University of Zurich[UZH]

Master Program in Biostatistics
www.biostat.uzh.ch

STA490: Statistical Consulting

Leonhard Held

EBPI, Department of Biostatistics

What are predictors of long-term work participation in patients with cystic fibrosis undergoing lung transplantation?

Analysis for Thomas Radtke, EBPI

Supervision by Julia Braun

Xijin Chen (*xijin.chen@uzh.ch*)

Version of October 10, 2019

1 Abstract

Background:

Cystic Fibrosis is a type of dysfunction of secretory glands, which affects lungs, digestive system and sweat glands. Lung transplantation is a standard of care for advanced lung diseases. Cystic fibrosis patients may be able to return to work after lung transplantation because of prolonged survival and improved quality of life. The aim of this study is to identify the predictors for long-term work participation in cystic fibrosis patients undergoing a lung transplant.

Methods:

This retrospective study included 99 cystic fibrosis patients having received lung transplantation between January 1996 and December 2016 at University Hospital Zurich. Factors describing patients' basic characteristics (age, sex, Body Mass Index, Six-minute Walk Test Distance, education, relationship status, living status and pre-employment status) were retrieved from the patients' charts. After evaluation tests, all of those 99 cystic fibrosis patients entried in the waiting list to wait for a healthy lung. After the lung transplantation, the follow-up time was divided into five periods: within one year, one to three years, three to five years, five to ten years and more than ten years after lung transplantation. Lung recipients may go through some rejection, infection and transplantation-related complications during the follow-up time.

To evaluate the work participation after lung transplantation, we investigated both work status and average of work percentage at each time period: one year, three years, five years, ten years after lung transplantation and the end of the study.

Predictors for work status at each time point were identified by logistic regression. Predictors for work percentage at each time point were identified by linear regression. Based on the results from 'simple' regression from the measurements on each time point, generalized linear mixed models and linear mixed models were used for longitudinal data analysis to identify predictors for long-term work participation.

The criterion for model selection in this study is the Bayesian Information Criterion (BIC).

Results:

When focusing on each specific time point after lung-transplantation, the results from logistic regression show that there are two possible time-independent predictors for work status after lung transplantation: *Pre-employment* and *Education*. One to three years after lung transplantation, the odds ratio for *Pre-employment* is 11.19, with 95% confidence interval from 3.67 to 42.71. It shows very strong evidence for the effect on work status. The odds ratio for *Education* is 5.34, (95% confidence interval from 1.22 to 37.82), showing moderate evidence for its effect on work status.

The results from linear regression show that the only predictor for work percentage after lung transplantation is *Pre-employment*. The strength of evidence varies between time periods.

From the results of longitudinal data analysis, besides *Pre-employment* and *Education*, there are some time-dependent factors that may have an effect on work participation.

From the results of the generalized linear mixed-effects model, *CLAD* (Chronic Lung Allograft Dysfunction) would have a negative effect on the post-transplantation work status. The odds ratio is 0.64, 95% confidence interval from 0.15 to 2.40, showing strong evidence of the negative effect on the work status after lung transplantation.

From the result of the linear mixed-effects model, *Kidney-Dialysis* would have a negative effect on the post-transplantation work percentage. The coefficient is -18.69, 95% confidence interval from -31.80 to -5.59, with moderate evidence for the association between *Kidney-Dialysis* and long-term work percentage.

Conclusions:

We determined time-independent factor *Pre-employment* and *Education* as the factors that would have an effect on both work status and work percentage after lung transplantation. Besides time-independent factor, we decided to use *CLAD* as another predictor for the long-term work status after transplantation. *Kidney-Dialysis* is the predictor that we chose for the long-term work percentage after transplantation.

In a word, patients who were employed, and with academic education before transplantation would be more likely to return to work. Chronic Lung Allograft Dysfunction is a kind of limitation for the cystic fibrosis patients to return to work after transplantation. Patients who had kidney transplantation or dialysis after transplantation, would work less than the others without renal dysfunctions.

2 Introduction

Cystic fibrosis, lung transplantation and work participation

Cystic fibrosis (CF) is a genetic disorder passed down from generations to generations. It is caused by the presence of mutations in both copies of the gene for cystic fibrosis transmembrane conductance regulator (CFTR) protein. Those with a single working copy are carriers and otherwise mostly unaffected When the protein is not working correctly, it's unable to help move chloride, a component of salt, to the cell surface. Without chloride attracting water to the cell surface, the mucus in various organs becomes thick and sticky. In the lungs, mucus clogs the airways and traps germs, e.g. bacteria, leading to infections, inflammation, respiratory failure, and other complications. Cystic fibrosis patients frequently can not work due to cancers of the digestive tract, diabetes, heart failure, kidney problems and lung infections.

There is currently no cure for cystic fibrosis since it is inherited, but several kinds of treatments can ease symptoms and reduce complications. Those treatments can help prevent and control infections that occur in the lungs, remove and loosen mucus from the lungs, treat, prevent intestinal blockage and provide adequate nutrition as well. Currently, there are many options for treatment, e.g. medication, chest physical therapy, pulmonary rehabilitation and surgery.

Lung transplantation (LTx) has become the standard of care for patients with end-stage chronic respiratory failure, which is necessary when medical management alone can no longer maintain lung health and physical function (**?**). Lung transplantation is a surgery removing a diseased lung and replacing it with a healthy lung from a donor. Transplantation can be divided into two types, cadaveric transplantation from deceased organ donors and living transplant from part of one of the lungs form healthy adults. More than 6,400 lung transplantations have been performed since the first successful operations in the early 1980s (**?**).

Patients underwent evaluation tests as a preparation process before lung transplantation, including a variety of medical tests that provide complete information about overall health of patients. Those medical tests aim at helping the lung transplant team to identify any potential problem before the transplant surgery and avoid potential complications after the surgery. From a practical perspective, patient selection involves analysis of standard investigations designed to identify comorbidities that may increase risk, and to balance those against likely outcomes. After the evaluation tests, patients would entry in the waiting list before transplantation.

Patients' work participation is a common clinical quality indicator being used to monitor the performance of healthcare services (**?**), which helps to assess the quality of care, identify and prioritize areas for improvement (**?**), (**?**). In addition to life expectancy, this simple measure of health outcome can also reflect the reduction of disability and improvement of health-related quality of life (HRQoL). As indicated in **?** and **?**, lung transplantation can alleviate some physiological barriers to employment. Therefore, work participation can be used as an indicator for long-term performance after lung transformation (**?**).

Evidence from recent studies

There is a great number of references related to work participation after organ transplantation. In the analysis of solid organ transplantation, there is a list of factors being positively associated with returning to work after organ transplantation. Being employed before transplantation (**???**). High level educated patients, male patients would be more likely to return to work after organ transplantation from the results in **?**.

In terms of lung transplantation for lung diseases in cystic fibrosis patients, better outcomes are limited by a list of factors (**???**). As shown across studies **????**, pre-employment is the most consistent and dominant predictor for work participation. There is a strong positive association between the pre-employment and post work status. From the study **?**, lack of work history before lung transplantation was identified as a barrier to post-transplant employment in cystic fibrosis patients. The result of influence of education levels should not be ignored by **?**. In this study for lung transplantation for cystic fibrosis patients, the level of education was associated with the work participation after lung transplantation. From **?**, not only education, but also best predicted FEV1(%) showed a significant effect on the post-transplantation work status in multivariate analysis.

Possible predictors in our study

There is also a list of possible predictors in our study. Patient attributes and physical characteristics are common basic information that can be retrieved or tested. Evidence shows that clinical decision making varies according to patients (**?**), as characteristics provide important information about prior probability that a given patient will experience a given condition or problem.

The level of education can tell us about the possible post-transplantation work participation to some extent. Well-educated patients are more likely to go into skilled labor as opposed to manual work. Therefore, highly-educated patients would be more likely to return to work.

Measurements like body mass index, six-minute walk test distance (**?**), reflect physical functions of patients, which are directly related to the performance after lung transplantation. It varies greatly, depending on age, blood type, the status of the lung of patients and the availability of a donor lung that is a good match. People who are unable to wait may be considered for lung transplant from a living donor. Up to 25 to 41 % of cystic fibrosis patients have died while awaiting a healthy lung as illustrated by **?**, **?**, **?** and **?**. Therefore, too long waiting time may impede the patient's utility gain from the treatment.

A list of risks after a lung transplantation could also bring some barriers to work participation. Rejection is a normal body reaction to a foreign object or tissue. When an organ is transplanted into a person's body, their immune system sees it as a threat and attacks the organ. From the study **?**, main limitation to better long-term survival after lung transplantation is chronic lung allograft dysfunction (CLAD).

After solid organ transplantation, the risk of cancer in the digestive tract is elevated 2-4 fold compared with population controls. The high risks for cancer are associated with viral infections, including Mb, Hodgkin and non-Hodgkin lymphoma (NHL) (caused by Epstein Barr virus [EBV]), anogenital cancer (human papillomavirus), and liver cancer (hepatitis C and B virus), but also for other malignancies such as lung, kidney, skin, and thyroid cancer.

Better management of medical complications may lead to improved long-term outcomes. Improvements in surgical techniques, lung preservation, immunosuppression, and management of infections have influenced the improvements in mortality by **?**.

Renal dysfunction is one of the most common long-term complications of lung transplantation and may have an effect on the post-transplantation work participation, with an incidence of 25.5% at 1 year after transplant and 37.8 at 5 years after transplant from **?**. The development of chronic renal failure increases the risk of death by four to five fold in patients who have undergone lung transplant (**?**). The progression of chronic renal failure ultimately leads to renal replacement therapy in the form of dialysis or renal transplantation.

Pulmonary function test results (forced expiratory volume in one second [FEV1]), were expressed as the percentage of the predicted value and the best values achieved after the lung transplantation. FEV1, as a result of breathing test, is normally used to show the severity of chronic obstructive pulmonary disease. The measurement of FEV1 for a given patient is compared to reference values. The reference value is based on healthy individuals with normal lung function and it shows the values that would be expected for someone of the same sex, age and height. If the predicted values of patients are within 80% of the reference value, the results are considered normal. As indicated in **?**, the value of FEV1 count for work participation.

Therefore, our study would investigate all the possible predictors listed above to identify the predictors for long-term work participation in cystic fibrosis patients after lung transplantation.

Aim of our study

In an attempt to characterize the impact of lung transplantation on the long-term work participation for cystic fibrosis patients, the aim of our study is to identify the predictors for long-term work participation after lung transplantation.

To our acknowledgement, there is no research about the longitudinal data analysis with respect to both work status and work percentage after lung transplantation.

Our study identified the predictors not only for work status, but also for work percentage, as a compensation analysis. After the conventional analysis at each time point, our study took the effect of time-dependent factors and time into consideration and analyzed outcomes for all time periods by longitudinal analysis. Finally, identifying the predictors for long-term work participation among cystic fibrosis patients.

3 Research Questions

1. What are the predictors for work status in cystic fibrosis patients after lung transplantation?
2. What are the predictors for work percentage in cystic fibrosis patients after lung transplantation?

4 Methods

Study Design

Type of study

Retrospective, single center study in the University Hospital Zurich.

Study population

The study population is 99 cystic fibrosis patients receiving lung transplantation, from January 1996 to December 2016.

Data collection

Social-demographic factors about the basic information of patients (age, sex, education, relationship status, BMI, living status, six-minute walk distance and pre-employment status) were recorded before the transplantation. Among those factors, education was dichotomized into academic or non-academic. Relationship status was dichotomized into single or divorced versus with relationship or married/engaged. Living status was dichotomized into alone and not alone. Pre-employment status were indicated whether the patient was employed or not before the lung transplantation.

The six-minute walk test (6MWD) measured the distance a patient was able to walk over a total of six minutes on a hard, flat surface. The measurement was a performance-based measure of functional exercise capacity. Body Mass Index was calculated based on the weight and height of patients.

