. Any criticism inferred was not intentional.

The legend on the figure should read DCD: “donation after *circulatory* death”, not “donation after *cardiac* death”.	p57	Thank you for your comments. The report will be amended prior to publication.
The figure title should read “Number of donors, transplants and people on the active transplant list from **1 April** 2004 to 31 March 2014”.	p58 Fig 3	See above
“This, in effect, means that people are back to where they started with their treatment, …” This is not the case. Patients with failed transplants are commonly sensitized to the mismatched HLA on the failed donor kidney, such that they cannot have a future transplant bearing the same HLA. Moreover this sensitization is often to a wider spectrum of HLA. The end result is a much longer wait on dialysis for a second transplant than for a first, and when they get that transplant they may have a poorer outcome on account of the sensitization.	p59 final paragraph	See above
“The cross---matching for donor---recipient compatibility will include an assessment on HLA matching” The cross---match is a test to detect antibodies in the recipient serum directed to donor HLA as a consequence of which hyperacute rejection would occur were the transplant to proceed. HLA matching aims to find a donor kidney with a minimal number of HLA that are not present on the recipient. It is not cross---matching. The sentence is meaningless as written.	p64	Thank you for your comments. The report will be amended prior to publication.
“The immunosuppression regime[sic] can then be tapered and withdrawn whilst the patient returns to dialysis and waits for a new kidney to become available.” In fact the immunosuppression may be continued after the patient returns to dialysis. This is done in patients who have not already formed antibodies to the donor HLA and aims to prevent the formation of said antibodies, and thus give the patient access to a wider pool of potential donors (and hence a shorter wait) for their next kidney graft. It is generally practised if there is a live donor being worked up, when stopping the immunosuppression with resultant HLA antibody formation may	p65	Thank you for your comments. This will be clarified prior to publication.

make the retransplant impossible. This highlights the importance of maximizing the life of the first transplant, particularly in young patients if they are to have a long life which will usually be achieved with a number of kidney transplants.		
"Clinician estimations of average frequency of outpatient visits have been reported as 34.3, 6.3 and 4.7 visits respectively for the first, second and third years posttransplant, with UK database figures suggesting 39.7,11.0 and 9.2 visits respectively for the first, second and third years posttransplant." This gives the false impression that the data are derived from a large database when in fact the "UK database" is the database at the Cardiff transplant unit.	p67	Thank you for your comments. This will be clarified prior to publication
"**Mycophenolate sodium.** MMF is also available as an enteric---coated formulation mycophenolate sodium (MPS)" MMF is a different drug to MPS, as mentioned previously. They are different salts of mycophenolic acid.	p70	Thank you for your comments. This will be amended prior to publication
"Donor type" ECD is variously used to mean extended or expanded criteria donors. This category is a definition of the United Network for Organ Sharing (UNOS) in the USA and follows work by Port et al (Transplantation 2002;74:1281) showing that donors 60 years or older, or those between 50 and 59 who have two of death from a stroke, creatinine over 1.5mg/dl, and a history of hypertension, have a relative risk of graft loss at least 1.7 fold that of an ideal donor (aged 10 to 39, no hypertension, normal creatinine, not dying from a stroke). It does not include donors who may pass on malignancy or infection (which may kill the recipient but are not otherwise associated with poor graft function). The term marginal is often used to cover these different donor--- associated risks, but it is probably best to adopt the phrase "less than ideal". Age is the most important factor influencing graft function and survival, but there are other factors, the most important of which is probably the cold ischaemic time. Both	p71	Thank you for your comments. This will be clarified prior to publication

DBD and DCD kidneys can be ECD kidneys – ECD are a subset of deceased kidney donors.		
"Immunosuppression" in the table title is spelt incorrectly. "'Symphony study' regimen using triple therapy irrespective of immunological risk or delayed graft function risk with: tacrolimus; mycophenolate mofetil or mycopehnolate[sic] sodium and a reducing course of prednisolone" Note *mycophenolate* is spelt incorrectly.	p73 Table 6	Thank you for your comments. These errors will be amended prior to publication
The Symphony Study regimen also included daclizumab; does this unit use basiliximab instead, and can this be stated?		Yes, basiliximab is used, not daclizumab
"In addition, the following websites were searched for background information…" Given this is an assessment of kidney transplantation I am surprised to note that **no** transplant website has been accessed, including the British Transplantation Society, the European Society for Organ Transplantation, or the American Society for Transplantation.	p80	We usually get exposure to the meeting/ conference abstracts through Embase and Web of Science, where we are purposefully looking for meeting abstracts. However, we agree it may a limitation that we did not access the transplant websites.
"The majority of trials report outcomes up to one year, with the period of induction therapy generally continued for up to 14 days, however, this was occasionally unclear since length of treatment varied according to participant trough levels." Neither of the two induction agents under study, basiliximab or rabbit ATG, are dosed by participant trough levels. This paragraph does not make sense.	p93	We agree. This will be clarified prior to publication.
The ERG has used incorrect drug acquisition costs within their model, driven by inconsistently applying discounts to the list price offered on tac-IR and tac-PR (as found on the 'drug costs' sheet of the ERG model, and subsequently throughout the health economic discussion). The EMIT database was used to inform the cost of tac-IR, and includes discounts offered to the NHS. However, the BNF publication was used to inform the cost of tac-PR (Advagraf) which does not include discounts		We do not accept that incorrect drug acquisition costs have been used. The NICE reference case states (5.5.2) that: "The public list prices for technologies (for example, pharmaceuticals or medical devices) should be used in the reference-case analysis. When there are nationally available price reductions, for example for medicines procured for use in secondary care through contracts negotiated by the NHS Commercial Medicines Unit, then the reduced price should be

