EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

CONFIDENTIAL
Doc Ref: EMADOC-360526170-1021482
Case No.: EMA/SA/0000077234
Human Medicines Division

Initial Qualification Procedure List of Issues

iBox Scoring System (Composite Biomarker Panel)

Summary

The proposed iBox Scoring System is intended to be used as a surrogate endpoint for efficacy in clinical trials evaluating the safety and efficacy of novel immunosuppressive therapies (ISTs) in kidney transplant recipients as a marker for the probability of long-term allograft survival. The proposed iBox Scoring System is a composite biomarker panel used at one-year post-transplant to predict the five-year risk of death-censored allograft loss (allograft failure) in kidney transplant recipients. It is intended for use in clinical trials to support evaluation of novel IST applications via CMA. Two iBox Scoring Systems are proposed, the full iBox and the abbreviated iBox scoring without biopsy data.

The Applicant aims to provide an alternative to the historically utilized primary endpoint in clinical trials. The endpoint is intended to be used throughout the development phases and the Applicant aims at regulatory endorsement of the iBox Scoring System as surrogate endpoint in studies with a broad population of renal transplant patients.

The historically accepted clinical trial endpoint for multi-national clinical trials of novel ISTs in kidney transplantation is the composite endpoint of death, graft-loss, biopsy-proven acute rejection (BPAR) and loss to follow-up, often assessed one year after transplantation. With current improved standard of care ISTs, the event rates may be low in settings of post-transplant trial of an attainable period. The composite endpoint approach results in equal weighting of all-cause death, BPAR and loss to follow-up. While trial endpoints do not target long-term outcomes, there is an unmet need for improvement in the long-term survival of the transplant recipient and graft. Current IST regimens have improved short-term outcomes, with one-year graft survival rates of approximately 91% after deceased donor transplant. However, long-term graft survival is considerably lower than short term graft survival. The 5- and 10-year graft survival rates after deceased donor kidney transplant based on one estimation are 77% and 56%, respectively (Gondos et al., Transplantation Journal 2013). Development of new ISTs is challenging, as demonstration of improved long-term outcomes would require trials of sufficient duration (e.g., five years) and likely a considerable number of subjects.

Scientific discussion

The proposed Context of Use (CoU) definition for the two composite biomarker panels is use as a surrogate endpoint to predict the five-year risk of death-censored allograft loss (allograft failure) in kidney transplant recipients for use in clinical trials to support evaluation of novel IST applications via CMA. The target population includes adult de novo kidney only transplant recipients from a living or deceased donor. Currently, SAWP is of the opinion that before the CoU can be determined, properties of the iBox scores as surrogate endpoint need to be established and critically reviewed. Further, the Applicant states that it will be necessary to determine if there is clinically meaningful decrease

in transplant recipient survival with new therapy. This notion and the Applicant proposed CoU for the additionally developed All Cause Endpoint score is noted; however, for CMA the clinical meaningfulness of differences should be seen also from benefit risk perspective and should be interpretable considering not only observed parallel but also diverging trends of risks, including all-cause mortality.

SAWP acknowledges the strengths of the model development and validation approach and the extensive and valuable work of the group led by Prof. Loupy (Loupy A et al, BMJ 2019). The initial prospective approach for derivation data collection led to a prediction model that has good predictive performance for clinical endpoint events based on several variables included in a biomarker panel proposed as iBox. The model was internally and externally validated. Based on previous feedback, the Applicant refined the approach and performed additional analysis. Analyses plans prospectively defining the analyses have not been provided. The differences to the initial approach by Loupy et al. are explained. These include a different approach to handling donor specific antibodies (DSA) and pertain to the fixed 1-year time point proposed by the Applicant for the CoU, which was addressed by imputing data for patients who die or lose graft during the first year. Imputation was also applied for UPCR data based on other proteinuria measurements. Inclusion of a new independent set of validation data with the re-development of the iBox score by TTC is welcomed.

It is possible that iBox could support CMA. However, other requirements for CMA need to be taken into consideration to fulfil the requirements for CMA as outlined in the relevant EMA guideline (EMA/CHMP/509951/2006, Rev.1). For the iBox to support CMA, it will have to be able to support a positive benefit-risk balance of the medicine and it needs to be justified that it is likely that comprehensive data post-authorisation will be generated. To achieve this, the timeframe to generate data post-authorization needs to be considered. The ability to generate post-authorization data should not be jeopardized by a very long study duration or other factors that impact on the conduct or completion of the study, e.g. availability of a newly approved medicinal products. Other requirements for CMA include the fulfilment of an unmet medical need and that the benefit of the medicine's immediate availability to patients is greater than the risk inherent in the fact that additional data are still required. There are also several other regulatory approaches available to address safety, and/or efficacy, post approval. Such post-authorisation measures (PAMs) may be aimed at collecting or providing data to enable the assessment of the safety or efficacy (e.g. 'Post-authorisation measures: questions and answers', EMA website).

The development of the iBox focused on a fixed time scheme envisaged by the Applicant (assessment of iBox panel at one-year, full marketing authorisation based on a survival endpoint after five years). A more flexible CoU may have advantages, e.g. in case of safety concerns of a candidate IST which could require a longer study duration before an acceptable risk profile could be concluded for CMA. On the other hand, use of the biomarker panel in phase 2 studies may require an earlier assessment time point. A model that includes 'time post-transplant' as variable for prediction may have advantages over a model with a fixed time point for assessment of variables. For application of the iBox scores in future trials that could include non-inferiority settings, derivation of a non-inferiority margin already at this stage would be desirable.