All of those 99 cystic fibrosis patients had finished evaluation tests. The date for waiting list and lung transplantation were recorded, which were needed for the waiting time. The bio-medical factor waiting time was transformed into weeks, unlike the dichotomy for waiting time **?**, to get more accurate result of the influence of waiting time on post-transplantation work participation.

Observations were obtained at five time points: one year, three years, five years, ten years after lung transplantation and at the end of the study. The best FEV1(%) out of all those repeated measurements for our patients was used. Besides, we got the exact time for chronic lung allograft dysfunction (CLAD), Kidney transplant, dialysis and cancer if these events occurred.

Work participation, were evaluated in two ways, work status and the average of work percentage between two separate time points of measurements after lung transplantation. Post-transplant work status was divided into not work or work after lung transplantation, post-transplant work percentage was from 0 to 100(%).

Statistical Analysis

Data preparation

We used the difference between the date of putting on the waiting list and date of lung transplantation for waiting time of each patient, the differences were transformed into weeks.

Similarly, we calculated the difference between the date of events and date of lung transplantation, and then compared the difference with the date of each time point: one year, three years, five years, ten years after lung transplantation and at the end of the study. The results of comparison were used to create binary variables for each time periods. Therefore, those time-varying variables, increased monotonously with time.

This time-varying variable takes two events into consideration: kidney transplantation and dialysis. Both of them are management for the complication of renal dysfunction, so we created one binary variable to reflect the effect of renal dysfunction on the work participation after lung transplantation.

work status and work participation, and those three time-varying variables were measured repeatedly at several time points, we transformed whole dataset into long-format data for the preparation of the longitudinal data analysis.

Since the period of waiting time vary a lot for each patient, not all of those 99 patients would go through the whole study period (20 years). Therefore, those 99 patients may have different number of repeated measurements. Some of those patients would have measurements for both outcomes and time-dependent factors for all five time points, while others may have only 1.

Statistical model

For work status after transplantation, logistic regressions were used for different time periods. Based on the result of initial analysis, generalized linear mixed effect model was used for longitudinal data analysis.

For work percentage after transplantation, simple linear regressions were used for different time periods. Based on the result of initial analysis, linear mixed effect model was used for longitudinal data analysis.

Visualization methods

We used a great number of visualization methods in our study:

1. Boxplot: Age, Sex, Living status, Relationship status, Post-transplantation work percentage. Histogram: Waiting time, BMI and Best FEV1(%).
2. Jitter plot: Two by two table visualization.
3. Heatmap: The distribution of time-varying variables.
4. Likert: Post-transplantation work status.
5. Barplot: Number of repeated measurement for patients.

Implementation

We used a great number of packages for implementation.

1. Data cleaning: *zoo* for the date variables, *stringr* for vectorization of characteristics variables and *tidyr* for gathering values.
2. Table making: *tableone*, *xtable*, *reporttools* and related function in *biostatUZH* from EBPI.
3. Plot: *likert*, *ggplot2*, *ggpubr*,*superheat*, *RColorBrewer* for colors for plots, *ggbeeswarm* and *sjPlot*,
4. Longitudinal data analysis: *lme4*, *lmerTest*.

5 Data Description

Data description for possible predictors

As is shown in section **??**, factors were collected before lung transformation or after transformation. We have repeated measurements on those factors we collected from different time points. However, not all of those post-transplantation factors would change with time. *Best FEV1% pred* is a factor that have repeated measurements, but being time-independent.

We classify the types of predictors based on the whether they change with time, in other words, time-dependent factors and time-independent factors.

Description for time-independent factors

	level	Overall
n		99
Age (mean (SD))		35.02 (10.30)
BMI (mean (SD))		18.32 (3.15)
6MWD (mean (SD))		394.27 (108.03)
Best FEV1 pred (mean (SD))		92.67 (17.84)
Waiting time (mean (SD))		41.83 (38.87)
Sex (%)	Female	49 (49.5)
	Male	50 (50.5)
Education (%)	Non-academic	78 (78.8)
	Academic	21 (21.2)
Relationship (%)	Not single	31 (32.0)
	Single	66 (68.0)
Living (%)	Not alone	79 (81.4)
	Alone	18 (18.6)
Pre-employment (%)	Non-employed	52 (52.5)
	Employed	47 (47.5)

Table 1: Time-independent factors

The main characteristics of the 99 patients are listed in table **??**. The mean age of our study population is 35 years old.

Most of the patients had non-academic education (78.79%), living not alone (81.44%) and being single (31.96%). With regard to sex, there is a balance in this study population.

The mean of the waiting time for a healthy lung in this study population is about 42 weeks (9.8 months). There is high variance of the waiting time among all those patients.

Figure **??** shows that for male and female patients, the distribution of age are very similar. The mean of age in these two groups, are both around 35 years old. The age of half of the patients range from 28 to 42 years old. There is one outlier, who was much older than the rest of patients. This is an individual case with age of 69 years old. The distribution of age of patients illustrate the fact that too old patients would not be proper candidates for lung transplantation.

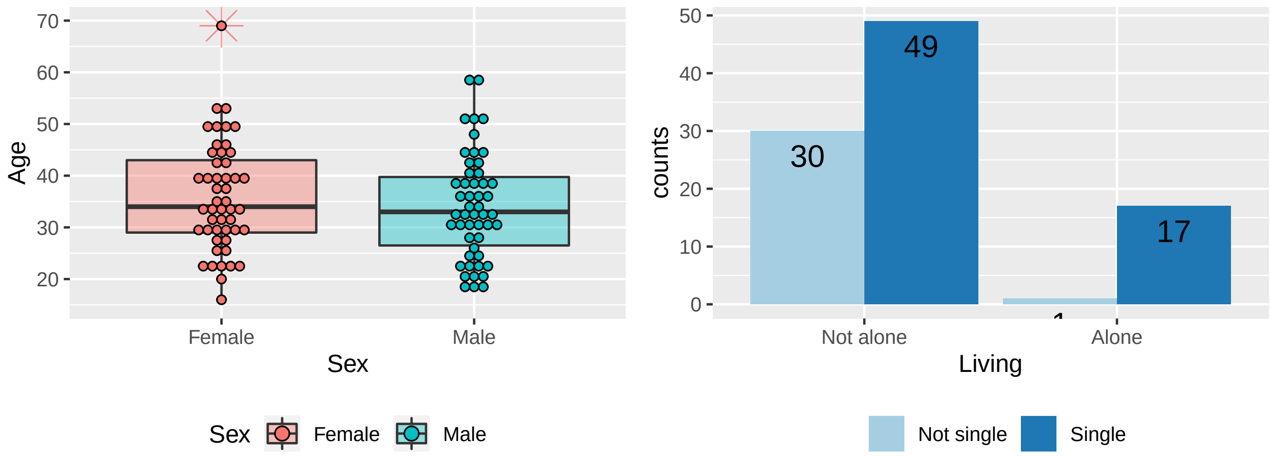

Figure 1: Distributions of population with respect to Sex, Age, Living and Relationship Status

As for relation status and living status, there is an obvious imbalance. Much more patients were living alone, and the proportion of single patients was twice as large as the proportion of the not single group.

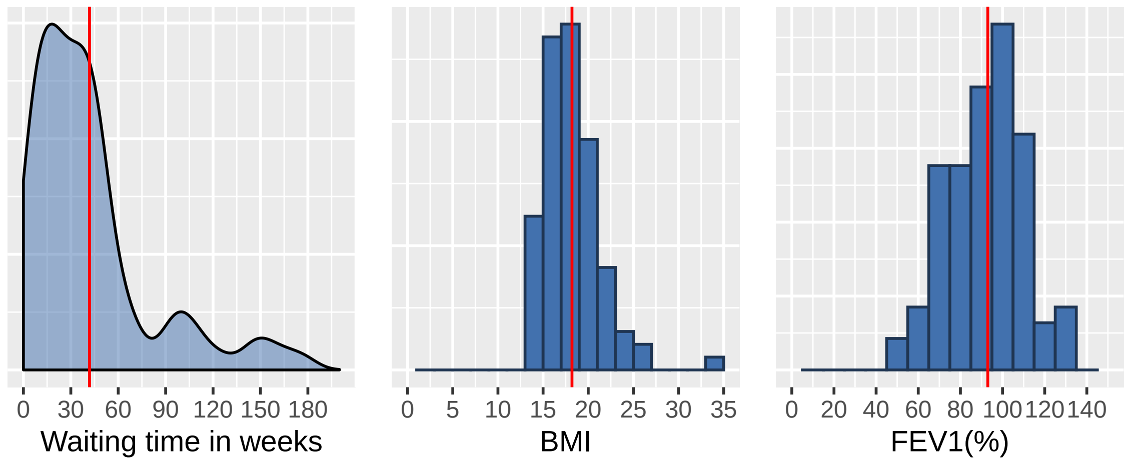

Figure 2: Distributions of some characteristics of physical functions

Figure **??** shows distributions of some characteristics of physical functions and waiting time. the longest waiting time for patients were about 177 weeks, while the lowest waiting time is very close to 0. In total, the majority of patients waited no longer than 60 weeks in this study.

As shown in figure **??**, the mean of BMI is about 18.32 kg/m^2, the distribution of BMI tells us that majority of those 99 patients in this study population had lower than 25kg/m^2 BMI, indicating the underweight problem of cystic fibrosis patients, as a result of dysfunction of digestive system problems.

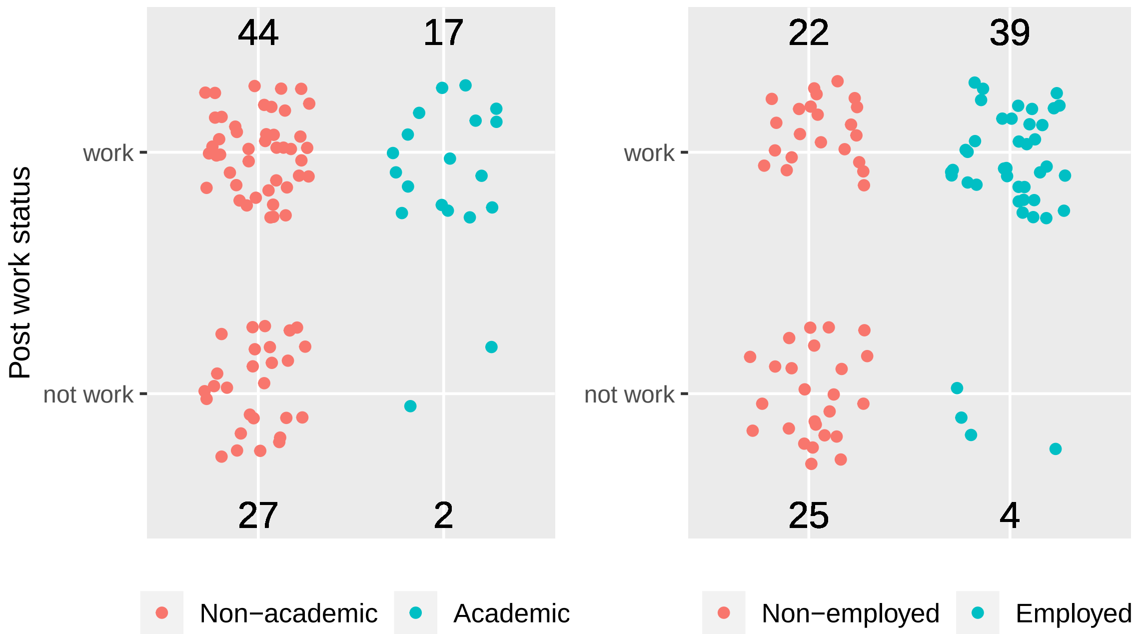

Figure 3: Distributions for Education and Pre-employment with respect to work status three years after LTx

We picked measurements from one specific time point to show the distribution of outcomes. The left figure in figure **??** shows that there are only 2 cases, with academic education background, not working after lung transplantation.