offered to the NHS. The consequence of this inaccuracy is that the modelled cost of tac-PR is substantially higher than that paid by the NHS. Astellas estimate that the actual acquisition cost of the two tacrolimus formulations to be similar once all discounts are applied, and would be happy to confidentially share this with the ERG and Committee.		used in the reference-case analysis to best reflect the price relevant to the NHS. The Commercial Medicines Unit publishes information on the prices paid for some generic drugs by NHS trusts through its Electronic Marketing Information Tool (eMIT); focusing on medicines in the National Generics Programme Framework for England." Our interpretation of this is that list prices for pharmaceuticals (from the BNF for secondary care) should be used unless one of the following applies: 1) The pharmaceutical is listed in the eMIT database 2) Details are supplied by the company of contracts negotiated with the NHS Commercial Medicines Unit In the case of immediate-release tacrolimus, (1) applies, but for prolonged-release tacrolimus neither (1) nor (2) applies. Following paragraphs are commercial in confidence (and were not included in our report): : [redacted] [redacted] [redacted]
Astellas would also like to register a concern with the ERG document, that the comparative efficacy of tac-IR and tac-PR	Table 57 to 59	Point estimates and confidence intervals from non-inferiority trials are just as valid as point estimates and confidence intervals from superiority

has been misrepresented, by a) the reliance on non-significant point estimate values taken from non-inferiority trials (Tables 57 to 59) despite acknowledgement that the two formulations are non-inferior, and b) the exclusion of evidence describing the benefits of once daily dosing i.e. Advagraf has demonstrated improved adherence and reduced variability in tacrolimus exposure, both of which have been shown to improve graft survival in multiple RCTs and the exclusion of real world evidence experiences with these two formulations.		trials – the study objective affects the power calculations but fundamentally the trial design is unchanged. Uncertainty in the relative effectiveness is appropriately propagated through the economic model in the PSA. To use adherence or variability in exposure as a surrogate relationship in an HTA, we suggest that recommendations are followed, e.g., Taylor et al. 2009 [Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports. Health Technol Assess 2009;13(8)] to ensure that high-quality evidence is informing the surrogate relationship, and that uncertainty is explored. Furthermore, as noted on page 351, it is necessary for the same surrogate relationship to be applied to all arms being compared, i.e., it is not legitimate to consider comparisons between prolonged-release tacrolimus and other agents except immediate-release tacrolimus.
Itemised differences between PenTAG and Novartis models and model re-run		See pro forma
EVL The PenTAG assessment has not considered the use of EVL in specific subpopulations [page 550, Section 8.2.2]. The Novartis systematic review demonstrates that the addition of EVL to a ciclosporin-based regimen allows for reduction in ciclosporin dose while demonstrating a lower risk of developing CMV infection compared with MMF and standard-dose ciclosporin, with no significant differences in efficacy and other safety parameters. This is of significance because CMV disease is associated with inferior long-term patient and graft survival (Hartmann 2006, Tedesco-Silva 2013). The published NICE scope makes provision for the analysis of interventions in 'people at high risk of complications from immunosuppression'. In our view, EVL patients considered by their physician to be at high risk from CMV infection (eg, donor+/recipient patients, patients with recurrent CMV infection, patients experiencing CMV infection despite prophylaxis, ATG induction immunosuppression) should be considered a specific		We conducted a network meta-analysis of studies reporting CMV infection comparing mTOR-I use with no mTOR-I use. This concluded that there is significant evidence to suggest that CMV infections are less frequent when mTOR-I replace an antimetabolite (e.g., MMF), OR 0.316 (95% CrI 0.124–0.888), and also evidence to suggest that CMV infections are less frequent when mTOR-I replace a calcineurin inhibitor, although this was not quite statistically significant. On this basis it was assumed that the probability of a KTR not receiving mTOR-I experiencing CMV infection would be 10.7%, while this would be 3.7% if mTOR-I replaced an antimetabolite and 5.1% if mTOR-I replaced a calcineurin inhibitor. We acknowledge that in the PenTAG economic model CMV disease is not modelled separately to CMV infection (which in some studies may include asymptomatic viraemia) and is not used to predict long-term patient or graft survival. We do not believe there to be any published RCTs comparing mTOR-I

<table>
<tr>
<td>subgroup in which EVL use is supported based on the following data:

In the randomised study, B201, CMV infection was significantly less likely at 36 months in both EVL groups than in the MMF group (5.7% and 8.1% in the EVL 1.5 mg/day and 3 mg/day groups versus 19.9% in the MMF group; p=0.0001) (Vitko 2005). Another randomised trial compared the effect of EVL with that of MMF plus reduced-dose tacrolimus on the incidence of CMV infection in renal transplant recipients for 12 months. Preliminary data up to 239 days indicate that patients receiving EVL are at lower risk of developing CMV infection compared with those receiving MMF, with no significant differences in efficacy and other safety parameters (Tedesco-Silva 2013).

Novartis, therefore, requests that NICE should consider the use of EVL in patients at high risk from CMV infection, because CMV disease is associated with inferior long-term patient and graft survival.</td>
<td></td>
<td>use to no mTOR-I use in patients at high risk of CMV disease. We note that Vitko 2005 do not present any subgroup analyses for their CMV infection results, and that Tedesco-Silva 2013 is in abstract form only (although it considers low immunological risk patients receiving rATG but no CMV prophylaxis).

As an exploratory analysis (not included in our report) , we increased the baseline rate of CMV infection to 23.1% (the rate of CMV infection in D+/R− KTRs in [Harvala 2013]) to simulate a population at high-risk of CMV infection.

In this exploratory analysis the following key results were observed:
• SRL continued to be dominated by TAC and CSA in combination with BAS+MMF, and by MMF and AZA in combination with TAC
• The ICER of EVL versus MMF (in combination with CSA) was £1.5m per QALY

In fact, these results were robust to any baseline rate of CMV infection.

This suggests that only structural uncertainty (i.e., the omission of CMV disease as a predictor of patient and graft survival) could result in mTOR-I being cost-effective at £20,000 to £30,000 per QALY.