It can be agreed that the clinical transplant population is heterogenous. This poses challenges to establishing the surrogacy. The proposed target population is 'adult de novo kidney only transplant recipients from a living or deceased donor', i.e. the broad population of adult transplants. The studies included in the qualification exercise represent subjects with varying underlying diagnoses, receiving living related as well as extended donor kidneys, receiving various induction therapies and either CNI or CNI free therapy. The efforts of TTC to acquire subject-level data for development of the proposed surrogate endpoint are acknowledged. Selecting studies (5 out of 31) which contained all variables of interest including biopsy information for the full iBox score is a reasonable approach for development of a score that aims at including biopsy data. The variables included in the composite panel appear to be of clinical relevance, while some thresholds for variables included may require discussion. Considerations on the data source period and change in treatment landscape are missing, and as the data come from a period of considerable duration, a discussion on potential change of the standard of care should be included.

In addition, given the difference (therapy, incidence of acute rejection etc.) between the derivation and validation cohorts important parameters might not be included in the model or need a different weight for the final model to predict treatment effect. This includes the pathological variables and stability of the functional parameters (i.e. eGFR and proteinuria).

A limitation of the data sources is the small number of patients included in therapeutic intervention trials that are important for assessing the change in treatment effects in the proposed surrogate and the clinical endpoint at 5 years. Outcome events derived from randomised controlled trials are currently too sparse to be fully informative for the surrogacy at trial level of the iBox biomarker panel. Such a relation is considered key for establishing full surrogacy of a biomarker-based endpoint.

It is currently unclear if a different development strategy, focusing on a score without biopsy data and an alternative validation exercise, would benefit from access to more data. The abbreviated iBox score shows good performance and contains only three parameters, i.e. eGFR, UPCR and donor specific antibody (DSA). A development strategy for the abbreviated score could have used more datasets available to support surrogacy of the abbreviated iBox, even if such a database would not support the development and assessment of the full-iBox score. With the current approach, the abbreviated iBox without biopsy information is supported by only a minimally larger number of subjects in the derivation data set and was not internally validated. In the external validation dataset, more data without biopsy information are available. Further, it is not clear if a development strategy of an abbreviated model from scratch instead of omitting biopsy related information would lead to the same variables included in the abbreviated model.

Overall, an extensive validation exercise has been performed, comprising internal validation based on prospectively collected data, external validation including randomised clinical studies and a trial-level surrogacy analysis. Results show that the proposed iBox score models are likely suitable for individual predictions of graft loss events with good performance based on c-statistics and with the ability to predict numbers of graft loss events with reasonable margins of error. The quality of the analyses for model development and diagnostics, and the validation including competing risks analysis for the death-censored scores are noted.

Some points regarding the validation exercise need further explanation and discussion before the questions from the Applicant can be answered. The performance of an iBox score restricted to functional renal parameters is not clear, as c-statistics results for the iBox scores for single components (figure 13) do not provide an analysis with functional renal parameters only. Regarding handling of functional renal variables, it is not fully clear if eGFR was derived using the same equation in the derivation and validation datasets and if the handling of this variable could have an impact on validation results. UPCR was imputed for use in the iBox calculation, as the validation datasets assessed either dip stick (BENEFIT, BENFIT-EXT and population from Helsinki University Hospital), 24-hour proteinuria (Mayo) or UACR (Mayo) rather than UPCR. 24-hour proteinuria, UACR and UPCR all reflect protein, or albumin, excretion over 24 hours. It is understood that the extrapolation of spot urine albumin by dipstick was based on a German population with both UPCR and dip stick results. However, fit of the data is not clearly presented. The IQR (middle 50%, figure 16) seems wide. It is not clear if the imputation approach for UPCR using dip stick albumin is fully adequate for external validation of the iBox scores from the derivation dataset. Concerning the use of DSA in the biomarker panel, it is proposed to classify DSA levels as binary variable, based on a threshold of MFI < 1400 or ≥ 1400. The relationship to graft failure (i.e. linearity) and a rationale to categorize DSA based on the threshold is not provided and not clear. The comparability of DSA levels regardless of (laboratory) source is also important. Regarding graft loss events in the included trials, it is noted that reclassification of graft loss outcome events was performed according to Levin et al. (Levin A et al., Kidney International 2020). As a relevant number of events was reclassified, this could have had an impact on overall results, and it is not stated if this approach was predefined.

List of issues to be addressed in writing and during the discussion meeting

Based on the coordinators' reports the Scientific Advice Working Party (SAWP) determined that the Applicant should discuss the following points, before advice can be provided:

Issues to be addressed in writing by 27 April 2022 and during the discussion meeting

Issues on Clinical development

1. **CHMP question:** Please discuss the restriction of the proposed context of use to CMA, considering that the iBox score could be used for proof of concept studies or dose ranging studies, e.g., in phase 2 development. Please also discuss the utility of clinical meaningful differences observed by iBox score for benefit risk assessment considering not only observed parallel but also diverging trends of risks, including all-cause mortality.

   Applicant Response: The analyses supporting this Qualification Opinion for the iBox Scoring System and the associated CoU were intended to support the iBox Scoring System as a surrogate endpoint at one-year post-transplant in a registration-driven Phase 3 trial of a novel IST in kidney transplant recipients. The iBox Scoring System without imputation, including the time post-transplant variable, could also be used in Phase 2 or proof of concept (POC) trials. The CoU could be modified to support use in both settings. It is essential that the Phase 3 surrogate be qualified to stimulate the introduction of novel ISTs to address the unmet needs of kidney transplant recipients, including long-term graft survival.

   The iBox Scoring System is intended to be used as a primary efficacy endpoint in a kidney transplant trial. The full consideration of benefit-risk will require an assessment of the risks associated with the new therapies as measured by adverse events and overall mortality. The iBox Scoring System with imputation accounts for overall mortality in the first-year post-transplant by imputing a worse-case iBox score for recipients who die within the first-year post-transplant. If diverging trends of death are seen, it should be considered in the overall benefit-risk assessment in the context of a superior iBox score. Also, overall mortality will continue to be monitored and assessed as a key outcome after one-year post-transplant.