Similarly, the right side in figure **??** illustrates that there are only 4 pre-employed cases, not working after lung transplantation.

The total number in both plots are not corresponding to what is shown in table **??**, since there are some patients that we had no information for the work status one to three years after lung transplantation.

Figure **??** indicates that there may be some relations between education or pre-employment and post-transplantation work status.

Description for time-dependent factors

Table 2: Number of events with respect to time-dependent factors

	1 year	3 years	5 years	10 years	End of Study	NA
CLAD	1	8	12	20	22	4
Cancer	2	2	4	7	8	4
Kidney-Dialysis	0	0	1	7	9	6

Table **??** shows the number of patients who got CLAD, Cancer or Kidney-Dialysis at each time point after lung transplantation. It is very clear that there are not too many events for all of those three complications and rejections (CLAD, Cancer and Kidney or Dialysis) at any time point, especially for the earlier time periods after lung transplantation. The first event of kidney transplantation or dialysis arose three years after lung transplantation.

Figure **??** shows the whole distribution of those three time-dependent factors among those 99 cystic fibrosis patients. The large orange part represents those patients who did not suffer from those

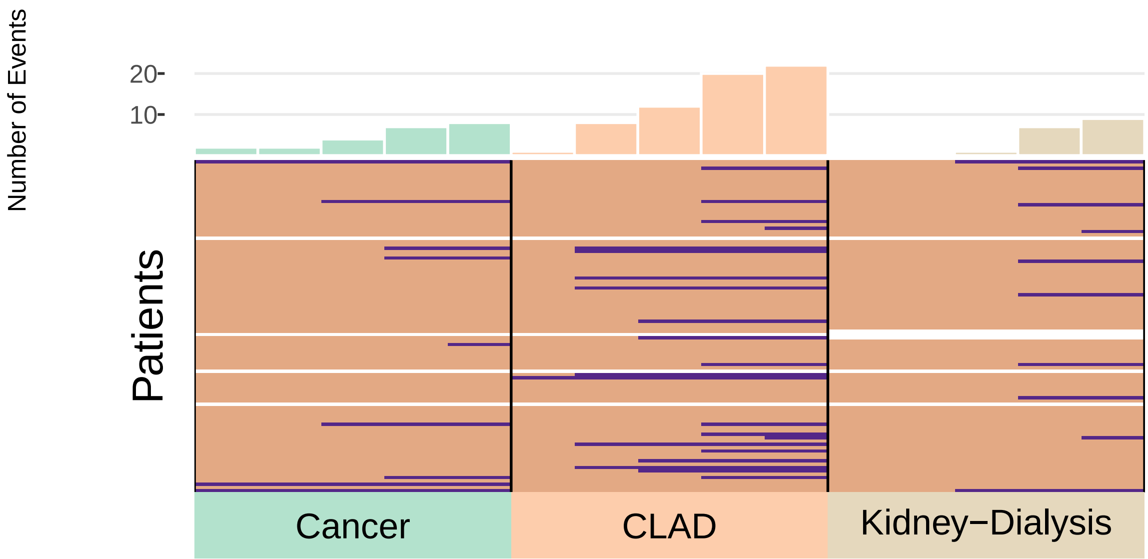

Figure 4: Time-dependent factors of the whole study population

three problems (Cancer, CLAD and Kidney Problems) during whole period of follow-up time. The corresponding proportion was very large, corresponding to the small number of events for any time period in table **??**.

Purple horizontal lines in figure **??** indicate those patients who got CLAD, Cancer or Kidney-Dialysis after lung transplantation. Similarly in the histogram above the heatmap, the number of patients has increased with time.

White horizontal lines showed there are some patients, whose information about *Cancer*, *CLAD* and *Kidney-Dialysis* were missing.

Number of missing values was illustrated in table **??**. We lost the information for all three events of 4 patients and additional 2 patients for *Kidney-Dialysis*.

From the bar on the top of this figure, *CLAD* is much more common and *Kidney-Dialysis* had the smallest number of events. Each bar illustrated the the number of events for each time point after transplantation. During the whole period of follow-up time, number of events for all those three diseases had increased with time. The number of events for all those three complication, rejection or transplantation-related disease were very small. In table **??**, the largest number of events was the number of patients for CLAD, 22 measured at the end of the study.

Data Description for outcomes

Data description for work percentage after LTx

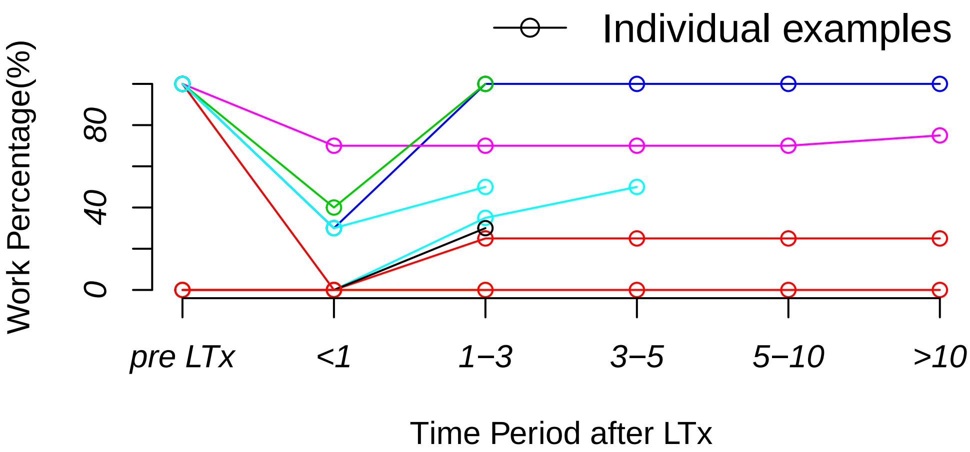

Figure 5: Work percentage for 10 samples during different time periods

In figure **??**, we selected ten patients randomly, each line in the figure represents a single patient. It showed how did the work percentage at each time point change with time. The first time point "pre LTx" indicates the pre-transplantation work percentage, being estimated by the Pre-employed or Pre-not employed status. Therefore, the information of the pre-transplantation work percentage is not biased. We treated those non-employed before lung transplantation before as 0 for work percentage and those employed before lung transplantation as 100(%) of work parentage. These estimates are not accurate, but it can help to show the change of work participation over time after lung transplantation.

It is natural that patients would work less or never work right after lung transplantation, to allow optimal healing. One year after lung transplantation, work percentage started to increase. Three years later, the work percentage of those 10 patients seems to be more or less stable. It should not been ignored that, there are more and more loss to follow-up with time, even though there are only ten patients as a sample there, we can also see that average of work percentage of some patients was not accessible for later period of time.

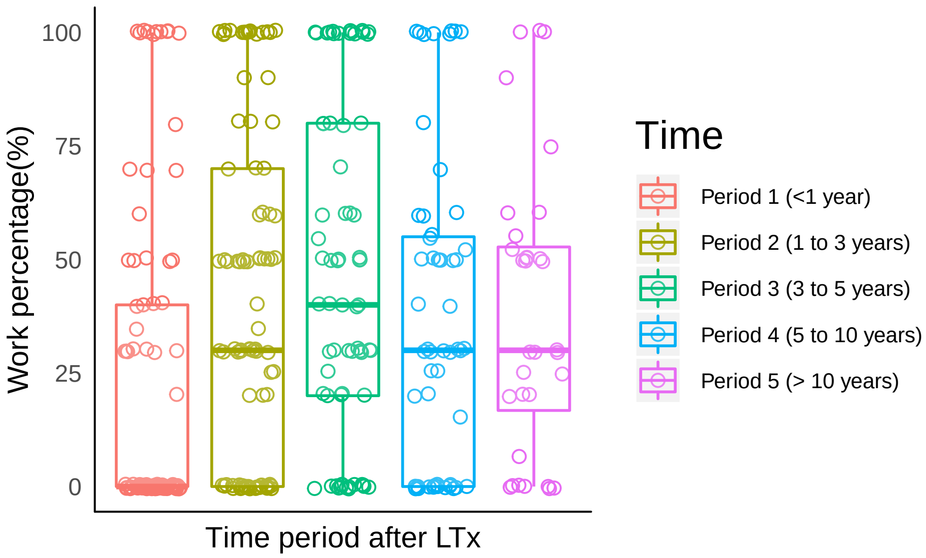

Figure 6: Post-transplantation work percentage of all the periods

We used the estimated pre-transplantation work percentage and only showed the post-transplantation work percentage of a sample of patients with size 10 in figure **??**. Figure **??** showed the work percentage after transplantation of whole study population. The dots in colors indicate the work percentage of patient at each time point after transplantation. The box for each time point shows distribution of work percentage. The upper and lower "hinges" correspond to the first and third quantile, and the middle line indicate the mean of work percentage. From the left to the right, the later of the follow-up period of time, the more patients were loss to follow-up. From

Similar to what we have seen in figure **??** of 10 samples. Within one year after lung transplantation, there were many patients not working at all. It seems to be stable for the mean of work percentage during later time periods, although the information of a large number of patients were not accessible.

Since there was a large proportion of patients not working at all within one year after lung transplantation, we would check the residuals for our final model.

Data description for work status after LTx

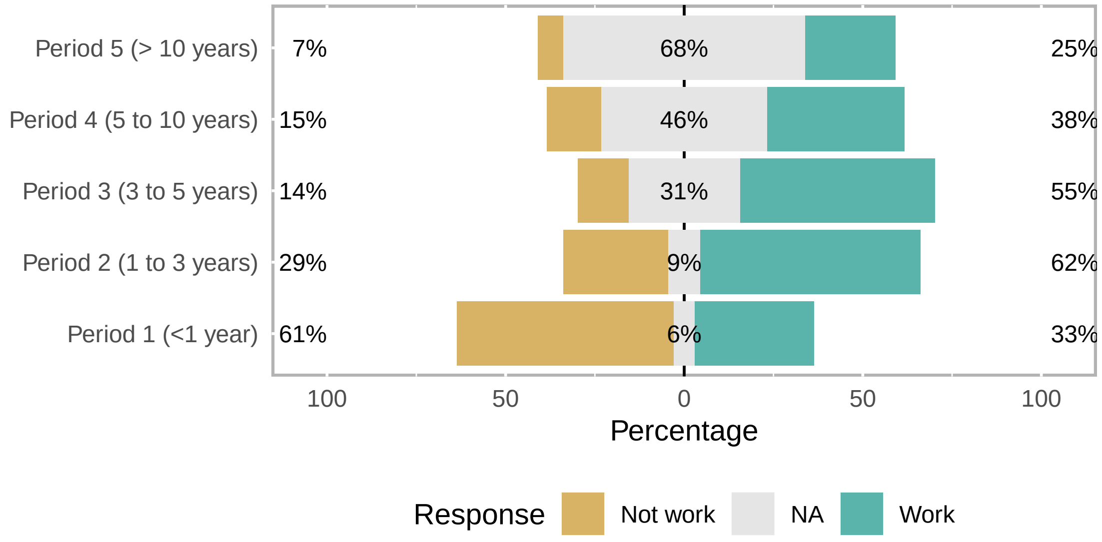

Figure 7: Work status after LTx for all time periods

Table 3: Post-LTx work status

	Period 1	Period 2	Period 3	Period 4	Period 5
Work	60	29	14	15	7
Not work	33	61	54	38	25
NA	6	9	31	46	67
Total	93	90	68	53	32

When it comes to work status after lung transplantation, figure **??** and table**??** showed that, the proportion of loss to follow-up has increased a lot during the follow-up time periods, from 6% of the whole study population to 68% of the whole study population. The proportion of not working patients after transplantation decreased monotonously, from 61% to 7%.

