Early allograft failure in liver transplant – Sample size calculation

**Fernando Palluzzi, PhD**

**1. Prior knowledge**

1.1. EAF incidence

* Incidence of Early Allograft Failure (EAF) according to the L-GrAFT study (2018): 11.1% 1.
* Incidence of EAF according to the EASE score study: 6.7% 2
* Incidence of EAF according to the L-GrAFT validation study (2020) in UCLA: 7% 4
* Incidence of EAF according to the L-GrAFT validation study (2020) in other US validation CENTERS: 11% 4
* Incidence of EAF according to the L-GrAFT validation study (2020) in European COPE cohort: 4% 4

Notably, the COPE cohort consists of 222 grafts.

Research hypothesis stratified by donor type:

* + 3% EAF in liver transplantation with living donors;
  + 7% EAF in liver transplantation with standard deceased donors;
  + 10% EAF in liver transplantation with high-risk deceased donors.

1.2. EASE and L-GrAFT score performance as AUC (95% confidence interval)

* EASE(10) AUC = 0.87 (0.83, 0.91) 2,3
* L-GrAFT10 AUC = 0.72 (0.65, 0.78) 2,3
* L-GrAFT7 AUC = 0.78 (0.75, 0.82) 4

**2. EASE score algorithm validation**

Current target sample size (n):

* 4000-5000 subjects overall;
* 1600-2000 for the prospective study;
* 2400-3000 for the retrospective study.

**Simulation 2.1: EASE(10) vs. L-GrAFT10**

An overall population size *n*0 = 5000 subjects (i.e., the highest target sample size for this study), with an EAF incidence1 of *p*0 = 0.111, is set for the first simulation. During the simulation, *n*0 subjects were sampled from a binomial distribution to build a random dichotomous outcome vector *y* = {1: EAF, 0: non-EAF}.

Two simulated predictors, namely *x*1 and *x*2, are then generated for L-GrAFT10 and EASE score, respectively. Predictor values are simulated by changing the *y* values (0-to-1 or 1-to-0) according to the probability *p* = P(*x* = *k* | *y* = *k*), where 1 – *p* is the probability for either the EASE or L-GrAFT10 score to yield a false positive or false negative (i.e., the prediction error). The value of *p* is such that the estimated AUC for either the EASE or the L-GrAFT10 score is equal to the observed AUC value. ROC curves for both EASE and L-GrAFT10 scores are generated using the *roc* function from the R package pROC 5.

AUCs and sample sizes for cases (EAF) and controls (non-EAF) are then estimated using the *power.roc.test* function from pROC 5. Taking the AUC1 = 0.72 of the L-GrAFT10 score as a reference, the sample sizes were estimated considering an AUC2 = 0.82 for the EASE score, under the null hypothesis of no difference between AUC1 and AUC2.

**CONCLUSION**

At a significance level = 0.05 and power = 0.8, the estimated sample size is **about 350 subjects** (36 EAF + 310 non-EAF subjects).

**Simulation 2.2: EASE vs. L-GrAFT7**

This simulation was conducted with the same input arguments as the previous one, but considering a baseline AUC1 = 0.78, corresponding to the L-GrAFT7 score AUC 4. This modification is introduced to consider the possibility of achieving a higher AUC at day 7 of follow-up, after the transplantation, rather than day 10 (the current lower boundary). This has also the desired effect of reducing the difference AUC2 – AUC1, leading to a more conservative (i.e., larger) sample size estimation.

**CONCLUSION**

Under the null hypothesis of no difference between AUC1 and AUC2, at a significance level = 0.05 and power = 0.8, the estimated sample size is **about 790 subjects** (83 EAF + 705 non-EAF subjects).

**3. Exploring different incidences of EAF**

Current target sample size (n):

* Stratum A: 800 living donors grafts (3% EAF);
* Stratum B: 1000 subjects with standard deceased donors (7% EAF);
* Stratum C: 200 subjects with high-risk deceased donors [DCD or MP grafts] (10% EAF).

The aim of this estimation is to provide the minimum sample size to achieve the baseline AUC, considering a different EAF incidence for each stratum. To reach a conservative sample size estimation, the baseline (i.e., lowest) AUC = 0.72 of the L-GrAFT10 score is considered, as a minimum performance requirement. The function *power.roc.test* from the R package pROC 5 is used also in this case.

**CONCLUSION**

Considering an AUC = 0.72, a significance level = 0.05, and a power = 0.8, the following sample sizes were estimated (with *k* = expected balance between non-EAF and EAF subjects):

* **442 grafts for stratum A**

**(living donors grafts, supposed incidence 3% EAF)**,

with *k* = (1 – *p*A)/*p*A

* **200 grafts for stratum B**

**(standard deceased donors, supposed incidence 7% EAF)**,

with *k* = (1 – *p*B)/*p*B

* **142 grafts for stratum C**

**(high risk deceased donors [DCD or MP grafts]**),

with *k* = (1 – *p*C)/*p*C

where *p*A = 0.03, *p*B = 0.07, and *p*C = 0.10, correspond to the EAF proportion for stratum A, B, and C, respectively.

**4. EASE score cut-off validation for class 5 subjects**

Let us first consider a *p*5 = 0.0336 proportion of EAF cases belonging to EASE class 5 (highest L-GrAFT or EASE score). The aim is to estimate the minimum required sample size to observe a proportion *p*5 of EAF cases in class 5 and validate/improve the cut-off between class 5 and the other classes (1 to 4).

**CONCLUSION**

To achieve a conservative sample size estimate, a minimum AUC = 0.7 is required for each class, at a significance level = 0.05, a power = 0.8, and a *k* = (1 – *p*5)/*p*5. Using the *power.roc.test* function from pROC 5, the minimum estimated sample size is 480.6 grafts (i.e., **more than 480 subjects**).

**References**

1. Agopian VG, Harlander-Locke MP, Markovic D, *et al*. Evaluation of Early Allograft Function Using the Liver Graft Assessment Following Transplantation Risk Score Model. JAMA Surg. 2018 May 1;153(5):436-444. doi: 10.1001/jamasurg.2017.5040.
2. Avolio AW, Lai Q, Cillo U, Romagnoli R, De Simone P. L-GrAFT and EASE scores in liver transplantation: Need for reciprocal external validation and comparison with other scores. J Hepatol. 2020 Dec 17; S0168-8278(20)33848-4. doi: 10.1016/j.jhep.2020.12.009.
3. Avolio AW, Franco A, Schlegel A, *et al*. Development and Validation of a Comprehensive Model to Estimate Early Allograft Failure Among Patients Requiring Early Liver Retransplant. JAMA Surg. 2020; 155(12): e204095. doi:10.1001/jamasurg.2020.4095.
4. Agopian VG, Markovic D, Klintmalm GB, *et al*. Multicenter validation of the liver graft assessment following transplantation (L-GrAFT) score for assessment of early allograft dysfunction. Journal of Hepatology 2021 vol. 74 j881–892. doi: 10.1016/j.jhep.2020.09.015.
5. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Mar 17; 12:77. doi: 10.1186/1471-2105-12-77.