   A modified CoU statement for consideration is below:

   The iBox Scoring System (Composite Biomarker Panel) ~~used at one-year post-transplant~~ is a surrogate endpoint for long-term ~~the five-year risk of~~ death-censored allograft loss (allograft failure) in kidney transplant recipients for use in clinical trials to support evaluation of novel immunosuppressive therapy applications ~~via conditional marketing authorisation~~.

2. **CHMP question:** Please discuss alternative development approaches and the potential value of extending the database to determine surrogacy of:

   a. the abbreviated iBox, i.e. not including assessment of the full iBox score. Please elaborate on shortcomings of the approach undertaken for the abbreviated iBox score and if relevant model variables may have been omitted with the procedure (i.e., focusing on a full score also for selection of data sets and omitting biopsy variables later),

   b. an iBox score using 3-year graft failure as the outcome measure,

   c. an iBox score with a flexible time for short-term assessment of iBox and flexible prediction of long-term graft loss outcomes.

   Applicant Response:

   a. The abbreviated iBox Scoring System, which includes estimated glomerular filtration rate (eGFR), proteinuria, and donor-specific antibody (DSA), consists of all the available variables that are modifiable

by new ISTs. Therefore, the Transplant Therapeutics Consortium (TTC) is confident that relevant model variables have not been omitted. The 31 candidate variables explored in the derivation of the iBox Scoring System are not consistently present in the qualification validation datasets. Therefore, re-derivation on the qualification validation datasets using the complete list of variables is not feasible. However, the TTC can explore re-deriving the abbreviated iBox Scoring System using the qualification validation datasets with a subset of candidate variables to compare coefficients in the different populations. Additionally, the datasets included in the qualification submission are all the datasets within TTC's Kidney Transplant Database with the core iBox variables for the abbreviated iBox Scoring System (i.e., excluding kidney biopsy data), the necessary follow-up, and demographic/baseline variables, including IST regimen information. Therefore, restricting the analysis to an abbreviated iBox Scoring System will not increase the available data for analyses.

b. Three-year follow-up could be assessed in the current datasets. However, the TTC anticipates that this will reduce prognostic/predictive performance since 1) the number of graft loss events will be reduced compared to the five-year follow-up assessment, and 2) the difference in death-censored graft survival between treatment arms will be reduced, summarized in Table 1 below.

Table 1. BENEFIT and BENEFIT EXT randomized controlled trials (RCTs) 3 year versus 5-year death-censored graft loss by treatment group

Treatment Arm	Belatacept more intensive No. of events	Belatacept less intensive No. of events	Cyclosporine No. of events
BENEFIT RCT			
Death-censored graft loss at 3 years	**3**	**2**	**3**
Death-censored graft loss at 5 years	**3**	**3**	**9**
BENEFIT-EXT RCT			
Death-censored graft loss at 3 years	**2**	**5**	**4**
Death-censored graft loss at 5 years	**5**	**7**	**11**

c. The iBox was originally designed by Loupy et al., 2019 as a patient prognostic tool with flexibility for time (including a short-term assessment) and flexibility in long-term follow-up. The iBox Scoring System can continue to be used in this approach for either individual prognosis or as a Phase 2/POC endpoint. To translate this work into a Phase 3 primary surrogate efficacy endpoint, it is necessary to fix the time of assessment and follow-up, and account for deaths and graft losses that occur before that fixed time of assessment (i.e., one-year post-transplant).

3. **CHMP question:** Please provide information on the performance of an iBox score restricted to functional renal parameters eGFR and proteinuria. Please discuss the feasibility of using functional renal parameters for long-term prediction of graft loss.

Applicant Response: Analyses testing the performance of the iBox Scoring System restricting to functional renal parameters (i.e., eGFR and proteinuria, and eGFR alone) were included in the Briefing Package re-

submitted on 16 February 2022. These results can be found in section 1.4 in the document titled "REVISED-Supporting results." Under the two iBox models with only functional renal parameters, it was found that the hazard ratios for eGFR and proteinuria remained similar, summarized in Tables 2 and 3 below. Therefore, functional renal parameters could be used for the prediction of graft loss. However, there is some loss of discrimination and calibration compared to the full (with biopsy) or abbreviated (without biopsy) iBox Scoring System, summarized in Tables 4 -6 below. The red text colour in Table 4 highlight lower C-statistics, below 0.7. The red text colour in Table 6 highlights statistically significant differences in observed vs. predicted events.

Table 2. Functional renal parameter iBox Scoring System (eGFR and Proteinuria)

Factor	No of subjects	No of events	Hazard ratio	P-value
Time from transplant to evaluation (years)	4000	549	1.16 (1.1 to 1.21)	<0.0001
eGFR (mL/min/1.73 m^2)	4000	549	0.95 (0.95 to 0.96)	<0.0001
UPCR Proteinuria (log g/g)	4000	549	1.6 (1.49 to 1.72)	<0.0001

Table 3. Functional renal parameter iBox Scoring System (eGFR only)

Factor	No of subjects	No of events	Hazard ratio	P-value
Time from transplant to evaluation (years)	4000	549	1.19 (1.13 to 1.25)	<0.0001
eGFR (mL/min/1.73 m^2)	4000	549	0.94 (0.94 to 0.95)	<0.0001

Table 4. C-statistic values for the qualification validation datasets

Dataset	full (+ biopsy) iBox Scoring System (SE)	abbreviated (- biopsy) iBox Scoring System (SE)	Functional renal parameter iBox Scoring System (eGFR and Proteinuria) (SE)	Functional renal parameter iBox Scoring System (eGFR only) (SE)
Helsinki University Hospital	0. 78 (0.06)	0. 77 (0.06)	0.76 (0.06)	0.74 (0.06)
Mayo Clinic Rochester	0.93 (0.03)	0.84 (0.03)	0.80 (0.04)	0.75 (0.04)
BENEFIT RCT	0.70 (0.09)	0.70 (0.08)	0.69 (0.08)	0.69 (0.08)
BENEFIT-EXT RCT	0.81 (0.07)	0.78 (0.06)	0.78 (0.06)	0.78 (0.06)