6 Results

The first step was logistic regression for post-transplantation work status and linear regression for post-transplantation work percentage with respect to each time point after lung transplantation.

From table **??**, time-dependent factors *CLAD*, *Cancer* and *Kidney-Dialysis* did not have enough number of events. Therefore, we did not take those three time-dependent factors into consideration before the analysis for repeated measurements.

We used BIC for model selection.

Logistic regression for post-transplantation work status

Post-transplantation work status within one year

Table 4: Univariate models for work status within one year after LTx

	Odds Ratio	95%-confidence interval	p-value	AUC	BIC
Age	0.98	from 0.94 to 1.03	0.48	0.54	105.15
Best FEV1(%)	1.00	from 0.97 to 1.02	0.81	0.59	104.97
BMI	0.89	from 0.75 to 1.03	0.14	0.59	102.81
Education(Academic)	2.17	from 0.79 to 6.02	0.13	0.57	103.40
Living(Alone)	0.35	from 0.08 to 1.20	0.12	0.57	102.56
Pre-employment(Employed)	8.67	from 3.33 to 25.12	< 0.0001	0.74	88.84
Relationship(Single)	1.42	from 0.56 to 3.73	0.47	0.55	104.70
Sex(Male)	0.99	from 0.42 to 2.34	0.99	0.54	105.09
6MWD	1.00	from 0.99 to 1.00	0.43	0.56	104.57
Waiting time	1.00	from 0.99 to 1.01	0.94	0.58	104.93

AUC and BIC were calculated based on the same dataset, which removed patients with missing values. Odds ratios and corresponding 95% confidence interval were calculated from the whole dataset.

Within one year after lung transplantation, there are 60 patients did not work and 33 worked. Our selected model would have at most three predictors.

The univariate model with predictor *Pre-employment*, has the much smaller BIC (88.84) than other univariate models. It may have much more strong effect on the post-transplantation work status than other factors. We would like to take it as a predictor in any of the selected models.

Table 5: Model selection for work status within one year after LTx

	BIC
Pre-employment+Living+Education	92.30
Pre-employment+Education	91.96
Pre-employment+Living	89.46
Pre-employment	88.84

Table 6: Selected model for work status less than one year after LTx

	Odds Ratio	95%-confidence interval	p-value
Pre-employment(Employed)	8.67	from 3.33 to 25.12	< 0.0001

Based on the the odds ratio from those univariate models, factors *Education* and *Living* also seem to have relatively strong effect on the post-transplantation work status. We compared the BIC values of all possible models, as is illustrated in table **??**.

With the smallest value of BIC (88.84), we determine *Pre-employment* as the only predictor for post-transplantation work status one year after lung transplantation. The result from table **??** shows that there is strong evidence for association between *Pre-employment* and work status within one year after lung transplantation. *Pre-employment* is the predictor we determined for work status within one year after lung transplantation.

Post-transplantation work status one to three years

Table 7: Univariate models for work status one to three years after LTx

	Odds Ratio	95%-confidence interval	p-value	AUC	BIC
Age	0.95	from 0.91 to 1.00	0.04	0.63	95.97
Best FEV1(%)	1.00	from 0.98 to 1.03	0.92	0.62	98.92
BMI	0.91	from 0.78 to 1.05	0.21	0.45	97.37
Education(Academic)	5.22	from 1.35 to 34.53	0.04	0.62	92.92
Living(Alone)	0.78	from 0.26 to 2.54	0.67	0.51	99.53
Pre-employment(Employed)	11.08	from 3.73 to 41.39	< 0.0001	0.80	72.59
Relationship(Single)	1.61	from 0.63 to 4.06	0.31	0.58	97.88
Sex(Male)	0.51	from 0.20 to 1.25	0.15	0.59	97.52
6MWD	1.00	from 1.00 to 1.00	1.00	0.54	99.52
Waiting time	1.01	from 1.00 to 1.03	0.13	0.55	98.21

Similarly, from the result of univariate models one to three years after lung transplantation, the one with *Pre-employment* as the only predictor has much lower BIC than other univariate models. The number of patients work work after lung transplantation is 29. Therefore, we would also take at most 3 predictors into consideration.

The results of odds ratios of univariate models show that, besides *Pre-employment*, model with *Education* is the only one whose odds ratio is far from 1. Therefore, *Education* and *Pre-employment* are possible predictors for work status one to three years after lung transplantation.

Table 8: Model selection for work status one to three years after LTx

	BIC
Pre-employment	72.59
Pre-employment+Education	70.33

The results determined *Pre-employment* and *Education* both as the predictors for work status one to three years after lung transplantation.

Table 9: Selected model for work status one to three years after LTx

	Odds Ratio	95%-confidence interval	p-value
Pre-employment(Employed)	11.19	from 3.67 to 42.71	< 0.0001
Education(Academic)	5.34	from 1.22 to 37.82	0.045

Table **??** shows there is also strong evidence that *Pre-employment* has an effect on the work status. The evidence for *Education* as a predictor is moderate.

Post-transplantation work status three to five years

Figure 8: Work status with respect to Pre-employment and Education for three to five year after LTx

There are some extreme cases as shown in figure **??**. We could not analysis the association between *Education* and *Pre-employment*, no academic-educated or pre-employed not working three to five years after transplantation.

It shows association between those two factors, *Education* and *Pre-employment* and the post-transplantation work status during this time period.

From the univariate models in **??**, it is clear that, in spite of *Education* and *Pre-employment*, odds ratio for any of those factor are not far from 1, indicating not very strong effect on the work status.

Since we would do further analyses, we do not choose any kind of correction ways to do regress for work status for those time periods.

Table 10: Univariate models for work status three to five years after LTx

	Odds Ratio	95%-confidence interval	*p*-value	AUC	BIC
Age	0.96	from 0.91 to 1.02	0.16	0.66	60.28
Best FEV1(%)	0.98	from 0.95 to 1.02	0.41	0.67	62.74
BMI	0.95	from 0.80 to 1.16	0.60	0.63	63.22
Living(Alone)	0.39	from 0.10 to 1.71	0.19	0.58	63.98
Relationship(Single)	1.05	from 0.29 to 3.53	0.94	0.59	64.54
Sex(Male)	0.87	from 0.26 to 2.84	0.82	0.57	64.64
6MWD	1.00	from 0.99 to 1.00	0.75	0.44	64.61
Waiting time	1.00	from 0.99 to 1.02	0.80	0.55	64.62

Post-transplantation work status five to ten years

Table 11: Univariate models for work status five to ten years after LTx

	Odds Ratio	95%-confidence interval	p-value	AUC	BIC
Age	0.94	from 0.88 to 1.00	0.07	0.64	57.33
Best FEV1(%)	0.98	from 0.94 to 1.02	0.34	0.64	58.91
BMI	0.90	from 0.71 to 1.11	0.33	0.51	58.65
Education(Academic)	6.46	from 1.09 to 123.87	0.09	0.64	54.93
Living(Alone)	0.44	from 0.10 to 2.07	0.28	0.57	58.22
Pre-employment(Employed)	26.92	from 4.64 to 515.70	0.0025	0.79	44.87
Relationship(Single)	1.00	from 0.26 to 3.52	1.00	0.55	58.78
Sex(Male)	0.88	from 0.26 to 2.91	0.83	0.54	58.96
6MWD	1.00	from 0.99 to 1.00	0.47	0.45	58.85
Waiting time	1.00	from 0.98 to 1.03	0.85	0.53	58.93

Table 12: Model selection for work status five to ten years after LTx

	BIC
Pre-employment	44.87
Pre-employment+Education	41.71

Table 13: Selected model for work status five to ten years after LTx

	Odds Ratio	95%-confidence interval	p-value
Pre-employment(Employed)	36.51	from 5.84 to 724.49	0.001
Education(Academic)	11.56	from 1.61 to 240.46	0.036

For time period three to five years after lung transplantation, there are 15 patients did not work and 38 worked. We took at most 2 predictors into consideration for selected model. It is very obvious that *Education* and *Pre-employment* are the two possible predictors for selected model, as the univariate models for those two show much more strong effect on post-transplantation work status than other factors.

From table **??** and **??**, we determine *Pre-employment* and *Education* as the predictors for the work status five to ten years after lung transplantation. The evidence for those two predictors are strong for *Pre-employment* and moderate for *Education*.

Post-transplantation work status longer than ten years

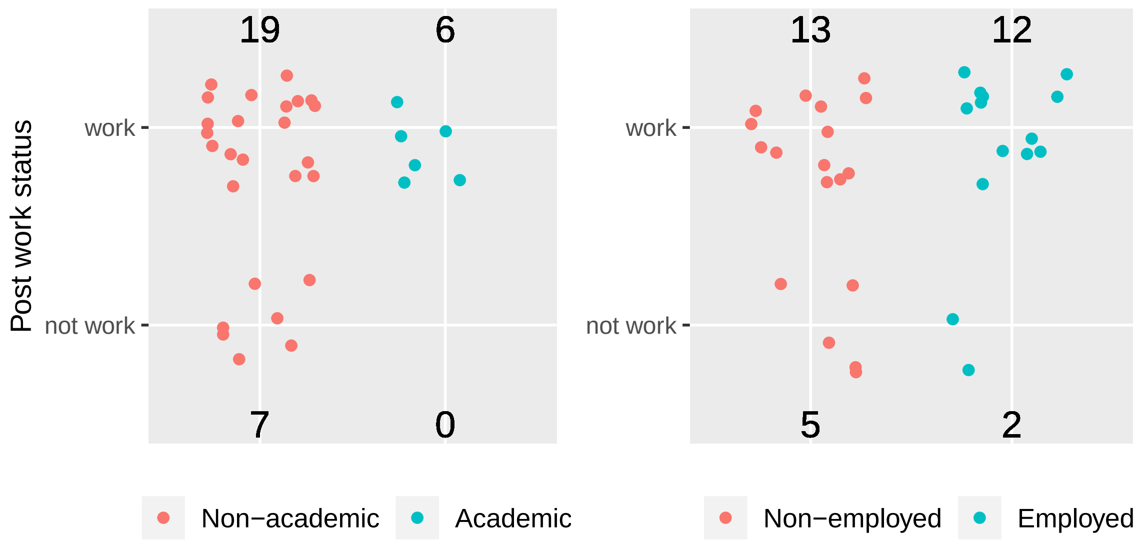

Figure 9: Work status with respect to Pre-employment and Education for longer than ten years after LTx

Table 14: Univariate models for work status longer than ten years after LTx

	Odds Ratio	95%-confidence interval	p-value	AUC	BIC
Age	0.92	from 0.82 to 1.01	0.11	0.66	32.74
Best FEV1(%)	0.98	from 0.93 to 1.03	0.47	0.67	34.31
BMI	0.78	from 0.51 to 1.00	0.13	0.63	32.16
Living(Alone)	0.24	from 0.02 to 2.37	0.20	0.58	33.72
Relationship(Single)	2.70	from 0.42 to 16.86	0.28	0.59	33.83
Sex(Male)	1.23	from 0.23 to 7.35	0.81	0.57	34.19
6MWD	1.00	from 0.99 to 1.01	0.48	0.44	34.27
Waiting time	1.00	from 0.96 to 1.06	0.88	0.55	34.50

Similarly to the time period three to five years after lung transplantation, In figure **??**, there were few cases for academic-educated and pre-employed patients not working after transplantation.

We did not analyse the association between *Education*, *Pre-employment* and work status after transplantation.

From the univariate models in **??**, it is clear that, those factors, in spite of *Education* and *Pre-employment*, do not have very strong effect on the outcome.