[Harvala 2013] Harvala H, Stewart C, Muller K, et al. High risk of cytomegalovirus infection following solid organ transplantation despite prophylactic therapy. J Med Virol 2013; 85:893-898</td>
</tr>
<tr>
<td>Systematic review inclusion/exclusion criteria and study selection
Only randomised controlled trials (RCTs) or systematic reviews of RCTs were included as part of the study selection process to assess clinical effectiveness. While Novartis appreciates that this is a standard and well-recognised approach undertaken for systematic reviews in the context of a health technology assessment (HTA), by only including RCTs the evidence appraised does not provide the complete, long-term, real-world evidence (RWE) database available (eg, open-label extension studies, conversion studies, registry data, etc). Differences</td>
<td>4.1.3.1 p80</td>
<td>The NICE methods guidance is clear about the types and quality of evidence that should be used to inform its guidance to the NHS

<table>
<tr><th>First author and year</th><th>Reason for exclusion</th></tr>
<tr><td>Budde et al. 2004[271]; Budde et al. 2005[272]; Budde et al.</td><td>Intervention
Study design – RCT was at least 6 months post-transplant</td></tr>
</table>
</td>
</tr>
</table>

between EC-MPS and MMF, such as incidence of adverse events and effect on quality of life, are more likely to be evident in RWE, as such studies are designed to capture these outcomes. Indeed, it is becoming more widely recognised that RCTs are not the only evidence base for decision-making – assessing the value of a drug or technology also requires an understanding of its impact on current management in a practical, real-life setting (Annemans 2007). We propose that the exclusion of key studies from the PenTAG systematic review and exclusion of data from these key studies in the NMA has impacted clinical and cost-effectiveness outputs when EC-MPS is considered in the full renal transplant population. Table 140 from the PenTAG report [page 313, Section 4.5.4] lists the studies included in the Novartis submission and states whether they were included or excluded from the PenTAG report. Many have been excluded, some of which we consider should be included, as they present valuable data that support key differences between MMF and EC-MPS. These studies and Novartis' rationale for inclusion are listed in the Appendix. We request more clarity around exclusion decisions for these studies.		2006[273]		
		Shehata et al. 2009[274]	Study design – only patients with GI symptoms eligible. Also not known when transplant was in relation to randomisation.	
		Ortega et al. 2011[275]	Study design – subpopulation already with GI symptoms	
		Langone et al. 2013[276]; Chan et al. 2013[277]	Study design – not an RCT	
		Shah et al. 2013[278]	Study design – participants received different induction therapies	
		Langone et al. 2011[279]	Study design – subpopulation of individuals with GI symptoms, also unclear point of randomisation	
		Chan et al. 2008[280]	Study design – comparison of TAC doses, identical intervention and comparator	
		Hwang et al. 2010[281]	Study design – not an RCT	
		Paoletti et al. 2012a[286]; Paoletti et al. 2012b[287]	Study design – no independent comparator - EVL+CsA vs CsA	
		Favi et al. 2009a[288]; Favi et al. 2009b[289]; Favi et al. 2010[290]; Favi et al. 2013a[291]	Study design – not randomized – consecutive patients prospectively assigned.	
		Gonzalez et al.	Study design – not an RCT	

		2010[292]	
		Miserlis et al. 2008[293]	Abstracts are not included in the systematic review
		Watarai et al. 2013[294]	Abstracts are not included in the systematic review
		Loriga et al. 2010[295]	Study design – not an RCT
		Tedesco et al. 2012[299] ; Tedesco-Silva et al. 2013[300]	Abstracts are not included in the systematic review
		Favi et al. 2012[257]; Favi et al. 2013b[301]	Abstracts are not included in the systematic review
		Kamar et al. 2005[302]; Rostaing et al. 2001[303]	Study design – not an RCT
Four studies have been inappropriately included as Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) studies. Samsel (2008), Kyllönen (2007), and Sheashaa (2008) used ATG-Fresenius and Sollinger (2001) used equine-derived antithymocyte globulin (ATGAM), not Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) and ATGAM are anti T and B lymphocyte immunoglobulin products derived from thymus lymphocyte immunisation, but rATG is rabbit derived whereas ATGAM is horse-derived. ATG-Fresenius is an anti-T-lymphocyte	Section 4.2 Page 94: Table 9	Please see addendum	

immunoglobulin product derived from rabbits immunised with Jurkat cells, a lymphoblastoid cell line. Both ATGAM and ATG-Fresenius differ inherently from Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]), and the three products cannot be considered interchangeable. The doses, activity and targets of the three products have been assessed seperately, and although they may be used in various induction protocols in a similar manner, there are numerous pieces of evidence to suggest that that they are not biosimilar products, and are in fact inherently different. Additionally, it has been suggested that successful induction with these products may not just be due to lymphocyte-depleting properties, but may also involve immunomodulation of other cell types, including regulatory T cells and dendritic cells (Leitner 2010; LaCorcia 2009). It is not difficult to conceive that even subtle differences in antibody content between the different products could significantly and differentially affect their activities as induction therapies. The efficacy and safety of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) and ATGAM have been compared in a prospective randomised double-blind single-site study (Hardinger 2008), in which significant outcome differences were identified, including the composite endpoint of freedom from death, graft loss or acute rejection at 10 years (in favour of Thymoglobulin). To our knowledge, there have been no randomised controlled studies that have directly compared the efficacy and safety of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) and ATG-Fresenius for induction of kidney transplant patients, to discern similarity or otherwise. However, one randomised comparative study of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) versus ATG-Fresenius, assessing haematologic adverse effects at one month, has been reported in kidney transplant patients (Rostaing 2010). This study concluded that reduced platelet and reticulocyte counts occurred less frequently with Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin,		

[Sanofi]), accompanied by a reduced requirement for erythropoietin-stimulating agents versus ATG-Fresenius. Another recent publication demonstrated that Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) and ATG-Fresenius differ significantly in terms of regulation of the intrarenal transcriptome (Urbanova 2012). Two retrospective series also have indicated significant differences in terms of white blood cell count recovery rates (Cicora 2013) and rates of delayed graft function and acute rejection (Chen 2015). Finally, in regard to the idea that Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) may benfically have immunomodulatory activity in addition to lyphocyte-depleting, Leitner 2010 determined that Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) and ATG-Fresenius targets dendritic cell antigens with differential activity. All four non-Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) studies addressed in the assessment group's report have been incorporated into subsequent network meta-analyses for a variety of outcome measures.There is no rationale or justification provided in the assessment group's report for combining data from studies involving different lymphocyte-depleting products. The European Public Assessment report for paediatric studies of Anti-human T-lymphocyte immunoglobulin from rabbits states that although Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) and ATG-Fresenius have the same ATC code they actually constitute two different drug products (EMA 2014). Furthermore, please find the following quotation from an editorial piece (Mohty 2012); "Despite sharing some common properties, these different products are strictly different drugs, and have never been compared thus far in a controlled setting." A recent review (Hardinger 2012) also stated; "A recent meta-analysis of six randomized studies including 853 patients showed no differences between ATG and basiliximab for the outcomes including BPAR, delayed graft function (DGF), graft loss, and patient death [Liu 2010]. A major limitation to this meta-analysis was that authors included all preparations of		