Table 5. Poisson calibration by dataset, functional renal parameter iBox Scoring System (eGFR and Proteinuria)

Dataset	Observed graft loss	Predicted graft loss	Observed / predicted	z score for observed / predicted	p value
Helsinki University Hospital	22	17.22	1.28	1.15	0.25
Mayo Clinic Rochester	45	57.93	0.78	-1.69	0.09
BENEFIT RCT	15	20.39	0.74	-1.19	0.23
BENEFIT-EXT RCT	23	24.25	0.95	-0.25	0.80

Table 6. Poisson calibration by dataset, functional renal parameter iBox Scoring System (eGFR only)

Dataset	Observed graft loss	Predicted graft loss	Observed / predicted	z score for observed / predicted	p value

Helsinki University Hospital	22	19.40	1.13	0.59	0.56
Mayo Clinic Rochester	51	76.50	0.67	-2.90	<0.01
BENEFIT RCT	17	25.66	0.66	-1.70	0.09
BENEFIT-EXT RCT	23	29.59	0.78	-1.21	0.23

4. **CHMP question:** Please discuss if the time period over which development data were collected could have an impact on future use of the iBox, considering potential change of the treatment landscape.

   Applicant Response: The data collection period for the derivation dataset was from 2005 to 2014. This time period was necessary to ensure adequate inclusion of data with 1) long-term follow-up and 2) various IST regimens. Most kidney transplant patients in the United States (US) and the European Union (EU) currently receive a calcineurin inhibitor (CNI)-based regimen combined with mycophenolic acid (MPA)/mycophenolate mofetil (MMF) with selected patients and centres using mammalian target of rapamycin (mTOR) inhibitors and/or belatacept. Tacrolimus and MMF/MPA were approved in the mid-1990s. Therefore, the datasets in the qualification submission represent the current IST regimens (i.e., most recipients receiving a CNI-based regimen with MPA/MMF) and a subset receiving mTOR inhibitor or belatacept-based regimens. The performance of the iBox Scoring System was assessed in all these regimens and demonstrates that the iBox Scoring System is robust regardless of the immunosuppressive mechanism of action.

5. **CHMP question:** Please discuss if establishing full trial level surrogacy could be an attainable goal and provide an update on efforts to obtain additional data from RCTs for validation.

   Applicant Response: The TTC led an extensive global data collaboration effort across the field of kidney transplantation. Unfortunately, establishing full trial-level surrogacy is not an attainable goal within the next five years based on the currently available RCTs that have the core abbreviated iBox Scoring System variables taken at one-year post-transplant (i.e., eGFR, proteinuria, and DSA) as well as a follow-up period sufficient to evaluate long-term graft survival (i.e., at least five years). A summary detailing the datasets in the TTC's Kidney Transplant Database with associated information describing why datasets were included/excluded in the qualification submission was included in the Briefing Package re-submitted on 16 February. This summary can be found in Tables 1 and 2 in the document titled "REVISED-Transplant Therapeutics Consortium's Kidney Transplant Database." The best approach to obtaining further data in RCTs would be the qualification of the iBox Scoring System as a surrogate endpoint, which would help ensure it is incorporated into future registration RCTs for new ISTs.

6. **CHMP question:** Please discuss how a non-inferiority margin for the iBox scores could be derived.

   Applicant Response: The iBox Scoring System was developed to measure the superiority of a new IST compared to controls, thus addressing the needs for early endpoints that are predictive of long-term survival. There is currently insufficient information on the operating characteristics of the iBox Scoring System to support the derivation of a non-inferiority margin. However, the TTC can envisage that the iBox Scoring System, as a measure of the superiority of the new IST, could be coupled with an assessment of non-inferiority on the historically accepted composite of efficacy failure (death, graft loss, biopsy-proven acute rejection, and loss-to-follow-up).

7. **CHMP question:** Please provide additional information on the reclassifications of graft loss events in the BENEFIT and BENEFIT-EXT studies according to Levin et al. (Levin A et al., Kidney International 2020). The impact on results is not clear and it is not stated if this approach was predefined.

   Applicant Response: After examining the raw patient-level data from the BENEFIT and BENEFIT-EXT RCTs, it was evident that a small number of subjects had poor renal function that would otherwise be considered graft loss but were not coded as such in the database. This was not anticipated and therefore was not predefined.

Therefore, to harmonize the definition of graft loss used across the five historical datasets, TTC examined the relevant literature available for defining graft loss in the context of clinical trials based on objective criteria for renal failure.

Levin A et al., Kidney International 2020 graft loss definitions were used to reclassify graft loss across the qualification datasets. Table 2 of this publication summarizes the international consensus definitions of clinical trial outcomes for kidney failure, in which sustained low GFR and sustained percent decline in GFR were used for reclassifying graft loss in this qualification submission. For the sustained percent decline in GFR, the most rigorous decline of 57 percent was chosen as this corresponds to a doubling of serum creatinine, a well-established (putative) surrogate.

The qualification derivation dataset did not have longitudinal eGFR data to review and therefore was excluded from potential reclassification. The BENEFIT and BENEFIT-EXT RCTs, Mayo Clinic Rochester, and Helsinki University Hospital, were reviewed for potential reclassification of graft loss. The reclassification process impacted the BENEFIT and BENEFIT-EXT RCTs more than the other datasets. No subjects from the Mayo Clinic Rochester dataset met the criteria necessary for a reclassification.