Post-transplantation work status for all time periods

Table 15: Result for logistic regression for all possible periods after LTx

	Odds Ratio	95%-confidence interval	p-value
Pre-employment(<1)	8.67	from 3.33 to 25.12	< 0.0001
Pre-employment(1-3)	11.19	from 3.67 to 42.71	< 0.0001
Pre-employment(5-10)	36.51	from 5.84 to 724.49	0.001
	-	-	-
Education(3-5)	5.34	from 1.22 to 37.82	0.045
Education(5-10)	11.56	from 1.61 to 240.46	0.036

Table **??** showed all of the results of logistic regression for three time periods, within one year, one to three year and five to ten years after transplantation.

There are two possible predictors for post-transplantation work status. *Pre-employment* is the dominant one with very strong or strong of evidence. *Education* is another predictor with moderate evidence for association with the work status after lung transplantation.

The width of 95% confidence intervals for both predictors increased with time, due to more and more loss to follow-up with time.

Linear regression for post-transplantation work percentage

Post-transplantation work percentage within one year

Table 16: Univariate models for work percentage less than one year after LTx

	Coefficient	95%-confidence interval	p-value	BIC
Age	-0.78	from -1.46 to -0.09	0.04	742.15
Best FEV1(%)	-0.20	from -0.59 to 0.20	0.11	743.70
BMI	-2.23	from -4.46 to 0.00	0.08	743.10
Education(Academic)	2.20	from -15.02 to 19.42	0.94	746.35
Living(Alone)	-16.98	from -34.32 to 0.37	0.09	743.35
Pre-employment(Employed)	30.71	from 18.06 to 43.36	< 0.0001	728.85
Relationship(Single)	14.36	from 0.11 to 28.61	0.05	742.32
Sex(Male)	-0.87	from -15.03 to 13.29	0.49	745.86
6MWD	-0.03	from -0.10 to 0.04	0.30	745.26
Waiting time	-0.06	from -0.24 to 0.13	0.22	744.78

Table 17: Model selection for work percentage less than one year after LTx

	BIC
Pre-employment+Living+Education	733.38
Pre-employment+Education	732.99
Pre-employment+Living	729.21
Pre-employment	728.85

Table 18: Selected model for work percentage one to three years after LTx

	Coefficient	95%-confidence interval	p-value
Pre-employment(Employed)	30.71	from 18.06 to 43.36	< 0.0001

For those univariate models for linear regression for work percentage, we choose *Pre-employment*, *Living*, *Education* for model selection, since those three seem to have much more strong effect on the outcome than the others from the coefficients from table **??**.

After model selection based on BIC, we determined *pre-employment* as the predictor for work percentage one year after lung transplantation. Form the p-value in the table **??**, we can also see strong evidence for this predictor.

Post-transplantation work percentage one to three years

Table 19: Univariate models for work percentage one to three years after LTx

	Coefficient	95%-confidence interval	p-value	BIC
Age	-1.20	from -1.91 to -0.48	0.0013	732.13
Best FEV1(%)	0.00	from -0.44 to 0.45	0.99	740.81
BMI	-3.40	from -5.85 to -0.94	0.0072	735.00
Education(Academic)	15.73	from -3.11 to 34.56	0.10	740.01
Living(Alone)	-20.07	from -39.89 to -0.25	0.05	738.99
Pre-employment(Employed)	35.41	from 21.69 to 49.12	< 0.0001	714.92
Relationship(Single)	13.66	from -2.58 to 29.90	0.10	737.90
Sex(Male)	-6.16	from -21.79 to 9.46	0.44	740.38
6MWD	-0.00	from -0.08 to 0.07	0.94	741.17
Waiting time	0.13	from -0.07 to 0.32	0.20	740.02

Table 20: Model selection for work percentage one to three years after LTx

	BIC
Pre-employment+Education	718.51
Pre-employment+Living+Education	718.44
Pre-employment+Living	715.06
Pre-employment	714.92

Table 21: Selected model for work percentage one to three years after LTx

	Coefficient	95%-confidence interval	p-value
Pre-employment(Employed)	35.41	from 21.69 to 49.12	< 0.0001

The possible predictors for work percentage one to three years after lung transplantation, are *Pre-employment*, *Living* and *Education*. Although the model with *Pre-employment* and *Education* as the predictors have the smallest BIC, we would also choose the model with only *Pre-employment* as the predictor. Since the difference between these two model is smaller than 2, and we would like to choose the one with smaller number of predictors.

P-value in table **??** also showed strong evidence for *Pre-employment* as a predictor, but the 95% confidence interval is wider than before due to smaller sample size.

Post-transplantation work percentage three to five years

Table 22: Univariate models for work percentage three to five years after LTx

	Coefficient	95%-confidence interval	p-value	BIC
Age	-1.27	from -2.05 to -0.49	0.0019	528.53
Best FEV1(%)	-0.30	from -0.82 to 0.22	0.26	534.18
BMI	-3.46	from -6.25 to -0.66	0.02	531.87
Education(Academic)	20.00	from -0.38 to 40.38	0.05	536.27
Living(Alone)	-20.91	from -44.13 to 2.31	0.08	536.50
Pre-employment(Employed)	38.69	from 23.66 to 53.73	< 0.0001	514.05
Relationship(Single)	13.28	from -5.04 to 31.59	0.15	536.42
Sex(Male)	-3.65	from -21.48 to 14.18	0.68	538.32
6MWD	-0.01	from -0.10 to 0.07	0.77	538.25
Waiting time	0.13	from -0.12 to 0.39	0.31	537.75

Table 23: Model selection for work percentage three to five years after LTx

	BIC
Pre-employment+Living+Education	515.77
Pre-employment+Education	515.60
Pre-employment+Living	515.12
Pre-employment	514.05

Table 24: Selected model for work status three to five years after LTx

	Coefficient	95%-confidence interval	p-value
Pre-employment(Employed)	38.69	from 23.66 to 53.73	< 0.0001

For work percentage three to five years after lung transplantation, model with only *Pre-employment* has the lowest BIC. The evidence for the association between *Pre-employment* and work percentage is very strong during this time period.

Post-transplantation work percentage five to ten years

Table 25: Univariate models for work percentage five to ten years after LTx

	Coefficient	95%-confidence interval	*p*-value	BIC
Age	-0.96	from -1.89 to -0.03	0.04	394.83
Best FEV1(%)	-0.20	from -0.78 to 0.37	0.48	396.56
BMI	-2.31	from -5.58 to 0.95	0.16	396.01
Education(Academic)	20.38	from -0.92 to 41.69	0.06	395.24
Living(Alone)	-24.89	from -48.01 to -1.76	0.04	393.15
Pre-employment(Employed)	32.92	from 16.34 to 49.50	0.00021	379.69
Relationship(Single)	4.62	from -14.91 to 24.15	0.64	396.86
Sex(Male)	-7.63	from -26.49 to 11.24	0.42	396.7
6MWD	-0.04	from -0.12 to 0.05	0.41	395.98
Waiting time	0.08	from -0.25 to 0.40	0.64	396.83

Table 26: Model selection for work percentage five to ten years after LTx

	BIC
Pre-employment+Living+Education	378.99
Pre-employment+Education	379.85
Pre-employment	379.69
Pre-employment+Living	379.20

Table 27: Selected model for work status five to ten years after LTx

	Coefficient	95%-confidence interval	*p*-value
Pre-employment(Employed)	32.92	from 16.34 to 49.50	0.0002

For work percentage five to ten years after lung transplantation, model with predictors, *Pre-employment* and *Living*, has the lowest BIC, but the BIC for model with only *Pre-employment* as the predictor, is not so much larger than this one. Therefore, the predictor we determined is *Pre-employment*.

Post-transplantation work percentage longer than ten year

Table 28: Univariate models for work percentage longer than ten years after LTx

	Coefficient	95%-confidence interval	*p*-value	BIC
Age	-1.01	from -2.24 to 0.22	0.10	234.5
Best FEV1(%)	-0.04	from -0.72 to 0.64	0.90	236
BMI	-2.65	from -5.75 to 0.45	0.09	234.38
Education(Academic)	29.04	from 1.58 to 56.50	0.04	233.62
Living(Alone)	-28.62	from -62.83 to 5.60	0.10	233.41
Pre-employment(Employed)	20.23	from -1.74 to 42.20	0.07	229.41
Relationship(Single)	8.25	from -19.21 to 35.71	0.54	235.81
Sex(Male)	2.00	from -21.08 to 25.07	0.86	236.03
6MWD	-0.02	from -0.16 to 0.12	0.74	235.84
Waiting time	0.23	from -0.36 to 0.81	0.44	235.57

Table 29: Model selection for work percentage longer than ten years after LTx

	BIC
Pre-employment+Living	231.79
Pre-employment+Living+Education	231.47
Pre-employment	229.41
Pre-employment+Education	229.36

Table 30: Selected model for work status longer than ten years after LTx

	Coefficient	95%-confidence interval	*p*-value
Pre-employment(Employed)	20.23	from -1.74 to 42.20	0.07

Though the model with lowest value of BIC is the one with *Pre-employment* and *Education*, we would still determine *Pre-employment* as the only predictor. The difference between values of BIC of the those two models is smaller than two, so we choose the one with smaller number of predictors.

For this time period, we get the same result as the previous time periods, *Pre-employment* is the predictor that we determined.

Post-transplantation work percentage for all time periods

Table 31: Result for linear regression for periods after LTx

	Coefficients	95%-confidence interval	p-value
Pre-employment(<1)	30.71	from 18.06 to 43.36	< 0.0001
Pre-employment(1-3)	35.41	from 21.69 to 49.12	< 0.0001
Pre-employment(3-5)	38.69	from 23.66 to 53.73	< 0.0001
Pre-employment(5-10)	32.92	from 16.34 to 49.50	0.0002
Pre-employment(>10)	20.23	from -1.74 to 42.20	0.07

From table **??**, we get the conclusion that, results for work percentage for all time periods are consistent. *Pre-employment* is the only predictor that we determined for work percentage after lung transplantation. The evidence is very strong for the fist three time periods and relatively weaker for the last time period. 95% confidence interval also gets wider and wider.

Longitudinal data analysis for post-transplantation work status

The aim of our study is to evaluate the long-term work participation after lung transplantation. In an attempt to take time, time-varying variables and repeated measurements of outcomes into one model, we did the longitudinal data analysis.

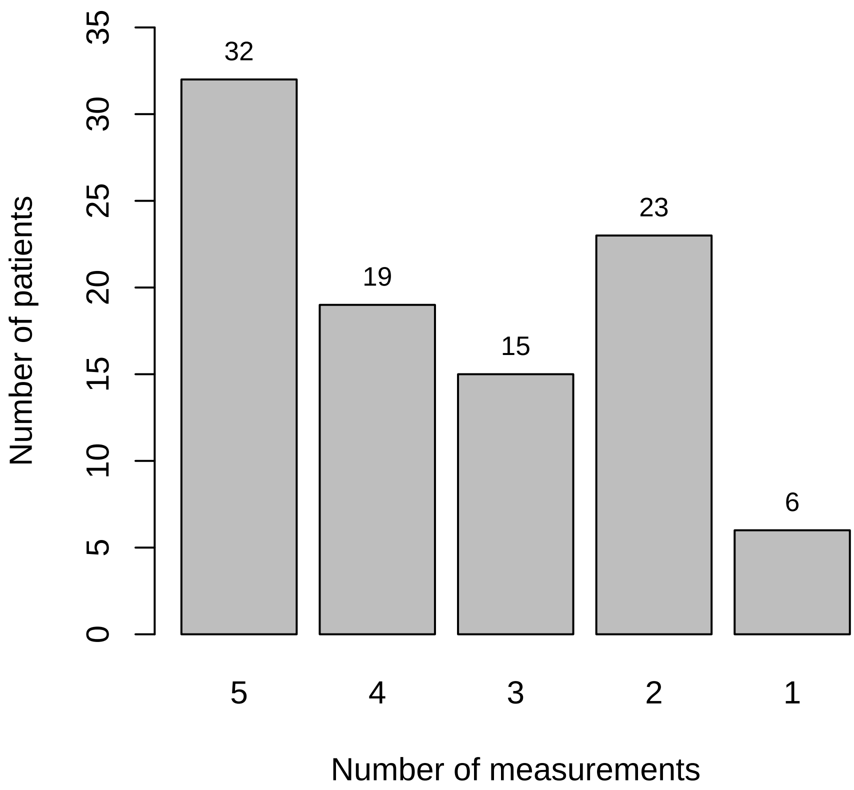

Figure 10: The number of patients corresponding to number of measurements

From figure **??**, we can see the number of repeated measurements are different for different patients. Since the time they got the lung transplantation are different, the study period for each patient are different. There are 32 patients, undergoing lung transplantation very early and have measurements for all five time points. 6 patients got the lung transplantation too late, we only had access to their measurements of work participation and the possible post-transplantation rejections and complications at the end of the study.