ATG in the analysis". The Liu meta-analysis was not considered in the assessment group report. We believe that it is clearly incorrect to combine analyses from studies involving different products that are considered non-interchangeable. We therefore recommend that all relevant analyses should be rerun with the exclusion of Samsel (2008), Kyllönen (2007), Sheashaa (2008) and Sollinger (2001).		
Although the scope for this MTA included immunological risk as a relevant parameter in defining subgroups, the assessment group did not present any subgroup analyses. Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) is a potent immunosuppressing agent, used as induction therapy to suppress acute rejection of kidney transplants. However, the implications of over-suppression are potentially increased infection and malignancy rates. Therefore, it is not only important to balance safety and benefits with appropriate dosing, it is also critical to acknowledge that the risk:benefit profile may differ for patients with different levels of rejection risk. It is our understanding that in clinical practice Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) is predominantly used in patients with a moderate/high immunological risk status, where a powerful immunosuppressant may be most justified. Use of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) in this setting is supported by the 2009 KDGIO guidelines and a more recent review by Hardinger (2013). Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) induction may be used in low-risk kidney transplant patients on occasion, to support calcineurin inhibitor or steroid-sparing regimens. Randomised trials of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) versus no induction or placebo in populations selected according to higher immunological risk have not been performed. However, one randomised trial has shown a lower rate of acute rejection for Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin,	4	Thank you for your comments. Unfortunately, none of the included trials were reported in such a way that subgroup analysis was possible. With regard to level of immunological risk, we will clarify the differences between the RCTs prior to publication. We agree the report is limited due to a lack of evidence in this area.

[Sanofi]) versus basiliximab induction in patients at moderate to high immunological risk (Brennan 2006). Although studies of daclizumab (another IL-2 receptor antagonist) are not in scope for this appraisal, it should be noted that comparisons of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) with daclizumab are consistent with the Brennan findings (Noël 2009). Three other studies in lower risk patients found no significant benefit of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) versus IL-2 receptor antagonist induction (Lebranchu 2002; Mourad 2004; Abou-Ayache 2008). The benefits of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) are therefore likely to be concentrated in moderate or high-risk populations. The EMA guideline on clinical investigation of immunosupressants for organ transplantation (2009) states that best attempts should be undertaken to define immunological risk at baseline (e.g. low/medium/high). The Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) studies included in the assessment group's report recruited patients with different immunological risks, potentially resulting in different outcomes. Of the five Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) studies included in the report, two studies recruited patients with low/medium immunological risk (Lebranchu 2002; Mourad 2004) , one study selectively recruited high risk patients (Brennan 2006) and two studies were non-selective (Charpentier 2001; Charpentier 2003). In order to provide clinically relevant guidance to the NHS we believe it is important to compare the immunosuppressive efficacy of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) versus basiliximab according to the category of immunological risk.		
The dose of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) used in some of the included studies no longer reflects clinical practice. Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin,	4.2	Thank you for your comments. These concerns will be reflected in the report prior to publication. We agree that a more detailed consideration of dose would be beneficial.

[Sanofi]) is a potent immunosuppressing agent, used as induction therapy to suppress acute rejection of kidney transplants. However, the implications of over-suppression are potentially increased infection and malignancy rates. Therefore, it is important to balance safety and benefits by appropriate dosing. It is understood that the high cumulative dose (12.5 mg/kg) of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) used in Charpentier (2001) and Charpentier (2003) is no longer used in kidney transplantation, as the increased risk may not be offset by the benefits. Based on pharmacodynamic studies (Büchler 2013; Kho 2012), a cumulative Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) dose of 6 mg/kg is generally considered appropriate for induction therapy related to kidney transplantation. For example, the large ongoing SAILOR study is currently administering a total dose of only 5 mg/kg (Ekberg 2014). Thus, these early studies in which a dose of 12.5 mg/kg was utilised no longer reflect current practice. Given the dose-dependency of possible side effects related to Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) induction, that could potentially affect graft survival, such as post-transplant lymphoproliferative disease (Marks 2011; Hertig 2015), infection and malignancy, outcomes of Charpentier (2001) and Charpentier (2003) should not be considered relevant to current practice.		
We believe that two studies have been cited in Table 30 in error. Margreiter (2002) is a study of tacrolimus versus cyclosporine. Charpentier (2003) is a study of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) versus no induction. It is possible that Mourad (2004) was intended to be cited here, since the publication states that all episodes of BPAR in both treatment arms were borderline or grade I. Another reference which the authors may have intended to cite is Sollinger 2001. However, Sollinger (2001) analysed ATGAM and therefore does not apply to this analysis of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) versus basiliximab. Thus, the comment ‘A statistically significant difference is seen for the most severe Banff	Section 4.3 Page 145: Table 30	Thank you. These errors will be rectified prior to publication.