In the REVISED-Briefing Dossier, the methodology for reclassification is described in section, Methods 4.3.8.2. Also, in this section, the flow charts of subjects who met the reclassification strategy are described for the validation datasets and summarized below in Figures 1-3. Specifically, regarding Helsinki University Hospital, a typographical error was found in the flowchart in which there is only one reclassified subject (not three) that meets the CoU. The old typographical error has been highlighted in red and struck out and the new updated value is in black in Figure 3 below. Despite this typographical error in the flowchart, no analyses were impacted. Alignment of the graft loss definition across qualification datasets was essential for conducting analyses to support this qualification submission.

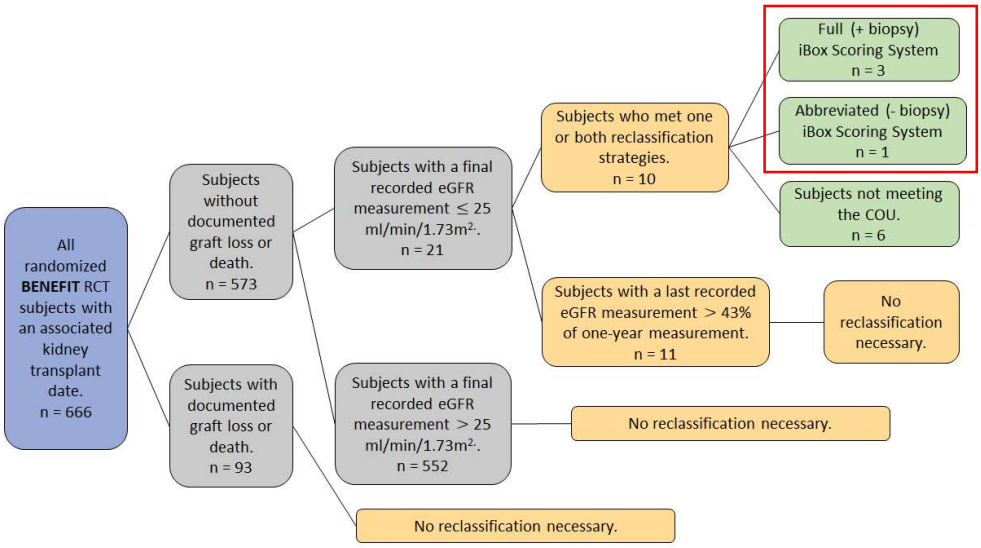

Figure 1. Flow chart describing reclassification of graft loss in the BENEFIT RCT

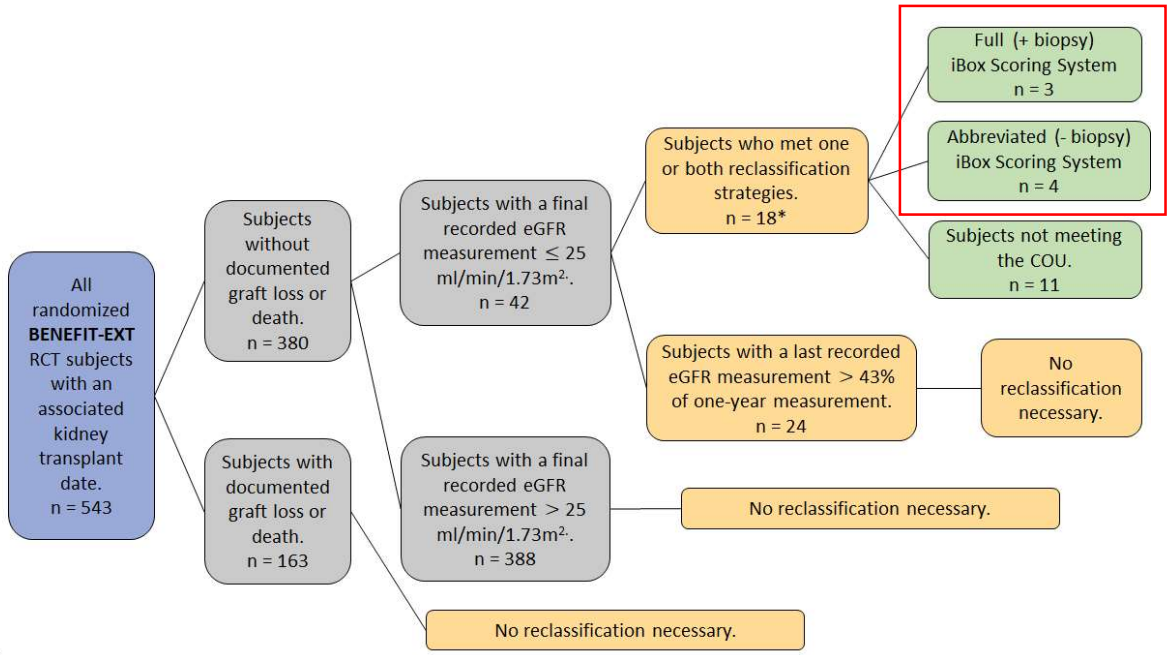

Figure 2. Flow chart describing reclassification of graft loss in the BENEFIT-EXT RCT