Based on those repeated measurements and the exact time for each patient, we got the long-format data. From the results from the previous results from linear regression and logistic regression for separate time points, we would be able to construct mixed-effects models for work status and work percentage after lung transplantation.

Generalized linear mixed effect model for work status after LTx

From the results of logistic model, there are two possible predictors for the work status after lung transplantation, *Pre-employment* and *Education*. We select model for longitudinal data analysis based on the result and three transformation of Time.

There a great number of ways to transform time, we chose the most common ones, log(Time), Time and Time^2 , we selected the form of transformation with the lowest BIC.

Table 32: Time transformation for work status after LTx based on BIC

	Pre-employment	Education + Pre-employment
log(Time)	352.37	343.07
Time	371.11	361.61
$(\text{Time})^2$	383.84	373.83

Table **??** shows that we would choose log(Time) as the transformation for time. Both of the time-independent factors, *Pre-employment* and *Education* would go into our model for generalized linear mixed-effects model.

Table 33: Model selection for work status based on Time-dependent factors

	CLAD	Cancer	Kidney-Dialysis	BIC
Model 1	0	0	1	333.75
Model 2	0	1	0	336.48
Model 3	1	0	0	337.00
Model 4	0	1	1	339.00
Model 5	1	0	1	339.48
Model 6	1	1	0	342.18
Model 7	1	1	1	344.69

From table **??**, among all those three time-dependent factors, we chose *CLAD* as the only predictor for our model, and Time variable is transformed into as log(Time).

Table 34: Result for Generalized Linear Mixed-effects Model for Post-LTx Work Status

	Odds Ratio	95%-confidence interval	*p*-value
Pre-employment	32.16	from 9.48 to 177.09	< 0.0001
Education	4.40	from 1.07 to 20.55	0.04
CLAD	0.64	from 0.15 to 2.40	0.49
log(Time)	4.94	from 2.95 to 9.38	< 0.0001

The result from generalized linear fixed effect model **??** showed that there is moderate evidence that time-dependent factor, *CLAD*, is the predictor for post-transplantation work status.

Linear mixed-effect model for work status after LTx

Analogously, the first thing we did for linear mixed-effects model was time transformation.

Table 35: Time transformation for work percentage after LTx

	Pre-employment	Education + Pre-employment
log(Time)	3291.53	3288.63
Time	3302.13	3298.95
$(\text{Time})^2$	3307.03	3303.80

Table **??** shows that the time transformation for long-term work percentage is also log(Time), and *Education* which we determined as a predictor for post-transplantation work status during the analyses for each time period, would also go into this model with respect to work percentage after lung transplantation.

Table 36: Model selection for work percentage based on Time-dependent factors

	CLAD	Cancer	Kidney-Dialysis	BIC
Model 1	0	0	1	3192.72
Model 2	0	1	0	3197.34
Model 3	1	0	0	3197.52
Model 4	0	1	1	3198.28
Model 5	1	0	1	3198.46
Model 6	1	1	0	3203.09
Model 7	1	1	1	3204.05

Based on the BIC values for all possible choices of combination of time-dependent factors, we determined to include *Kidney-Dialysis* as an additional predictor into the linear mixed-effect model.

The result from linear fixed-effect model **??** shows that there is moderate evidence for *Kidney-Dialysis* as a predictor.

Table 37: Result for Linear Mixed-effects Model for Post-LTx Work percentages

	Coefficients	95%-confidence interval	*p*-value
Pre-employment	31.66	from 21.00 to 42.31	< 0.0001
Education	7.82	from -5.04 to 20.64	0.24
Kidney-Dialysis	-18.69	from -31.80 to -5.59	0.0056
log(Time)	6.91	from 4.00 to 9.91	< 0.0001

7 Conclusion

From the results of longitudinal data analysis, we determine *Pre-employment* and *Education* as the predictors for long-term work status. The odds ratio for *Pre-employment* is 32.16, from 9.48 to 177.09. Odds ratio for *Education* is 4.4, 95% confidence interval from 1.07 to 20.55, showing moderate evidence of the effect on the post-transplantation work status. *CLAD* is also a predictor, with odds ratio 0.64, 95% confidence interval from 0.15 to 2.40.

Similarly for work percentage, *Pre-employment* showed strong evidence as a predictor, while the evidence for *Education* is weak. The evidence for the negative effect of *Kidney-Dialysis* is moderate.

The effect of lung transplantation on work participation would be better and better with time. From the our results, patients who were employed and with academic education would be more likely to return to work after lung transplantation.

In an attempt to improve the work participation after lung transplantation, measurements should be taken to avoid CLAD and also some renal dysfunctions.

R version and packages used to generate this report:

R version: R version 3.5.2 (2018-12-20)

Base packages: stats, graphics, grDevices, utils, datasets, methods, base

Other packages: lmerTest 3.1-0, huxtable 4.7.0, stringr 1.4.0, lme4 1.1-21, Matrix 1.2-17, sjPlot 2.7.2, tidyr 1.0.0, dplyr 0.8.3, reshape2 1.4.3, superheat 0.1.0, ggstance 0.3.3, jtools 2.0.1, ggbeeswarm 0.6.0, plyr 1.8.4, ggpubr 0.2, magrittr 1.5, reporttools 1.1.2, biostatUZH 1.8.0, survival 2.43-3, likert 1.3.5, ggplot2 3.2.0, xtable 1.8-4, tableone 0.10.0, RColorBrewer 1.1-2, zoo 1.8-6, knitr 1.23

This document was generated on October 10, 2019 at 20:47.

