4. Complete Case Only analysis

Because of randomly missing data in many variables, the number of patients eligible for complete-case analysis is substantially reduced:

Patients	N	No. events	No. deaths	No. graft losses
Complete cases	1108	236	104	132
Full analysis	2041	588	303	285

The study population of 2041 patients can be seen as "the population" for which the HR estimates for patient survival, actual graft survival and functional graft survival are "population parameters". By definition, a 95% confidence interval for the HR computed from 1108 randomly selected patients covers its population value with 95% probability. Non-random selection would be indicated if the confidence interval excludes the inferred population value.

Since statin use is available for all patients, a full data analysis can be performed to compute the crude hazard ratio of statin use.

Comparison of crude hazard ratio estimates and results from MSM analysis, using complete-case-only (CCO) analysis (1108 patients) and multiple imputation analysis (2041 patients):

Outcome	**Crude hazard ratio (95% confidence limits)**	**MSM: CCO hazard ratio (95% confidence limits)**	**MSM: multiple imputation hazard ratio (95% confidence limits)**
Patient survival	0.77 (0.59, 1.0)	0.89 (0.57, 1.41)	0.69 (0.50, 0.95)
Actual graft survival	0.77 (0.63, 0.93)	0.84 (0.63, 1.14)	0.73 (0.58, 0.92)
Functional graft survival	0.76 (0.58, 1.0)	0.79 (0.50, 1.27)	0.84 (0.61, 1.17)

Results from MSM/multiple imputation are closer to crude estimates than to their MSM/CCO counterparts. For analysis of actual and functional graft survival, differences between CCO and multiple imputation results are less dramatic. The confidence intervals of the MSM/CCO hazard ratio cover the MSM/multiple imputation estimates (which are considered population values in this comparison), suggesting random selection.

However, the reduced number of patients leads to an inflation of confidence intervals. Since statin use is completely documented, a complete-cases-only analysis seems a waste of resources in the present study and bears the danger of overfitting the models due to a prohibitively high ratio of covariates and events. Therefore, such an analysis was not

further pursued. Instead, the multiple imputation approach was subjected to an analysis of sensitivity of the randomly missing data assumption (section 5).