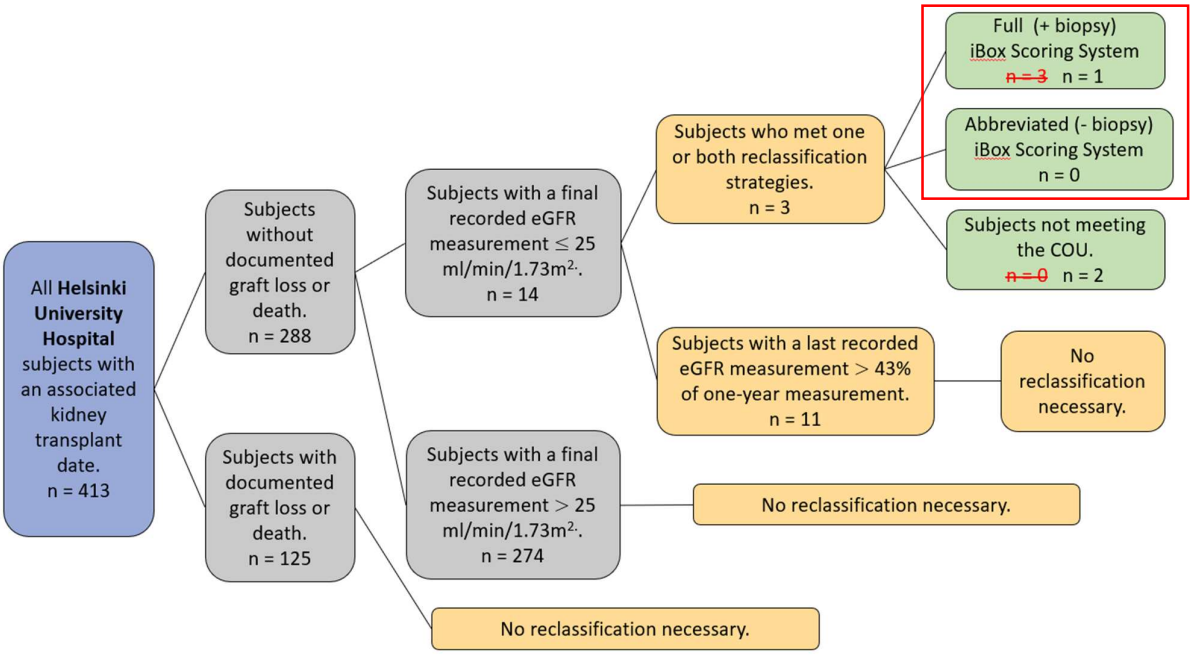

Figure 3. Flow chart describing reclassification of graft loss in Helsinki University Hospital

8. **CHMP question:** Please discuss the quality and comparability of the iBox variables, specifically:

   a. if eGFR was derived using the same equation in the derivation and validation datasets,

   b. the imputed UPCR value in the BENFIT, BENEFIT-EXT and Helsinki populations.

Applicant Response:

a. eGFR was derived using the same equation across the qualification derivation and validation datasets (4-variable Modification of Diet in Renal Disease (MDRD)-186 Study equation) (Levey et al., 2006). However, the TTC tested the performance of the iBox Scoring System using MDRD-175 Study equation and the Chronic Kidney Disease-Epidemiology Collaboration (CKD-EPI) 2021 equation (both shown in Table 2 of the Analytical Considerations document) and found that the results presented in section 6 of the submitted Briefing Dossier are nearly identical so long as the same eGFR equation is used in both the training and test datasets.

b. In clinical practice and clinical trials, dipstick tests for urinary protein are also used. They are inexpensive, easy to use, highly specific, and capable of providing rough estimates of the severity of proteinuria. After the global outreach of potential datasets to support this qualification submission, it was evident that the number of datasets to support the iBox Scoring System was limited. Therefore, it was decided to proceed with the datasets that can support this qualification submission and investigate all avenues to support the proteinuria conversions. This was not anticipated and therefore was not predefined.

   An extensive literature search of conversions from dipstick proteinuria to urine protein-to-creatinine ratio (UPCR) was conducted. Unfortunately, there were no conversions available to use for this qualification submission. The German cohort from Charité – Universitätsmedizin Berlin, comprising 1,387 subjects with 6,169 dipstick and UPCR values, was used to develop an algorithm for converting the dipstick proteinuria categorical results to continuous UPCR values. As shown in the figure below, the median UPCR values per dipstick category were used for conversion since they provided a better representation of the central location of the data points.

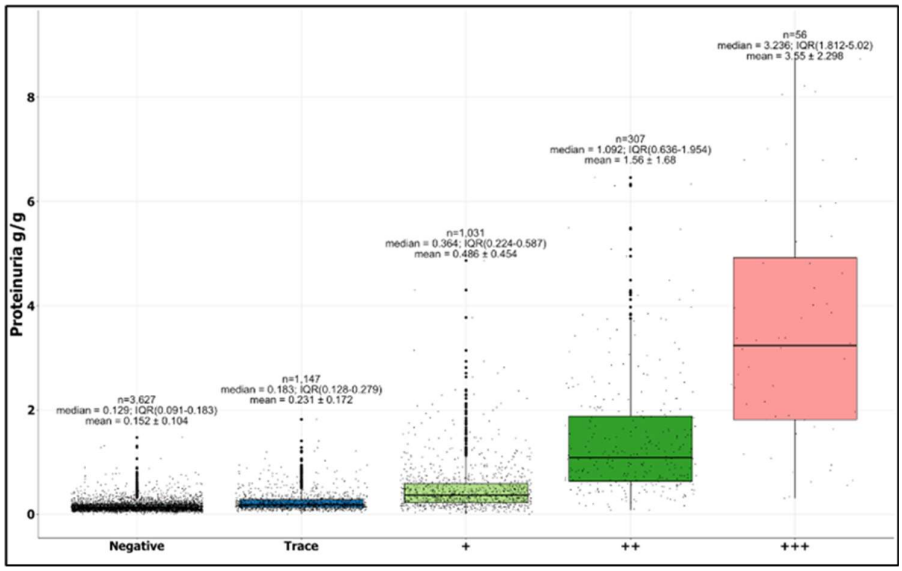

Figure 4. Association of dipstick proteinuria with UPCR in the German cohort

The datasets used for external validation that include dipstick proteinuria values were Helsinki University Hospital, and the two RCTs, BENEFIT and BENEFIT-EXT. Trace dipstick result was present in the 1,387 subject German cohort and the BENEFIT and BENEFIT-EXT RCTs. However, trace dipstick result was not present in the Helsinki University Hospital dataset. This algorithm for converting dipstick proteinuria to UPCR was applied in the Helsinki University Hospital cohort without consideration for trace proteinuria, consistent with the dipstick proteinuria assay capabilities. For consistency between datasets, the value of zero from Helsinki University Hospital has been equated to "negative" from BENEFIT and BENEFIT-EXT RCTs. The dipstick-imputed median UPCR values in Table 7 below were used to calculate an iBox score for these qualification validation datasets.