8 Appendix

```
####### code for packages, settings, data prep
library(zoo)#manipulation of pate variables
library(RColorBrewer) # colors for plots
library(tableone) # for Table 1 functions
library(xtable) # formatting tables and generating the tex code
library(likert)
library(biostatUZH) # EBPI-written package
library(ggplot2) # customizable plots
library(reporttools)#to make table in report
library(ggpubr)
library(plyr)
library(ggbeeswarm)
library(jtools)
library(ggstance)
library(superheat)
library(reshape2)
library(dplyr)
library(tidyr)
library(sjPlot)
library(lme4)#linear mixed effects models
library(stringr)#manipulation of work percentage
library(huxtable)
library(lmerTest)


#source("../code/data_cleaning.R")
#Âfdata cleaning
## Import Data
## ------------
dat <- read.csv("../data/newdata.csv")


## sample size of the study population: 99
## ----------------
dat <- dat[,-1]
dat <- unique(dat)
dat <- dat[-nrow(dat),]
dim(dat)#99  60

## extract the useful ones
## ----------------
binary_pred <- (dat[,c(8,3,4,6,28,51,55,59)])
colnames(binary_pred) <- c("Sex","Education","Relationship","Living",
                           "Employment","CLAD","kidney_Tx","Cancer")
num_pred <- dat[,c(9,26,27,44)]
colnames(num_pred) <- c("Age","BMI","6MWD","FEV1")
employ_stat <- dat[,c(30,32,34,36,38)]
colnames(employ_stat) <- c("less_one","one_three","three_five",
                           "five_ten","longer_ten")
employ_percent <- dat[,c(31,33,35,37,39)]
colnames(employ_percent) <- c("less_one_perc","one_three_perc",
                              "three_five_perc","five_ten_perc",
                              "longer_ten_perc")

##relevel binary variables
```

```r
## ----------------
binary_pred[] <- lapply(binary_pred, as.factor)
employ_stat[] <- lapply(employ_stat,as.factor)

## creating binary variables
## ----------------
pred <- cbind(binary_pred, num_pred)
out_percent <- employ_percent
out_stat <- employ_stat

levels(pred[,"Sex"]) <- c("male", "female")
levels(pred[,"Education"]) <- c("Non-academic", "Academic ")

levels(pred[,"Relationship"])[levels(pred[,"Relationship"])=="Divorced"|
                              levels(pred[,"Relationship"])=="Single"] <- "single"
levels(pred[,"Relationship"])[levels(pred[,"Relationship"])=="Engaged"|
                              levels(pred[,"Relationship"])=="Relationship"|
                              levels(pred[,"Relationship"])=="Married"] <- "not single"
levels(pred[,"Relationship"])[levels(pred[,"Relationship"])!="single"&
                              levels(pred[,"Relationship"])!="not single"] <- NA
#two missing values for relationship status


pred$Living <- ifelse(pred$Living=="allein","alone","not alone")
pred$Living <- as.factor(pred$Living)

pred$Employment <- as.factor(ifelse(pred$Employment==0,"Non-employed","Employed"))

pred$CLAD <- as.factor(ifelse(pred$CLAD==0,"no","yes"))
pred$kidney_Tx <- as.factor(ifelse(pred$kidney_Tx==0,"no","yes"))
pred$Cancer <- as.factor(ifelse(pred$Cancer==0,"no","yes"))


## work percentage(from nomial variables to continuous ones)
## ----------------

#less than one year work percentage
factor1 <- levels(out_percent$less_one_perc)[out_percent$less_one_perc]
factor1 <- str_replace(factor1, "40-60", "50")
factor1 <- str_replace(factor1, "20 - 50", "35")
factor1 <- str_replace(factor1, "30 - 50", "40")
factor1 <- str_replace(factor1, "60-80", "70")
out_percent$less_one_perc <- as.numeric(factor1)

#one to three years work percentage
factor2 <- levels(out_percent$one_three_perc)[out_percent$one_three_perc]
factor2 <- str_replace(factor2, "40-60", "50")
factor2 <- str_replace(factor2, "20 - 50", "35")
factor2 <- str_replace(factor2, "30 - 50", "40")
factor2 <- str_replace(factor2, "60-80", "70")
factor2 <- str_replace(factor2, "80 - 100", "90")
factor2 <- str_replace(factor2, "20-30", "25")
factor2 <- str_replace(factor2, "20 - 30", "25")
factor2 <- str_replace(factor2, "30 - 40", "35")
out_percent$one_three_perc <- as.numeric(factor2)

#three to five years work percentage
```

```
factor3 <- levels(out_percent$three_five_perc)[out_percent$three_five_perc]
factor3 <- str_replace(factor3, "20-30", "25")
factor3 <- str_replace(factor3, "50-60", "55")
out_percent$three_five_perc <- as.numeric(factor3)

#five to ten years work percentage
factor4 <- levels(out_percent$five_ten_perc)[out_percent$five_ten_perc]
factor4 <- str_replace(factor4, "20-30", "25")
factor4 <- str_replace(factor4, "50-60", "55")
factor4 <- str_replace(factor4, "50 - 60", "55")
out_percent$five_ten_perc <- as.numeric(factor4)

#longer than ten years work percentage
factor5 <- levels(out_percent$longer_ten_perc)[out_percent$longer_ten_perc]
factor5 <- str_replace(factor5, "40-80", "60")
factor5 <- str_replace(factor5, "20-30", "25")
factor5 <- str_replace(factor5, "80-100", "90")
factor5 <- str_replace(factor5, "(5-10)", "7")
factor5[47] <- "7"
factor5 <- str_replace(factor5, "50-60", "55")
factor5 <- str_replace(factor5, "20-30", "25")
out_percent$longer_ten_perc <- as.numeric(factor5)

## consistency btw work percentage and work status
## ----------------
outcome <- cbind(ind=1:99,pred,out_percent,out_stat)
outcome[which(outcome[,"less_one"]==0),][,"less_one_perc"]
outcome[which(outcome[,"less_one"]==1),][,"less_one_perc"]
outcome[which(is.na(outcome[,"less_one"])==TRUE),][,"less_one_perc"]
#no wired data for percentage


outcome[which(outcome[,"one_three"]==0),][,"one_three_perc"]
#1 wired data for 1-3 years percentage
outcome[which((outcome[,"one_three"]==0)&
                (outcome[,"one_three_perc"] > 0)),][,"one_three"] <- 1
outcome[which(outcome[,"one_three"]==1),][,c("ind","one_three_perc")]
outcome[81,"one_three"] <- 0
outcome[which(is.na(outcome[,"one_three"])==TRUE),][,c("ind","one_three_perc","one_three")]
# the 4 and 40 indicator patient
outcome[4,"one_three"] <- 0
outcome[40,"one_three"] <- 1


outcome[which(outcome[,"three_five"]==0),][,c("ind","three_five_perc","three_five")]
#two wired data for 3-5 years percentage
outcome[which((outcome[,"three_five"]==0)
              &(outcome[,"three_five_perc"] > 0)),][,"three_five"] <- 1
outcome[which(outcome[,"three_five"]==1),][,c("ind","three_five_perc","three_five")]
outcome[which(is.na(outcome[,"three_five"])==TRUE),][,c("ind","three_five_perc","three_five")]


outcome[which(outcome[,"five_ten"]==0),][,c("ind","five_ten_perc","five_ten")]
# one wired data for percentage
outcome[52,"five_ten_perc"] <- 0
outcome[which(outcome[,"five_ten"]==1),][,c("ind","five_ten_perc","five_ten")]
outcome[which((outcome[,"five_ten"]==1)
```

```
              &(outcome[,"five_ten_perc"]==0)),][,"five_ten"] <- 0
outcome[which(is.na(outcome[,"five_ten"])==TRUE),][,c("ind","five_ten_perc","five_ten")]
outcome[58,"five_ten"] <- 1
outcome[68,"five_ten"] <- 1


outcome[which(outcome[,"longer_ten"]==0),][,c("ind","longer_ten_perc","longer_ten")]
outcome[4,"longer_ten_perc"] <- 0
outcome[which(outcome[,"longer_ten"]==1),][,c("ind","longer_ten_perc","longer_ten")]
outcome[which(is.na(outcome[,"longer_ten"])==TRUE),][,c("ind",
                                                        "longer_ten_perc","longer_ten")]
outcome[78,"longer_ten"] <- 1

date <- dat[,c("LTx_date","CLAD_ED",
               "Kidney_Tx_Date","Start_Dialyse_Date","Cancer_ED","Waiting_list_date")]
dat <- cbind(outcome,date)


## creat time-varying variables by the date of LTx and case date
## ---------------

##CLAD
CLAD_day <- dat[which(dat[,"CLAD"]=="yes"),][,c("ind","LTx_date","CLAD_ED")]
CLAD_day$LTx_date <- as.character(CLAD_day$LTx_date)
CLAD_day$LTx_date <- as.Date(CLAD_day$LTx_date,na.action=TRUE,format="%d/%m/%Y")
CLAD_day$CLAD_ED <- as.character(CLAD_day$CLAD_ED)
CLAD_day$CLAD_ED  <- as.yearmon(CLAD_day$CLAD_ED,"%m/%y")
CLAD_day$CLAD_ED <- as.Date(CLAD_day$CLAD_ED,na.action=TRUE,format="%m/%y")


CLAD_time <-  difftime(CLAD_day$CLAD_ED,CLAD_day$LTx_date, units = "days") # days
CLAD_time_year <- cbind(CLAD_day$ind,(CLAD_time/365))#approximate years
CLAD_day <- cbind(CLAD_day,duration=CLAD_time_year[,2])


#CLAD: creat binary variable for each time period
#1 year later
CLAD_1 <- matrix(dat$CLAD,nrow=length(dat[,"CLAD"]),ncol=1)
CLAD_1[-CLAD_day[which(CLAD_day["duration"]<1),][,"ind"],] <- "no"
CLAD_1[dat[which(is.na(dat$CLAD)==TRUE),][,"ind"]] <- NA
#3 years later
CLAD_3 <- matrix(dat$CLAD,nrow=length(dat[,"CLAD"]),ncol=1)
CLAD_3[-CLAD_day[which(CLAD_day["duration"]<3),][,"ind"],] <- "no"
CLAD_3[dat[which(is.na(dat$CLAD)==TRUE),][,"ind"]] <- NA
#5 year later
CLAD_5 <- matrix(dat$CLAD,nrow=length(dat[,"CLAD"]),ncol=1)
CLAD_5[-CLAD_day[which(CLAD_day["duration"]<5),][,"ind"],] <- "no"
CLAD_5[dat[which(is.na(dat$CLAD)==TRUE),][,"ind"]] <- NA
#10 years later
CLAD_10 <- matrix(dat$CLAD,nrow=length(dat[,"CLAD"]),ncol=1)
CLAD_10[-CLAD_day[which(CLAD_day["duration"]<10),][,"ind"],] <- "no"
CLAD_10[dat[which(is.na(dat$CLAD)==TRUE),][,"ind"]] <- NA
dat <- cbind(dat,CLAD_1,CLAD_3,CLAD_5,CLAD_10)



#CLAD
```

```r
Cancer_day <- dat[which(dat[,"Cancer"]=="yes"),][,c("ind","LTx_date","Cancer_ED")]
Cancer_day$LTx_date <- as.character(Cancer_day$LTx_date)
Cancer_day$LTx_date <- as.Date(Cancer_day$LTx_date,na.action=TRUE,format="%d/%m/%Y")
Cancer_day$Cancer_ED <- as.character(Cancer_day$Cancer_ED)
Cancer_day$Cancer_ED[4] <- "04/11"
Cancer_day$Cancer_ED  <- as.yearmon(Cancer_day$Cancer_ED,"%m/%y")
Cancer_day$Cancer_ED <- as.Date(Cancer_day$Cancer_ED,na.action=TRUE,format="%m/%y")
Cancer_time <-  difftime(Cancer_day$Cancer_ED,Cancer_day$LTx_date, units = "days") # days
Cancer_time_year <- cbind(Cancer_day$ind,(Cancer_time/365))#approximate years
Cancer_day <- cbind(Cancer_day,duration=Cancer_time_year[,2])
#1 year later
Cancer_1 <- matrix(dat$Cancer,nrow=length(dat[,"Cancer"]),ncol=1)
Cancer_1[-Cancer_day[which(Cancer_day["duration"]<1),][,"ind"],] <- "no"
Cancer_1[dat[which(is.na(dat$Cancer)==TRUE),][,"ind"]] <- NA

#3 year later
Cancer_3 <- matrix(dat$Cancer,nrow=length(dat[,"Cancer"]),ncol=1)
Cancer_3[-Cancer_day[which(Cancer_day["duration"]<3),][,"ind"],] <- "no"
Cancer_3[dat[which(is.na(dat$Cancer)==TRUE),][,"ind"]] <- NA

#5 year later
Cancer_5 <- matrix(dat$Cancer,nrow=length(dat[,"Cancer"]),ncol=1)
Cancer_5[-Cancer_day[which(Cancer_day["duration"]<5),][,"ind"],] <- "no"
Cancer_5[dat[which(is.na(dat$Cancer)==TRUE),][,"ind"]] <- NA

#10 year later
Cancer_10 <- matrix(dat$Cancer,nrow=length(dat[,"Cancer"]),ncol=1)
Cancer_10[-Cancer_day[which(Cancer_day["duration"]<10),][,"ind"],] <- "no"
Cancer_10[dat[which(is.na(dat$Cancer)==TRUE),][,"ind"]] <- NA

dat <- cbind(dat,Cancer_1,Cancer_3,Cancer_5,Cancer_10)


#Kidney_dialysis
kidney_Dialysis_day <- dat[which(dat[,"kidney_Tx"]=="yes"),][,c("ind",
                                                                "LTx_date","kidney_Tx",
                                                                "Kidney_Tx_Date",
                                                                "Start_Dialyse_Date" )]

######9 cases
#two missing both date(kidney_TX and Dialyse)
dat[47,"kidney_Tx"] <- NA
dat[49,"kidney_Tx"] <- NA

kidney_Dialysis <- dat$kidney_Tx
#1 year later
kidney_Dialysis_1 <- matrix(dat$kidney_Tx,nrow=length(dat[,"kidney_Tx"]),ncol=1)
kidney_Dialysis_1[which(kidney_Dialysis_1=="yes")] <- "no"
#no case
kidney_Dialysis_1 <- kidney_Dialysis_1
#3 years later
kidney_Dialysis_3 <- matrix(dat$kidney_Tx,nrow=length(dat[,"kidney_Tx"]),ncol=1)
kidney_Dialysis_3[which(kidney_Dialysis_3=="yes")] <- "no"
#no case
kidney_Dialysis_3 <- kidney_Dialysis_1
#5 years later
kidney_Dialysis_5 <- matrix(dat$kidney_Tx,nrow=length(dat[,"kidney_Tx"]),ncol=1)
```

```
kidney_Dialysis_5[which(kidney_Dialysis_5=="yes")] <- "no"
kidney_Dialysis_5[1,] <- "yes"
#1 case
#10 years later
kidney_Dialysis_10 <- matrix(dat$kidney_Tx,nrow=length(dat[,"kidney_Tx"]),ncol=1)
kidney_Dialysis_10[17,] <- "no"
kidney_Dialysis_10[79,] <- "no"
dat <- cbind(dat,kidney_Dialysis,kidney_Dialysis_1,
             kidney_Dialysis_3,kidney_Dialysis_5,
             kidney_Dialysis_10)



## waiting time
## ---------------
dat$Waiting_list_date <- as.character(dat$Waiting_list_date)
dat$Waiting_list_date <- as.Date(dat$Waiting_list_date,format="%d/%m/%Y")
dat$LTx_date <- as.character(dat$LTx_date)