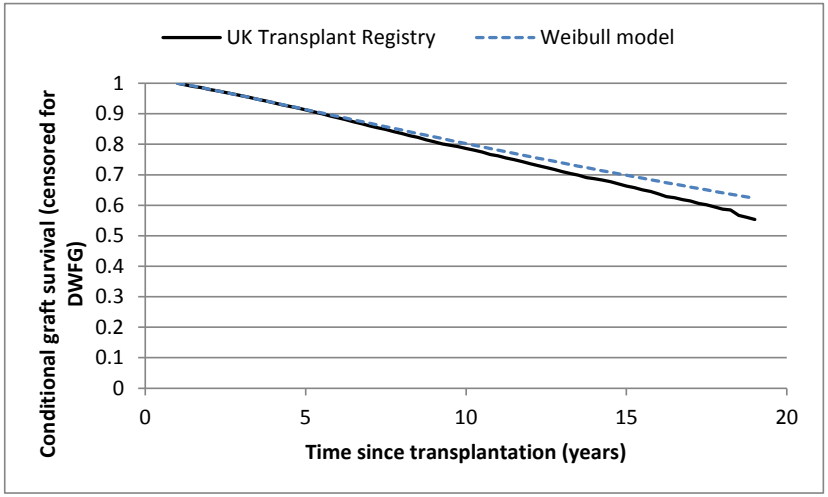
classification 3 in favour or BAS (OR 0.04, 95%CI 0.00 to 0.65), which is reported by Sollinger (2001) should be omitted since it does not relate to Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]).		
We would like to make the assessment group aware of a new "in press" article in Transplant Immunology regarding the risk of post-transplant lymphoproliferative disease (PTLD) associated with rabbit anti-human thymocyte immunoglobulin induction (Hertig and Zuckermann 2015). This pooled analysis concludes "Overall, the risk of PTLD following rATG induction therapy with modern dosing regimens and under current management conditions appears unlikely to make an important contribution to the risk:benefit balance."	Section 4.3.1	Thank you for your comment, which will be considered prior to publication of the report. I believe this paper refers to pooled data from a previous systematic review.
Inappropriate inclusion of studies. Given our concerns (detailed above) regarding the inappropriate inclusion of some studies for Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]), the effectiveness inputs and therefore the outputs of the assessment group's economic model are subject to considerable uncertainty. We recommend that the network meta-analyses used to inform the economic model and the subsequent deterministic and probabilistic results of the economic model are updated accordingly. Only five studies involve the use of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]), (Lebranchu 2002; Mourad 2004; Brennan 2006; Charpentier 2001; Charpentier 2003). Only three studies involve the use of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) at current-practice dose levels, (Lebranchu 2002; Mourad 2004; Brennan 2006).	7	See addendum
No consideration has been given to the effect of immunological risk status on the cost-effectiveness of the different treatment regimens. As discussed above, Thymoglobulin® (rabbit anti-human	7	We did not identify any published RCTs which performed subgroup analyses allowing us to estimate the impact of immunological risk on relative effectiveness of the treatments. To assume the results of a **single** trial are representative of relative

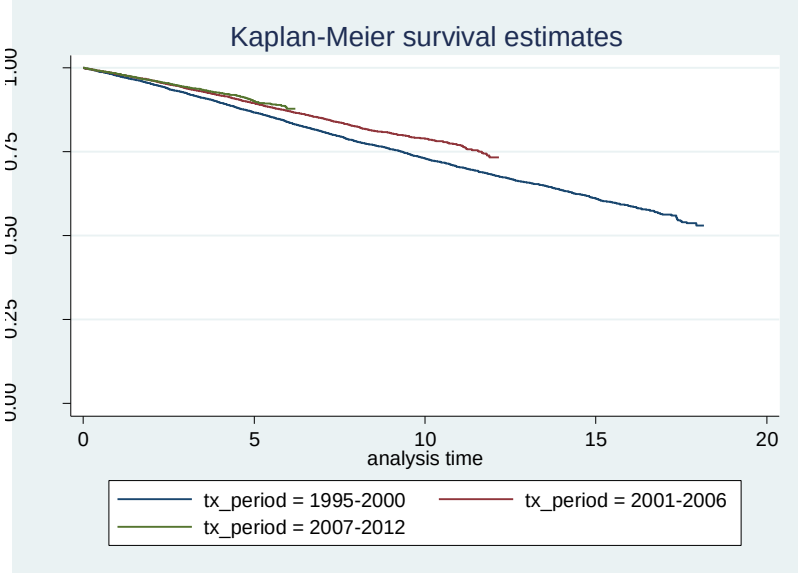
thymocyte immunoglobulin, [Sanofi]) is predominately used in moderate/high immunological risk patients on the basis of evidence demonstrating that it offers benefits versus basiliximab in this patient group. In order to provide clinically relevant guidance on the use of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]), an assessment of cost-effectiveness should be conducted separately for patients with low and moderate/high immunological risk. Only one study involves the use of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) at current-practice dose levels, and in moderate/high risk patients (Brennan 2006).		effectiveness in a particular subgroup and ignore the results of other trials, would be, we feel, speculative. If the Committee believes an exploratory analysis based on Brennan 2006 is appropriate to investigate efficacy in this subgroup, it should be noted that eGFR is not reported in Brennan 2006 and therefore could not be used as a surrogate predictor of graft survival.
No rationale has been provided for the higher costs of infection prophylaxis associated with Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]). In Brennan (2006) both Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) and basiliximab arms received prophylaxis.	Section 7 Page 392	We followed clinical expert advice, which is that CMV prophylaxis is indicated for KTRs at intermediate risk of CMV infection (i.e., seropositive) receiving T-cell depleting antibodies. This also accords with the British Transplant Society guidelines recommendations for CMV in solid organ transplants [BTS 2011]: “CMV seronegative recipients who receive a solid organ transplant from a donor who is seropositive should be offered prophylaxis against primary infection. The same should apply where either the donor or recipient is seropositive if the patient is treated with T-cell depleting antibodies.” [BTS 2011] British Transplant Society Guidelines for the prevention and management of CMV disease after solid organ transplantation. Third edition, August 2011.
It is not clear what CMV rates have been included within the model for regimens including Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) induction. In high risk patients groups, Brennan (2006) recorded a Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) mediated CMV infection rate of 7.8% versus 17.5% for basiliximab (with both treatment arms receiving prophylaxis).	Section 7 Page 427	In the model the same CMV rate is assumed regardless of induction therapy, i.e., 10.7%, except where mTOR-I are used, and these regimens are not used for any comparisons between induction therapies. See Table 177 (page 431). . We note that Mourad 2004, also comparing Thymoglobulin to basiliximab, found less CMV infection in the basiliximab arm.
The presentation of cost-effectiveness results is unconventional. An incremental analysis has been provided,	Section 7 Page 472:	The referent treatment for calculation of incremental net health benefit is the regimen with the least QALYs **on the cost-effectiveness frontier**,