Subjects in the qualification validation datasets were assigned a UPCR value based on Table 7. A summary of the distribution of dipstick proteinuria data across external qualification validation cohorts for the full and abbreviated iBox Scoring System models is in Tables 8-9, respectively.

Table 7. Dipstick proteinuria to UPCR proteinuria association

Dipstick result	Log transformed UPCR value (g/g)
Negative	0.129 IQR (0.091-0.183)
Trace	0.183 IQR (0.128-0.279)
+	0.364 IQR (0.224-0.587)
++	1.092 IQR (0.636-1.954)
+++	3.236 IQR (1.812-5.02)

Table 8. Distribution of dipstick proteinuria data across qualification validation datasets for full iBox Scoring System

	BENEFIT RCT	BENEFIT-EXT RCT	Helsinki University Hospital
	One year ± 28 days		
Negative	306 (73.56%)	160 (61.54%)	286 (83.14%)
Trace	55 (13.22%)	41 (15.77%)	Not available
+	36 (8.65%)	42 (16.15%)	48 (13.95%)
++	16 (3.85%)	11 (4.23%)	6 (1.74%)
+++	3 (0.72%)	6 (2.31%)	4 (1.16%)
Total	416	260	344

Table 9. Distribution of dipstick proteinuria data across qualification validation datasets for abbreviated iBox Scoring System

	BENEFIT RCT	BENEFIT-EXT RCT	Helsinki University Hospital
	One year ± 28 days		
Negative	374 (72.62%)	215 (60.22%)	286 (83.14%)
Trace	70 (13.59%)	56 (15.69%)	Not available
+	46 (8.93%)	59 (16.53%)	48 (13.95%)
++	20 (3.88%)	18 (5.04%)	6 (1.74%)
+++	5 (0.97%)	9 (2.52%)	4 (1.16%)

Total	515	357	344

9. **CHMP question:** Please discuss the biological plausibility of the proposed binary DSA variable and threshold in predicting graft failure.

Applicant Response: The threshold of DSA at 1400 mean fluorescence intensity (MFI) was determined after a review of the literature as described in section 1.4 in the document titled "REVISED-Clinical Considerations". The optimal MFI cut-off for classification rates for Class I and Class II kits was 1400 (Reed et al., 2013a). MFI cut-off ranges from 1000-1500 maximized agreement (ranging from 86-93% [positive agreement] classifications) between the two kit manufacturers using Receiver-Operating Characteristic (ROC) analyses. These MFI cut-offs were subsequently adopted by the Sensitization in Transplantation: Assessment of Risk (STAR) group, a collaborative effort between the American Society for Histocompatibility and Immunogenetics (ASHI) and the American Society of Transplantation (AST) (Tambur et al., 2018). Additionally, the 2019 Expert Consensus from the Transplantation Society Working Group indicated a positive cut-off of 1000-1500 in their recommendations for treatment of antibody-mediated rejection after transplantation (Schinstock et al., 2020). Additional support for an MFI cut-off at 1400 was provided by Senev et al. in multivariate Cox proportional hazards models for graft survival on biopsies showing antibody-mediated rejection (Senev et al., 2019).

Based on the data, a cut-off of 1400 was used in the datasets for the validation and qualification of the iBox Scoring System and the all-cause endpoint (ACE) score because this value 1) optimizes classification rates for both Class I and Class II kits, 2) falls within the recommended cut-off ranges from STAR and the Transplantation Society Working group, and 3) has demonstrated discrimination regarding predicting graft survival.

One Lambda LABScreen™ SAB assay received Conformité Européenne (CE) marking by the European Economic Area (EEA) (CE_{0197}) and 510(k)-clearance by the US Food and Drug Administration (FDA) Center for Devices and Radiological Health (CDRH) (BK030069) as a qualitative test, further described in section 1.3 in the document titled "REVISED-Analytical Considerations." Changing the cut-off from four categories, as described in Loupy et al., 2019, to a binary presence or absence maintains consistency with the intended use of the assay.

The survival plot (below) shows the death-censored survival probability with the proposed binary DSA variable. This plot shows a clear distinction between groups for their risk of death-censored graft loss.

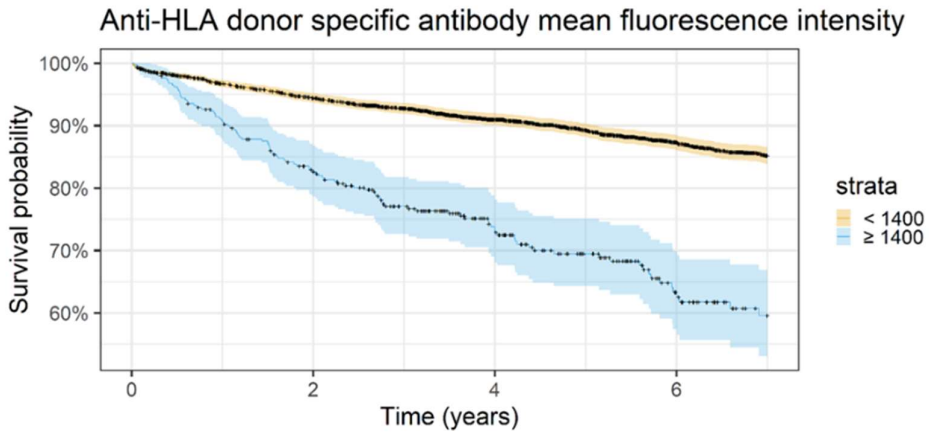

Figure 5. Survival plot based on binary DSA threshold of 1400 MFI

10. **CHMP question:** Please provide information on predefined analysis plans for the re-development of the iBox system by the Applicant and outline future plans if available.

