EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

CONFIDENTIAL
Doc Ref: EMADOC-360526170-1056649
Case No.: EMA/SA/0000077234
Human Medicines Division
Initial Qualification Procedure Second List of Issues
iBox Scoring System (Composite Biomarker Panel)

Summary
The proposed iBox Scoring System was according to the initially proposed Context of Use intended 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 iBox Scoring System is a composite biomarker panel used year post-transplant to predict 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. Two iBox Scoring Systems are proposed, the full iBox and the abbreviated iBox scoring without biopsy data. The endpoint is intended to be used throughout the development phases with a broad population of renal transplant patients.

The biomarker panel has been subject to assessment of the Scientific Advice Working Party (SAWP) and several issues have been addressed in a discussion meeting on 04 May 2022. Based on this meeting and further internal discussion within the Qualification team and at SAWP, it was concluded that issuing a Qualification Opinion or Letter of Support with a modified Context of Use (CoU) can be considered. However, further information for the practical application of the iBox scores e.g., more flexible time frame and the possibilities to further promote development of iBox score as a surrogate endpoint is needed before decisions on further steps can be made.

Scientific discussion
The CoU was subject to discussion at the meeting. A modified CoU for the two composite biomarker panels was proposed, suggesting more flexible use as an endpoint to predict the long-term risk of death-censored allograft loss (allograft failure) in kidney transplant recipients for use in clinical trials to support evaluation of novel IST applications. The target population, i.e., adult de novo kidney only transplant recipients from a living or deceased donor, remains unchanged.

SAWP is of the opinion that support for iBox as a surrogate endpoint and use as single primary endpoint in confirmatory clinical trials is unlikely based on the data provided by the consortium, considering that trial level surrogacy has not been shown and it is likely that CHMP will require that surrogacy is established for acceptance as primary endpoint. This would most likely pertain to situations with application for conditional marketing authorisation and full approval.

The proposal presented during the first discussion meeting by the Applicant that would allow a CoU with a more flexible time frame of iBox application instead of a fixed assessment at 1 year to predict 5-year graft loss is welcomed. Further, the considerations on application of iBox for benefit-risk assessment including also overall mortality as separate endpoint and imputation of a variant of a 'worst-case' iBox score for patients who die early are acknowledged. SAWP notes the better performance of the death-censored iBox scores compared to the additionally developed All Cause Endpoint score in terms of predicted risks and prediction of the correct number of graft loss or all cause events, respectively.

However, it is currently unclear if a model that includes' time post-transplant' as variable to generate predictions of long-term outcomes would allow the necessary flexibility if iBox scores would be used as important secondary or exploratory endpoint as discussed at the first meeting. A more flexible CoU may require a longer study duration to

establish the benefits and the safety profile of a candidate IST in case of safety issues which could require a study duration of e.g., 2 years. On the other hand, use of the biomarker panel in phase 2 studies may require an earlier assessment time point. The Applicant should outline how the inclusion of a variable' time post-transplant' for prediction could be achieved, which impact on performance of the model is expected and if the data sets allow support of the 'time post-transplant' variable for an iBox score if the study duration would be considerably longer than 1 year. It should be commented which validation steps to assess impact of a more flexible time frame for application of a model with 'time post-transplant' as variable are deemed necessary. Distribution of assessment time points for the variables included in the iBox scores in the dataset should be presented.

SAWP acknowledges that even though full trial level surrogacy was not established, overall, the predictive performance of the iBox biomarker panels as a risk predictor is good. Results demonstrate that the proposed death-censored iBox score models are likely suitable for individual predictions of graft loss events with good performance based on c-statistics together with the ability to predict numbers of graft loss events with reasonable margins of error. The strengths of the model development process and validation procedure are noted. The quality of the whole approach and results presented would likely allow support of the iBox as full and abbreviated score with a modified context of use. The Applicant is asked to define an alternative context of use that does not focus on use as primary endpoint. A discussion on a new context of use should consider utility and limitations of a more flexible iBox score.