however the treatment regimen with the least QALYs has been used as the referent treatment as opposed to the treatment regimen that is the least costly. It should be noted that the large ICERs associated with Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) result from a very small denominator (incremental QALYs) and should be interpreted in this context. It would be helpful if the probabilistic analysis was used to provide confidence intervals for these results.	Table 207	which is equivalent to the least costly regimen on the frontier. The referent treatment for calculation of the ICER is the next cheapest regimen on the frontier. The ordering of treatments on the frontier is unchanged from what it would be if ordered by cost. The departure from convention is extremely minor and solely presentational. We agree that incremental QALYs are small, at <0.1 QALYs for rATG versus BAS. Rather than provide confidence intervals for the ICERs (which are not always helpful due to nonlinearity of the ICER, especially with small denominators), we suggest consideration of the 95% coverage interval of the INHB (at £20,000 per QALY) of rATG vs BAS: • With CSA+MMF: −0.4506 [−0.9410, +0.0265] • With TAC+MMF: −0.4482 [−0.9414, +0.0148] • With CSA+AZA: −0.4203 [−0.9498, +0.0948]
The assessment group's model assumes that there is no difference in monitoring costs across treatment regimens. As acknowledged in the report, clinic visits are likely to be a key driver of costs. Popat (2014) report cost savings associated with the use of Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]) compared to IL-2 receptor antagonists. This is partly driven by a lower number of clinic visits for patients who received Thymoglobulin® (rabbit anti-human thymocyte immunoglobulin, [Sanofi]). We acknowledge that there are important limitations of this study (as highlighted in the report); however given the relatively small and uncertain difference in cost between the induction regimens in the assessment group's cost-effectiveness analysis we believe that this study should be taken into consideration.	Section 7 Page 450	The study by Popat 2014 was a before and after study with an imbalanced distribution of gender and race characteristics between the IL-2 and ATG induction arms. The study reported healthcare costs up to 1 year post-transplantation and although this was not clearly reported, the per patient cost due to clinic visits in the ATG was found to be 36.2% lower than those of the IL-2 receptor antagonist arm. Since the study did not adjust for baseline differences in patient characteristics, the extent to which this difference is due to confounding is unknown. This and the issues discussed in the report related to small samples (n=24 in IL-2 group and N=21 in ATG), limit our ability to derive a robust estimate of relative effect on costs of clinic visits. Nevertheless, evidence from a micro-costing study conducted in the UK that was submitted by one of the companies (BMS) for this appraisal, shows that of the costs of outpatient surveillance incurred for the average patient over the first three years post-transplantation, 68% occurred in the first year. Thus, we have explored the scenario of a relative reduction in clinic visit costs (leaving other monitoring costs unchanged, e.g., therapeutic drug monitoring, graft function monitoring) of rATG induction over basiliximab of 36.2% that lasted only for one year, as an optimistic alternative scenario for ATG. In this scenario analysis, the deterministic results are as follows:

		With CSA+AZA • No induction: £101,595 / 10.7711 QALYs / Dominated • Rabbit ATG: £103,093 / 10.8182 QALYs / Dominated • Basiliximab: £98,244 / 10.9029 QALYs / (Referent) With CSA+MMF: • No induction: £97,429 / 10.9145 QALYs / Dominated • Rabbit ATG: £100,453 / 10.9281 QALYs / Dominated • Basiliximab: £95,219 / 11.0247 QALYs / (Referent) With TAC+MMF: • No induction: £92,226 / 10.8884 QALYs / Dominated • Rabbit ATG: £95,660 / 10.9047 QALYs / Dominated • Basiliximab: £90,405 / 10.9880 QALYs / (Referent) Thus, basiliximab remains the dominant option (figures are based on the new network meta-analysis results).
Vanquoral (ciclosporin) will lower the average net price to the NHS for the same strength and formulation of ciclosporin, although we are not sure whether the price point would result in it being as cost effective as Immediate release tacrolimus.		We included the average acquisition cost for ciclosporin and immediate-release tacrolimus from the eMit database. Just as it may be the case that savings could be made by always using the lowest cost CSA (which may or may not be Vanquoral), this is also true for TAC. See Appendix 9: Table 238 (page 733). We also note that the cost difference between CSA and TAC regimens is determined by both differences in maintenance immunosuppression costs, dialysis costs and NODAT costs.
The authors do not recognise some flaws in trials that limit interpretation. For example the CsA dose and levels in the Belatacept trials (BENEFIT, BENEFIT-ext) are high, and likely to be associated with poor graft function and therefore any apparent improved graft functions in these trials may not be relevant is not relevant. Moreover, graft function with Belatacept is no better than that one sees with MMF and Tac in clinical practice.		Thank you. Where possible, the lower dose of interventions was used for analysis, for example the 'Less Intensive' dose in the BENEFIT trial.

I do not understand the decision to exclude malignancy in the long term outcomes given that this is the leading cause of death in renal transplant recipients and most immunosuppressive agents increase the risk of malignancy. The lack of data on CV disease is a limitation, although post transplant diabetes is included. The analysis is limited to short term follow up, reflecting the paucity of long term data (except in Registries), in clinical trials.		The inclusion of NODAT is justified on the basis that there is a significant increase in incidence in diabetes in the first year post-transplantation and as such can be detected in RCTs with limited follow up, unlike malignancy. Inclusion of malignancy would only make a significant difference to cost-effectiveness if certain agents increase the incidence of malignancy.
The most important omission is of Registry data which shows improving outcomes despite increasing comorbidity and immunological risk, which might usefully have been combined with the analyses performed to make practical recommendations.		We acknowledge the value of registry data, in particular the UK Transplant Registry standard dataset and UK Renal Registry annual reports (which were used as a key data sources for baseline disease natural history). The decision problem addressed in our report relates to the comparative effectiveness and cost-effectiveness of different immunosuppressive agents used in combination; our understanding of the registries is that they have not historically recorded immunosuppressant use with sufficient detail to produce comparative effectiveness estimates. Registries will also generally only represent treatments in common usage, so it is unlikely that registry data could inform comparative effectiveness and cost-effectiveness estimates for, e.g., everolimus or belatacept.
An analysis of higher potency regimens or triple maintenance therapy for patients at higher immune risk (e.g. those with high panel reactive antibody/calculated reaction frequency; vascular rejection; antibody mediated rejection; repeat transplantation) would have been of value.		Thank you for your comment. Unfortunately the included RCTs did not provide evidence for these subgroups.
The Department of Statistics and Clinical Studies at NHS Blood and Transplant, had specific concerns about the validity of the statistical model applied in the appraisal i.e. the Weibull model for survival, conditional on survival to one year, for the following reasons:		Comments below.
It is argued that the Cox Snell residuals indicate no problem, but this is not considered to be a good way to validate the fit.		The authors have not previously heard any criticism of the use of Cox-Snell residuals to validate model fit. There are, of course, other methods of validation which could have