Applicant Response: Currently, there are no plans for re-development of the iBox Scoring System. However, as stated in 2a, the TTC will perform a sensitivity analysis to determine if the coefficients in the abbreviated iBox Scoring System would substantially change if the iBox Scoring System were retrained from the validation datasets rather than the derivation. At this time, we have curated and analysed all available data for the qualification of the iBox Scoring System. Future analyses will be based on incorporating the iBox Scoring System in sponsors' studies post qualification.

11. **CHMP question:** Please discuss if variables included in the score were documented as stable/unstable before deriving an iBox Score and if iBox scores using serial measures of the individual variables (i.e. serial eGFR, - proteinuria, - biopsy results and/or PRA) have been assessed. Please discuss feasibility of including more than one measurement timepoint.

    Applicant Response: The variables included in the iBox Scoring System were not documented as stable/unstable. Due to the dynamic nature of renal function and events in the first-year post-transplant, the definitions of "stable" versus "unstable" are not established for renal transplant recipients. However, fixing the assessment time point one-year post-transplant substantially increases the likelihood of "stable" allograft function assessments since this is a routine evaluation timepoint of iBox Scoring System variables. Moreover, biopsies are likely to be protocol/surveillance instead of indication/for-cause biopsies at the time of renal dysfunction. Serial measurements were not assessed, and this is not feasible in the derivation dataset since the eGFR, proteinuria, and DSA are only available at the time of biopsy and not serially through a patient's post-transplant course.

12. **CHMP question:** Please elaborate if important iBox model variables may have been omitted, explaining the lack of treatment effect by the iBox-Scoring system as determined by the Trial-level surrogacy results.

    Applicant Response: The variables that are modifiable by IST therapy have been included in the iBox Scoring System. Other baseline variables, e.g., recipient age, are prognostic but not modifiable. The formal analysis of trial-level surrogacy was limited not by the variable inclusion but by the number and size of the RCTs available. The TTC is reassured by examining the treatment effect in the BENEFIT and BENEFIT-EXT RCTs as described in section 6.6.3.1.3 in the document titled "REVISED-BRIEFING DOSSIER" since it shows that a significant effect on iBox Scoring System at one year translates into an improvement in long-term graft survival.

References

Levey, Andrew S., Josef Coresh, Tom Greene, Lesley A. Stevens, Yaping (Lucy) Zhang, Stephen Hendriksen, John W. Kusek, Frederick Van Lente, and for the Chronic Kidney Disease Epidemiology Collaboration*. 2006. "Using Standardized Serum Creatinine Values in the Modification of Diet in Renal Disease Study Equation for Estimating Glomerular Filtration Rate." Annals of Internal Medicine 145 (4): 247. https://doi.org/10.7326/0003-4819-145-4-200608150-00004

Levin, Adeera, Rajiv Agarwal, William G. Herrington, Hiddo L. Heerspink, Johannes F.E. Mann, Shahnaz Shahinfar, Katherine R. Tuttle, et al. 2020. "International Consensus Definitions of Clinical Trial Outcomes for Kidney Failure: 2020." Kidney International 98 (4): 849–59. https://doi.org/10.1016/j.kint.2020.07.013

Loupy, A., Aubert, O., Orandi, B. J., Naesens, M., Bouatou, Y., Raynaud, M., Divard, G., Jackson, A. M., Viglietti, D., Giral, M., Kamar, N., Thaunat, O., Morelon, E., Delahousse, M., Kuypers, D., Hertig, A., Rondeau, E., Bailly, E., Eskandary, F., … Lefaucheur, C. (2019). Prediction system for risk of allograft loss in patients receiving kidney transplants: International derivation and validation study. *BMJ*, *366*. https://doi.org/10.1136/bmj.l4923

Reed, E. F., Rao, P., Zhang, Z., Gebel, H., Bray, R. A., Guleria, I., Lunz, J., Mohanakumar, T., Nickerson, P., Tambur, A. R., Zeevi, A., Heeger, P. S., & Gjertson, D. (2013a). Comprehensive assessment and standardization of solid

phase multiplex-bead arrays for the detection of antibodies to HLA. American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons, 13(7), 1859–1870. https://doi.org/10.1111/ajt.12287

Schinstock, C. A., Mannon, R. B., Budde, K., Chong, A. S., Haas, M., Knechtle, S., Lefaucheur, C., Montgomery, R. A., Nickerson, P., Tullius, S. G., Ahn, C., Askar, M., Crespo, M., Chadban, S. J., Feng, S., Jordan, S. C., Man, K., Mengel, M., Morris, R. E., … O'Connell, P. J. (2020). Recommended Treatment for Antibody-mediated Rejection After Kidney Transplantation: The 2019 Expert Consensus From the Transplantion Society Working Group. *Transplantation*, *104*(5), 911–922. https://doi.org/10.1097/TP.0000000000003095

Senev, A., Lerut, E., Van Sandt, V., Coemans, M., Callemeyn, J., Sprangers, B., Kuypers, D., Emonds, M.-P., & Naesens, M. (2019). Specificity, strength, and evolution of pretransplant donor-specific HLA antibodies determine outcome after kidney transplantation. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *19*(11), 3100–3113. https://doi.org/10.1111/ajt.15414

Tambur, A. R., Campbell, P., Claas, F. H., Feng, S., Gebel, H. M., Jackson, A. M., Mannon, R. B., Reed, E. F., Tinckam, K., Askar, M., Chandraker, A., Chang, P. P., Colvin, M., Demetris, A.-J., Diamond, J. M., Dipchand, A. I., Fairchild, R. L., Ford, M. L., Friedewald, J., … Nickerson, P. (2018). Sensitization in Transplantation: Assessment of Risk (STAR) 2017 Working Group Meeting Report. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *18*(7), 1604–1614. https://doi.org/10.1111/ajt.14752