Provided that a suitable Context of Use could be defined, the Applicant may want to provide a view on how CHMP could best support further development of iBox. This should include considerations on which type of statement may best foster use of iBox in future trials to generate additional data to allow establishing trial level surrogacy. It should be noted that a qualification opinion statement with a suitable context of use would likely include a detailed assessment of the briefing documents, while a letter of support would include limited public statements on assessment and the assessment of questions from the Applicant would be provided separately as a qualification advice letter.

List of issues to be addressed 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 during the discussion meeting
Issues on Clinical development

1. Please propose a new Context of Use statement considering the discussion during the first meeting. Please discuss utility of iBox scores as secondary endpoint.

   C-Path response:
   C-Path agrees that the context-of-use (CoU) statement could be modified to address the following:

   a) Inclusion of a 'time post-transplant' variable in the iBox calculation to allow use in:
      - Phase 2/proof of concept (POC) studies with iBox assessments prior to the first year of transplant
      - Phase 3 studies where a study duration could be up to 2 years post-transplant

   b) Use of the iBox Scoring System as a co-primary or secondary endpoint, but not as a sole primary endpoint

   c) Recognizing the iBox Scoring System has not reached EMA qualification requirements as a "surrogate", but is prognostic of long-term death-censored allograft survival in kidney transplantation

   C-Path recognizes that use of the iBox Scoring System as a sole primary endpoint is not acceptable at this time based on the currently available data necessary to support full trial-level surrogacy (TLS) requirements for a validated surrogate endpoint with EMA. C-Path proposes that the iBox Scoring System is appropriate for use as a co-primary endpoint when used in conjunction with the currently accepted efficacy failure endpoint. This would ensure that the regulatory standard for approval based on the efficacy failure endpoint is not compromised, while also allowing the demonstration of potential superiority of a new IST based on the iBox Scoring System.

   Since the iBox Scoring System is under parallel review for qualification by FDA as a reasonable likely surrogate endpoint (RLSE), qualification as a co-primary endpoint by EMA would assist in harmonizing endpoints when conducting multinational trials. Acknowledging that the evidentiary standard for qualification of a RLSE by FDA is different from EMA standards for surrogacy, if the FDA qualifies the iBox Scoring System as a RLSE, a registration trial could be conducted in both the US and the EU using the iBox Scoring System as a co-primary endpoint. The conduct of multinational registration trials in the US and the EU has been the standard for registration of novel immunosuppressive therapies (ISTs) and will continue to be the optimal approach to efficient global drug development. Additionally, C-Path supports the option for sponsors to use the iBox Scoring System as a secondary endpoint.

   The proposed new CoU is as follows:

   General measurement:
   The iBox Scoring System (Composite Biomarker Panel) is a ~~surrogate endpoint~~ co-primary or secondary endpoint prognostic for ~~long term~~ death-censored allograft loss (allograft failure) in kidney transplant recipients to be used in clinical trials to support the evaluation of novel immunosuppressive therapy applications.

   Timing of iBox assessments:
   The iBox Scoring System is an acceptable co-primary endpoint (when used in conjunction with the currently accepted efficacy failure endpoint) or secondary endpoint measured between 6- and 24-months post-kidney transplantation in pivotal or exploratory drug therapeutic studies for regulatory purposes.

   Future use as a surrogate endpoint:

Although the use of the iBox Scoring System as a sole primary endpoint for pivotal trials in this setting is promising, more robust data gained with additional patients and studies could strength the surrogacy and predictive performance of surrogacy the iBox Scoring System.

2. Please discuss the impact of inclusion of a variable' time post-transplant' for prediction, considering
   a) how this variable was already taken into account during development,
   b) if it is expected to have impact on performance of the model,
   c) if new/which validation steps to assess impact of a more flexible time frame are deemed necessary,
   d) if the data sets allow support of the 'time post-transplant' variable longer than 1 year,
   e) the assessment time points. For this, please provide the data distribution of assessment time points for the variables included in the iBox scores in the datasets.