		provided further evidence of goodness-of-fit.
In general, their experience has shown that the Weibull is not a good fit for post-transplant outcome, as it tends to level off more slowly than survival data suggest (see fig in main doc)		The figure provided by NHSBT is a fit of Weibull curve to survival from time of transplant, whereas we have fitted to conditional survival after one year post-transplantation, to acknowledge the high initial rate of graft loss.  This shows graft survival censored for DWFG from the UK Transplant Registry standard dataset (black solid) versus the Weibull model (blue dashed) from 1 year onwards. As can be seen the Weibull model appears to over-predict survival from 10 years onwards, but this is likely to be because censoring is not uninformative – KTRs from more recent cohorts are more likely to be censored.

		
Use of the Weibull model could lead to higher predicted survival rates and higher medians. Fig 92 suggests a median graft survival of over 25 years, which seems rather high. The figures in Table 162 look more realistic, perhaps because the fit is better in shorter time periods.		It should be noted that Figure 92 and Table 162 give graft survival censored for DWFG. At shorter time periods the effect of censoring is less pronounced and graft survival closely matches “transplant survival”. It should also be noted that Hudson and Collett from NHSBT Statistics and Clinical Studies predict median survival for DBD/DCD grafts to be 21-22 years and higher for living donors (60% at 20 years). They also predict median transplant survival to be 13 years for DBD and 20 years for living donors – we predict median transplant survival of 15.5 years.[http://www.odt.nhs.uk/pps/estimating_kidney_survival_using_period_analysis.ppt] If the Appraisal Committee desires, it will be possible to conduct sensitivity analyses in which the parameters of the Weibull fit are adjusted to achieve desired median or mean death-censored graft survival, and therefore to examine the impact on cost-effectiveness. As a guide, if baseline graft survival is reduced, we would expect this increase costs and reduce QALYs for all regimens, plus attenuate the incremental costs and QALYs due to differences in graft survival

<table>
<tr>
<td></td>
<td></td>
<td>As an example, looking at TAC vs CSA vs SRL vs BEL (with BAS+MMF), and adjusting median baseline graft survival to 23 years (from 26.75 years, by adjusting the λ parameter) we find:
<table>
<tr><th>Regimen</th><th>Base case (costs / QALYs / ICER)</th><th>Reduced graft survival</th></tr>
<tr><td>SRL</td><td>£114.5k / 10.90 / Dominated</td><td>£118.6k / 10.76 / Dominated</td></tr>
<tr><td>TAC</td><td>£90.4k / 10.99 / (Referent)</td><td>£95.2k / 10.85 / (Referent)</td></tr>
<tr><td>CSA</td><td>£95.2k / 11.02 / £131k</td><td>£100.3k / 10.88 / £196k</td></tr>
<tr><td>BEL</td><td>£209.4k / 11.29 / £424k</td><td>£209.4k / 11.16 / £391k</td></tr>
</table>
In this scenario, the principal cost-effectiveness conclusions across all comparisons were unchanged from the base case – basiliximab induction and maintenance with immediate-release tacrolimus and mycophenolate mofetil are the only options found to be cost-effective at £20,000 to £30,000 per QALY.

To conclude, we do not believe that graft survival has been significantly overestimated, and we do believe the choice of Weibull extrapolation after the first year is justified. In the event that different decisions are made regarding the most appropriate baseline graft survival, we do not believe these are likely to affect cost-effectiveness results materially.</td>
</tr>
<tr>
<td>This appraisal, typically, focused upon the cost effectiveness of immunosuppressive therapy for adult kidney transplantation. The assessment included an extensive review of Randomised Controlled Trials, which may not reflect ‘real life’ nor the diversity and complexity of the clinical case-mix. Other regimens may, therefore, be required for certain patient groups.</td>
<td></td>
<td>Thank you for your comment. We agree, however, we are constrained to only using RCTs.</td>
</tr>
<tr><td colspan="3"></td></tr>
<tr>
<td>The terms ‘Cadaveric transplant’ and ‘Heart beating donor’ are outdated terms and need to be replaced with ‘Deceased donor transplant’ and ‘Donation after brain death donor’. Also should include a definition of the term, ‘Donation after circulatory</td>
<td>All Exec Summary P25</td>
<td>Thank you for your comments. This will be amended prior to publication.</td>
</tr>
</table>