   C-Path response:

   a) The original development of the iBox Scoring System by Loupy et al. 2019[1] in the Paris Transplant Group included time post-transplant to account for varying iBox assessments of an individual patient and to assist in patient care and prognosis estimation. The derivation dataset included in this qualification submission represents all 4,000 subjects for the abbreviated iBox Scoring System described in the Loupy et al., 2019 publication. Additional analyses were conducted on this derivation dataset in which the time of evaluation was fixed at one-year post-transplant to assess the performance as a trial endpoint for a typical Phase 3 study. C-Path supports the inclusion of the time post-transplant variable from 6-24 months post-transplant, recognizing the expanded potential use in Phase 2/POC and Phase 3 trials of longer duration. Tables 3-6 below are the calibration and discrimination analyses (external validation) on the qualification validation datasets with the varying times post-transplant. C-Path envisages that Phase 2/POC studies may include an endpoint at six-months post-transplant, whereas Phase 3 trials would be of 1-2 years duration to assess the co-primary endpoints.

   b) and c) The frequency chart in Figure 1 below shows the distribution of assessment time points for donor-specific antibody (DSA) measurements up to 2 years post-kidney transplant. DSA was selected for illustration since it is collected less frequently than eGFR and/or proteinuria and therefore will be the key limiting factor for the availability of iBox measurements at various time points post-transplant across the qualification validation datasets. Helsinki is excluded in this data exploration as the iBox Scoring System was only assessed at 1 year since there is no longitudinal proteinuria or DSA data.

   The number of transplant recipients with iBox assessments at varying times post-transplant in the external validation datasets are shown in Tables 1 and 2 below. In addition, the five-year post-transplant discrimination and calibration were assessed and are shown in Tables 3-6 below. The red colour in Tables 3 and 5 highlights c-statistics (c-stat) below 0.7. The green shading in Tables 4 and 6 indicates that the observed events are not significantly different than model predictions. Non-applicable (NA) reflects no assessments in Tables 1-2 or fewer than two events in Tables 3-6.

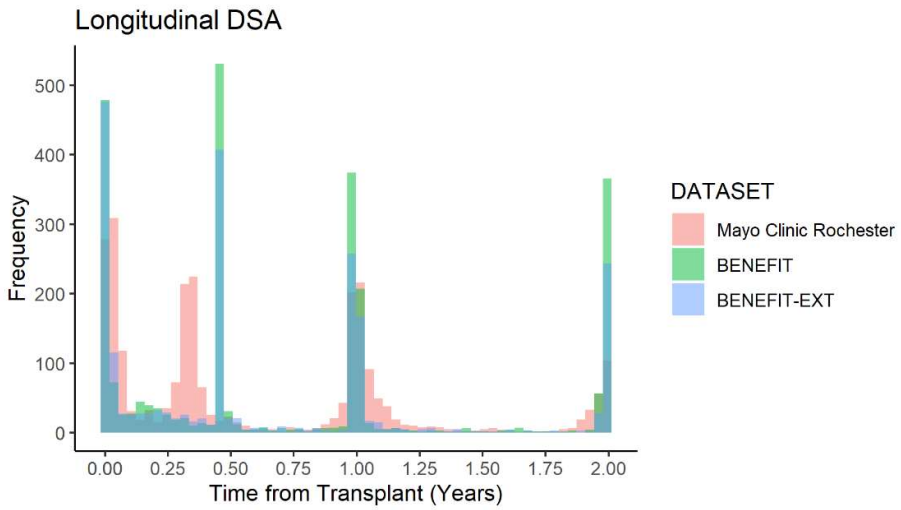

Figure 1. Frequency of DSA measurements across the qualification validation datasets

*Table 1. Number of subjects with **full** iBox Scoring System evaluations at varying time points*

Dataset	Time post-transplant			
	4-months (n)	6-months (n)	1-year (n)	2-years (n)
Helsinki University Hospital*	NA	NA	344	NA
Mayo Clinic Rochester	224	NA	483	NA
BENEFIT RCT	NA	30	416	12
BENEFIT-EXT RCT	NA	31	260	5
Total Subjects	**224**	**61**	**1,503**	**17**

* No longitudinal proteinuria or DSA data
NA reflects no assessments

*Table 2. Number of subjects with **abbreviated** iBox Scoring System evaluations at varying time points*

Dataset	Time post-transplant			
	4-months (n)	6-months (n)	1-year (n)	2-years (n)
Helsinki University Hospital*	NA	NA	344	NA
Mayo Clinic Rochester	231	NA	497	NA
BENEFIT RCT	NA	527	515	476
BENEFIT-EXT RCT	NA	383	357	328
Total Subjects	**231**	**910**	**1,713**	**804**

* No longitudinal proteinuria or DSA data
NA reflects no assessments

*Table 3. Five-year post-transplant c-statistics values for the **full** iBox Scoring System at varying time points*

	Time post-transplant															
Dataset	4-months				6-months				1-year				2-years			
	n	# Graft losses	c-stat	SE	n	# Graft losses	c-stat	SE	n	# Graft losses	c-stat	SE	n	# Graft losses	c-stat	SE

Observational	224	14	0.66	0.08	NA	NA	NA	NA	827	39	0.84	0.04	NA	NA	NA	NA
Helsinki University Hospital	NA	NA	NA	NA	NA	NA	NA	NA	344	21	0.78	0.06	NA	NA	NA	NA
Mayo Clinic Rochester	224	14	0.66	0.08	NA	NA	NA	NA	483	18	0.93	0.03	NA	NA	NA	NA
RCTs	NA	NA	NA	NA	61	8	0.78	0.06	676	24	0.76	0.06	NA	NA	NA	NA
BENEFIT	NA	NA	NA	NA	30	3	0.84	0.07	416	12	0.71	0.09	NA	NA	NA	NA
BENEFIT-EXT	NA	NA	NA	NA	31	5	0.71	0.11	260	12	0.81	0.07	NA	NA	NA	NA

- The red text colour highlight c-statistics < 0.7
- NA reflects fewer than two events

*Table 4. Poisson calibration for the **full** iBox Scoring System at varying time points*

Dataset	**Time post-transplant**															
	4-months				6-months				1-year				2-years			
	n	**Obs**	**Pred**	**p value**	**n**	**Obs**	**Pred**	**p value**	**n**	**Obs**	**Pred**	**p value**	**n**	**Obs**	**Pred**	**p value**
Observational	224	14	12.67	0.71	NA	NA	NA	NA	827	39	38.74	0.97	NA	NA	NA	NA
Helsinki University Hospital	NA	NA	NA	NA	NA	NA	NA	NA	344	21	14.40	0.08	NA	NA	NA	NA
Mayo Clinic Rochester	224	14	12.67	0.71	NA	NA	NA	NA	483	18	24.34	0.20	NA	NA	NA	NA
RCTs	NA	NA	NA	NA	61	8	5.11	0.20	676	24	27.90	0.46	NA	NA	NA	NA
BENEFIT	NA	NA	NA	NA	30	3	2.03	0.50	416	12	12.93	0.80	NA	NA	NA	NA
BENEFIT-EXT	NA	NA	NA	NA	31	5	3.08	0.28	260	12	14.97	0.44	NA	NA	NA	NA

- The green shading indicates that the observed events are not significantly different than model predictions
- NA reflects fewer than two events

*Table 5. Five-year post-transplant c-statistics values for the **abbreviated** iBox Scoring System at varying time points*