death'		
'medically suitable people' rather than 'people' as currently only 20-30% of patients on dialysis are considered medically suitable to transplantation	1.1 Para 2, line 2	Thank you. We will clarify this prior to publication
question the use of term 'healthy' kidney and what authors mean by this. Increasingly kidneys come from extended criteria kidney donors where there may be some disease or damage of the kidney, but where the risk/benefit ratio determine that transplantation is still the right option.	1.1 Para 3, line	See comment above.
Update activity date with 2014/15 figures, overall increase in transplant activity and follow up activity since 2008, and total number of patients with a functioning transplant. This data is available from the UK Transplant Registry (NHSBT) and Renal Registry.	1.1 Para 6.	Thank you. We will update figures prior to publication
Three service provision steps is an inadequate description of the transplant patient pathway. The recognised steps in the patient pathway are: Pre-transplant assessment of the recipient and listing; maintenance on the list, transplant episode, and transplant follow-up. Immunosuppression is part of the 'transplant episode' and 'transplant follow-up'. Alongside this pathway is the selection of the organ donor whether that be living or deceased.	1.1 page 29	Thank you for your comments. These additional steps will be included prior to publication.
We assume you have chosen to only assess those interventions with UK marketing authorisation at the time of compiling this report. However this is an incomplete list of interventions currently used for induction or maintenance treatment and other agents to be considered are: • Alemtuzumab – see www.3cstudy.org/results.html • Azathioprine – still in common use and referred to later in the executive summary; so should it not be included as an intervention? • Corticosteroids – see comments above for azathioprine • Envarsus - new prolonged release tacrolimus	1.1 Interventions p29-30	We have reviewed only RCT evidence of the interventions listed in the protocol, as detailed in the NICE scope.
Our understanding of the current literature is that most RCTs of immunosuppressive agents will be focussed on relative short term outcomes (up to 3-4 years, although most only one year) whereas what is becoming increasingly important is longer term outcomes for the patient and graft which are not usually	P31 Outcome Measures	Thank you. We agree this is a limitation of the report.

included within RCTs		
The list of maintenance therapies appears incomplete and this may reflect there were no studies identified for other combinations in routine use. The following observations are made: • All combinations will more often than not also incorporate a corticosteroid; although some will be corticosteroid free. • In clinical practice MMF or MPS can be used in all combinations where an antiproliferative is required • Within current UK practice Everolimus and Belatacept are not routinely used; and use has only been in the context of clinical trials or on a named patient basis. • Within the list of interventions on p29-30 a clear distinction is made between immediate release tacrolimus and prolonged release tacrolimus, yet there is no distinction between these in the list of maintenance therapies on p34. Should they not be looked at separately?	1.1 Data synthesis p34	Yes, we agree many regimens also included corticosteroids and this will be clarified prior to publication. However, for the sake of clarity where intervention regimens are displayed versus comparator regimens, this was not included. Thank you for your comment. We agree RCTs do not necessarily reflect clinical practice, however, we are required to report evidence for the interventions listed in the protocol. The list on p34 only refers to TAC, since there were an insufficient number of trial participants for TAC-PR to be included in the model.
Is it really valid to consider economic evaluations from other countries such as USA, Canada, Australia, and western Europe, where the funding models for healthcare are quite different?	Study selection, bullet 4, page 35	The evidence on cost-effectiveness of specific therapies in these other countries may not be generalisable to the UK, due to differences in the patient populations across these different healthcare systems, as well as in the economic costs of healthcare. However, the studies serve to complement the UK specific evidence reviewed in the report, pointing to key resource use drivers, relationships between surrogate markers and clinical outcomes, health related quality of life and resource use. In addition they help to inform the most robust modelling methodology applied to the evaluation of immunosuppressive therapy for kidney transplantation. The report has highlighted the issues of lack of generalisability of these studies and separately reviewed the UK specific cost-effectiveness evidence.
‘Corticosteroids were assumed to be used in all regimens but at a tapered dose.’ Can’t necessarily make this assumption unless clearly specified	P37	This refers to the modelled usage, rather than being an assumption about the trials informing the effectiveness estimates.
Need for consistency with earlier lists and assuming combinations used in clinical practice and not mentioned are because no published RCT in this area, but for completeness the following regimes should be considered as they are in	1.3.4.1.	Thank you for your comment. We agree that this report is limited by including only those combinations used in RCTs

widespread clinical practice: • Regimens following no induction immunosuppression: TAC-PR + Aza • Regimens following basiliximab induction and Regimens following rabbit ATG induction: should include all of the regimes included under those with no induction immunosuppression. We suspect the problem will be that a number of combinations in widespread clinical use will not have been part of a clinical trial and the results of this MTA will obviously be restricted to those combinations which have been subject to a RCT. Caution will therefore need to be exercised in terms of conclusions.		
There are two other issues that require to be factored in or assumptions made: • In clinical practice there will often be a switch between immunosuppressive agents based on episodes of rejection, adverse events, graft function etc. Some of these switches will have been permitted with a trial protocol and others will have resulted in the patient being withdrawn from the study. How will these changes in immunosuppressive factors be factored in? It may be too broad an assumption to assume that patients continue on the regimen they start on, particularly in looking at long term outcomes. • In clinical practice there has been an increasing use of generics which for the critical dose drugs e.g. ciclosporin and tacrolimus require 90-111% equivalence. As there will be very few, if any, RCTs looking at the generic versions of MMF, tacrolimus and ciclosporin (as detailed in the list of interventions) within transplant patients, then an assumption will presumably be made that the different brands of each agent can be viewed as bioequivalents.	1.3.4.3	Intention-to-treat analysis of clinical effectiveness in the RCTs reflects the issues of treatment switching or withdrawal, which may occur in clinical practice. We are restricted to RCTs and therefore have not had the opportunity to investigate the bioequivalence of generic preparations.
Drug costs do vary in practice from list price, which may add complexities to the analysis. Currently some drugs are on a national tender with significant discounts; the price of others are negotiated at a local level; whilst others are purchased at list price.	1.3.4.4 Costs	Where there was a clear indication that NHS hospitals can obtain drugs at a consistent discounted price (i.e., as shown in eMit), this was included in the analysis (following the NICE reference case). Likewise, where there was not a clear indication in the public domain that such discounts could be obtained, list prices were used (again, following the NICE reference case). Our understanding is that if manufacturers would like the cost-

		effectiveness of their drugs to be appraised at a lower than list price but they do not appear in eMit the usual course of action is negotiation through the Commercial Medicines Unit or to offer a Patient Access Scheme.
This section comments, 'the RCTs were of variable quality, but all appear to be flawed' and more than half the studies '...either did not report, or lacked clarity on, at least five of the ten items constituting the quality appraisal assessment.' This does raise concern regarding the validity of the conclusions of the benefits and risks in section 1.4.2 when in addition to the poor quality, the majority of studies only looked at one year data which is not a good surrogate marker for long term outcome.	1.4.1 and 1.4.2	Thank you for your comments. We agree that the report does have its limitations due to the frequently poor reporting, the quality of the trials and the lack of long term data.