Dataset	**Time post-transplant**															
	4-months				6-months				1-year				2-years			
	n	**# Graft losses**	**c-stat**	**SE**	**n**	**# Graft losses**	**c-stat**	**SE**	**n**	**# Graft losses**	**c-stat**	**SE**	**n**	**# Graft losses**	**c-stat**	**SE**
Observational	231	14	0.64	0.08	NA	NA	NA	NA	841	41	0.80	0.04	NA	NA	NA	NA
Helsinki University Hospital	NA	NA	NA	NA	NA	NA	NA	NA	344	21	0.77	0.06	NA	NA	NA	NA
Mayo Clinic Rochester	231	14	0.64	0.08	NA	NA	NA	NA	497	20	0.84	0.05	NA	NA	NA	NA

RCTs	NA	NA	NA	NA	910	45	0.73	0.05	872	38	0.75	0.05	804	24	0.75	0.06
BENEFIT	NA	NA	NA	NA	527	19	0.68	0.08	515	15	0.70	0.08	476	11	0.73	0.10
BENEFIT-EXT	NA	NA	NA	NA	383	26	0.72	0.06	357	23	0.78	0.06	328	13	0.76	0.07

- The red text colour highlight c-statistics < 0.7
- NA reflects fewer than two events

*Table 6. Poisson calibration for the **abbreviated** iBox Scoring System at varying time points*

Dataset	**Time post-transplant**															
	4-months				6-months				1-year				2-years			
	n	**Obs**	**Pred**	**p value**	**n**	**Obs**	**Pred**	**p value**	**n**	**Obs**	**Pred**	**p value**	**n**	**Obs**	**Pred**	**p value**
Observational	231	14	13.90	0.98	NA	NA	NA	NA	841	41	40.61	0.95	NA	NA	NA	NA
Helsinki University Hospital	NA	NA	NA	NA	NA	NA	NA	NA	344	21	16.19	0.23	NA	NA	NA	NA
Mayo Clinic Rochester	231	14	13.90	0.98	NA	NA	NA	NA	497	20	24.41	0.37	NA	NA	NA	NA
RCTs	NA	NA	NA	NA	910	45	51.30	0.38	872	38	39.92	0.76	804	24	32.38	0.14
BENEFIT	NA	NA	NA	NA	527	19	22.01	0.52	515	15	16.95	0.64	476	11	13.61	0.48
BENEFIT-EXT	NA	NA	NA	NA	383	26	29.29	0.54	357	23	22.97	1.00	328	13	18.77	0.19

- The green shading indicates that the observed events are not significantly different than model predictions
- NA reflects fewer than two events

b) As shown in Tables 1 and 2 above, there are 17 and 804 subjects with full and abbreviated iBox assessments at two-years post-transplant, respectively. The discrimination and calibration analyses support the inclusion of time post-transplant in the iBox Scoring System from 6 months up to 2 years post-transplant.

c) Tables 1 and 2 show the data distribution for the iBox assessments in the qualification validation datasets from 4 months up to 2 years post-transplant.

3. Please discuss if there are plans to further develop the iBox Scoring System for Qualification as a surrogate endpoint and if/how either publication of a Qualification Opinion with an amended Context of Use, or a Letter of Support could facilitate further evidence generation.

C-Path response:
The next step to further develop the iBox Scoring System for Qualification as a surrogate is the inclusion of this endpoint in future randomized controlled trials (RCTs). In order for the iBox to be included in such trials, C-Path believes that a Qualification Opinion as a co-primary and secondary endpoint is essential. When used in this way, sponsors and investigators will be able to assess and promote the potential advantages and superiority of novel ISTs when measured using the iBox Scoring System. Further, the iBox Scoring will be included in the summary of product characteristics (SmPC), claims, and other product labelling. Without this regulatory endorsement, it is unlikely that sponsors will include the iBox Scoring System in trials and commit to the longer-term follow-up needed for Qualification as a surrogate. Additionally, the harmonization of multinational trials will be facilitated by this Qualification. The ability to conduct trials in both the US and EU with similar endpoints is critical to advancing the field and is consistent with previous pivotal trials in kidney transplantation.

References

1. Loupy A, Aubert O, Orandi BJ, Naesens M, Bouatou Y, Raynaud M, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. BMJ. 2019 Sep 17;l4923.