dat$LTx_date <- as.Date(dat$LTx_date,format="%d/%m/%Y")

Waiting_time <-  difftime(dat$LTx_date, dat$Waiting_list_date, units = "weeks") # weeks
Waiting_time <- as.numeric(Waiting_time)
dat <- cbind(dat,Waiting_time)

levels(dat$less_one) <- c("not work","work")
levels(dat$one_three) <- c("not work","work")
levels(dat$three_five) <- c("not work","work")
levels(dat$five_ten) <- c("not work","work")
levels(dat$longer_ten) <- c("not work","work")


#check the names and levels of the datset

dat$Education <- relevel(dat$Education,ref="Non-academic")
dat$Employment <- relevel(dat$Employment,ref="Non-employed")
dat$Relationship <- relevel(dat$Relationship,ref="not single")
levels(dat$Relationship) <- c("Not single","Single")
dat$Living <- relevel(dat$Living,ref="not alone")
levels(dat$Living) <- c("Not alone","Alone")
levels(dat$Sex) <- c("Female","Male")

write.csv(dat,"../data/shortformat.csv")

## Long Format Dataset
## ---------------

#1st choose the mean of the distance from LTx for longer than ten years period
end <- as.Date("01/12/2016","%m/%d/%Y")
duration <- difftime(rep(end,nrow(dat)), dat$LTx_date, units = "days")/365
mean((as.numeric(duration[(as.numeric(duration)>10)])),na.rm=TRUE)

dat1 <- reshape(dat,
                varying = c("less_one", "one_three", "three_five",
                            "five_ten","longer_ten"),
                v.names = "Workstatus",
                timevar = "Time",
                times = c(1,3,5,10,14),
```

```
                # new.row.names = 1:1000,
                direction = "long")
dat2 <- reshape(dat,
                varying = c("less_one_perc", "one_three_perc",
                            "three_five_perc","five_ten_perc",
                            "longer_ten_perc"),
                v.names = "Percentage",
                timevar = "Time",
                times = c(1,3,5,10,14),
                # new.row.names = 1:1000,
                direction = "long")
dat3 <- reshape(dat,
                varying = c("CLAD_1", "CLAD_3",
                            "CLAD_5","CLAD_10","CLAD"),
                v.names = "CLAD.l",
                timevar = "Time",
                times = c(1,3,5,10,14),
                # new.row.names = 1:1000,
                direction = "long")

dat4 <- reshape(dat,
                varying = c("Cancer_1", "Cancer_3",
                            "Cancer_5","Cancer_10","Cancer"),
                v.names = "Cancer.l",
                timevar = "Time",
                times = c(1,3,5,10,14),
                # new.row.names = 1:1000,
                direction = "long")

dat5 <- reshape(dat,
                varying = c("kidney_Dialysis_1", "kidney_Dialysis_3",
                            "kidney_Dialysis_5","kidney_Dialysis_10",
                            "kidney_Dialysis"),
                v.names = "Kidney_Dialysis.l",
                timevar = "Time",
                times = c(1,3,5,10,14),
                # new.row.names = 1:1000,
                direction = "long")
longdata<-cbind(dat1,Percentage=dat2$Percentage,
                CLAD=dat3$CLAD.l,Cancer=dat4$Cancer.l,
                Kidney_Dialysis=dat5$Kidney_Dialysis.l
)
longformat <-
  longdata[,c("ind","Sex","Education","Relationship","Living",
              "Employment","Age","BMI", "6MWD",
              "FEV1","Waiting_time","Time","Workstatus",
              "Percentage","CLAD","Cancer","Kidney_Dialysis")]




## ---------------
#From the comments after final presentation, use the exact time for patients for the final period of time
#1st still creat the the one with estimated percentage for the final period

end <- as.Date("12/01/2016","%m/%d/%Y")
duration <- difftime(rep(end,nrow(dat)), dat$LTx_date, units = "days")/365
```

```
duration <- as.numeric(duration)
sum(is.na(duration))#4 NA

dat.dura <- cbind(dat,duration)
patient_longerten <- dat.dura[which(duration>10),c("ind","duration")]
#28 patients who end at the time period 5.
patient_longerfive <- dat.dura[which((duration>5)&(duration<10)),c("ind","duration")]
#19 patients who end at the time period 4.
patient_longerthree <- dat.dura[which((duration>3)&(duration<5)),c("ind","duration")]
#15 patients who end at the time period 3
patient_longerone <- dat.dura[which((duration>1)&(duration<3)),c("ind","duration")]
#23 patients who end at the time period 2
patient_withinone <- dat.dura[which((duration<1)&(duration>0)),c("ind","duration")]
#6 patients who end at the time period 2

#the exact time for the final time period for each patient
longdata2 <- longformat
Num5 <- patient_longerten$ind+4*99
Num4 <- patient_longerfive$ind+3*99
Num3 <- patient_longerthree$ind+2*99
Num2 <- patient_longerone$ind+1*99
Num1 <- patient_withinone$ind
longdata2[Num5,]$Time <- patient_longerten$duration
longdata2[Num4,]$Time <- patient_longerfive$duration
longdata2[Num3,]$Time <- patient_longerthree$duration
longdata2[Num2,]$Time <- patient_longerone$duration


longformat <- longdata2


dat$X6MWD <- dat$`6MWD`

####### code for results showing in abstract
#results from the final part, for the abstract
final1_3  <- glm(one_three ~ Employment + Education,family=binomial,data=dat)
ci_emp <- formatCI(exp(confint(final1_3))[2,],text = "english")
ci_edu <- formatCI(exp(confint(final1_3))[3,],text = "english")

longformat$ind <- factor(longformat$ind)
fit_GLMM <- glmer(Workstatus ~ Employment + Education  + (1|ind) + CLAD + log(Time) ,
                data=longformat, family="binomial")
OR_GLMM <- exp(summary(fit_GLMM)$coefficients[,1])[-1]
test <- confint(fit_GLMM)
ci_GLMM <-formatCI(exp(test),text = "english")[-c(1:2)]

fit_LMM <-lme4::lmer(Percentage ~ Employment  + Education +
(1|ind) + Kidney_Dialysis +log(Time) ,
data=longformat)
coef_fit <- lmerTest::lmer(fit_LMM,data=longformat)
coef_LMM <- (summary(coef_fit)$coefficients[,1])[-1]
test <- confint(coef_fit)
ci_LMM <-formatCI((test),text = "english")[-c(1:3)]



####### code for Data Description
```

```
dat.ind <- dat[,c("Age","BMI","6MWD","FEV1","Waiting_time","Sex",
                  "Education","Relationship","Living", "Employment")]
colnames(dat.ind) <- c("Age","BMI","6MWD","Best FEV1 pred",
                       "Waiting time","Sex","Education",
                       "Relationship",
                       "Living", "Pre-employment")
Vars <- c("Age","BMI","6MWD","Best FEV1 pred","Waiting time",
          "Sex","Education","Relationship","Living", "Pre-employment")
tableOne <- CreateTableOne(vars = Vars, data = dat.ind)
tab1 <- print(tableOne, showAllLevels = TRUE,
              printToggle = FALSE, noSpaces = TRUE)
kable(tab1, format = "latex",booktabs = T)
mu <- ddply(dat, "Sex", summarise, grp.mean=mean(Age))
agesex_plot <- ggplot(dat, aes(x=Sex, y=Age,fill=Sex)) +
  geom_boxplot(alpha=0.4,outlier.colour="red", outlier.shape=8,
               outlier.size=6)+
  geom_dotplot(binaxis='y', stackdir='center',
               position=position_dodge(1))+
  theme(legend.position="bottom")

dat.choice <-  dat[-which(is.na(dat$Relationship)|
                            is.na(dat$Living)),]

df2 <- data.frame(table(dat.choice$Relationship,dat.choice$Living))
levels(dat.choice$Living) <- c("living alone","not alone")
colnames(df2) <- c("Relationship","Living","counts")

status2 <- ggplot(df2,aes(x=Living,y=counts, fill = Relationship)) +
    geom_bar(stat="identity", position=position_dodge()) +
  geom_text(aes(label=counts), vjust=1.6,
            position = position_dodge(0.9), size=5)+
  scale_fill_brewer(palette="Paired")+
  theme(legend.position="bottom",legend.title = element_blank())

ggarrange(agesex_plot,status2)
barfill <- "#4271AE"
barlines <- "#1F3552"

wait <- ggplot(dat,aes(x=Waiting_time))+
  geom_density(fill=barfill,alpha=0.5)+
  scale_x_continuous(name = "Waiting time in weeks",
                     breaks = seq(0, 200, 30),
                     limits=c(0, 200))+
  labs(x = "Waiting time in weeks",y="density")+
  theme(axis.title.y = element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())+
  geom_vline(xintercept = mean(dat$Waiting_time,na.rm=T),
           vline.data,col="red")

bmi <- ggplot(dat,aes(x=BMI))+
 geom_histogram(aes(y = ..density..),binwidth=2,
                colour = barlines, fill = barfill)+
  scale_x_continuous(name = "BMI",
                     breaks = seq(0, 35, 5),
                     limits=c(0, 35))+
  theme(axis.title.y = element_blank(),
```

```
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())+
  geom_vline(xintercept = median(dat$BMI,na.rm=T),
             vline.data,col="red")

sixMWD <- ggplot(dat,aes(x=X6MWD))+
  geom_histogram(aes(y = ..density..),binwidth=50,
               colour = barlines, fill = barfill)+
  scale_x_continuous(name = "6MWD(m)",
                     breaks = seq(0, 800, 150),
                     limits=c(0, 800))+
  theme(axis.title.y = element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())+
  geom_vline(xintercept = median(dat$X6MWD,na.rm=T),
             vline.data,col="red",na.rm=T)

FEV <- ggplot(dat,aes(x=FEV1))+
  geom_histogram(aes(y = ..density..),binwidth=10,
               colour = barlines, fill = barfill)+
  scale_x_continuous(name = "FEV1(%)",
                     breaks = seq(0, 150, 20),
                     limits=c(0, 150))+
  theme(axis.title.y = element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())+
  geom_vline(xintercept = median(dat$FEV1,na.rm=T),
             vline.data,col="red",na.rm=T)
ggarrange(wait,bmi,FEV,ncol=3,nrow=1)


CLAD <- cbind(sum(dat$CLAD_1=="yes",na.rm = T),
              sum(dat$CLAD_3=="yes",na.rm = T),
              sum(dat$CLAD_5=="yes",na.rm = T),
              sum(dat$CLAD_10=="yes",na.rm = T),
              sum(dat$CLAD=="yes",na.rm = T),
              sum(is.na(dat$CLAD)))

Cancer <- cbind(sum(dat$Cancer_1=="yes",na.rm = T),
                sum(dat$Cancer_3=="yes",na.rm = T),
                sum(dat$Cancer_5=="yes",na.rm = T),
                sum(dat$Cancer_10=="yes",na.rm = T),
                sum(dat$Cancer=="yes",na.rm = T),
                sum(is.na(dat$Cancer)))

KD <- cbind(sum(dat$kidney_Dialysis_1=="yes",na.rm = T),
            sum(dat$kidney_Dialysis_3=="yes",na.rm = T),
                      sum(dat$kidney_Dialysis_5=="yes",na.rm = T),
            sum(dat$kidney_Dialysis_10=="yes",na.rm = T),
                      sum(dat$kidney_Dialysis=="yes",na.rm = T),
            sum(is.na(dat$kidney_Dialysis)))

df.timedep <- rbind(CLAD,Cancer,KD)
rownames(df.timedep) <- c("CLAD","Cancer",
                          "Kidney-Dialysis")
colnames(df.timedep) <- c("1 year",
                        "3 years",
```

```
                        "5 years",
                        "10 years",
                        "End of Study",
                        "NA")
res.table<-xtable(df.timedep,
                  caption = 'Number of events with respect to time-dependent factors',
                        table.placement ="", label = "tab:tab2")
print(res.table, scalebox=0.8, caption.placement = "top")
dat.heat <- dat
df <- dat.heat[,c("CLAD_1","CLAD_3","CLAD_5","CLAD_10","CLAD",
             "Cancer_1","Cancer_3","Cancer_5","Cancer_10","Cancer",
             "kidney_Dialysis_1","kidney_Dialysis_3",
             "kidney_Dialysis_5","kidney_Dialysis_10","kidney_Dialysis")]
cases <- 1:ncol(df)
for(i in 1:ncol(df)){
  levels(df[,i]) <- c(0,1)
  df[,i]<-as.numeric(as.character(df[,i]))
  cases[i] <- sum(df[i],na.rm = T)
}
dat.slice <- slice(df, 1:nrow(dat),na.action=T)
col.cluster <- as.vector(c(rep("CLAD",5),
                           rep("Cancer",5),rep("Kidney-Dialysis",5)))
superheat(dat.slice,left.label = "none",
          yt = cases,heat.na.col = "white",
                                         yt.plot.type = "bar",
                                         yt.axis.name = "Number of Events",
                                         yt.obs.col =  c(rep("#fdcdac",5),
                                                                                                                   rep("#b3e2c
                                                                                                                   rep("#e5d8b

          heat.pal = c("#b35806", "white", "#542788"),
          legend = F,
          membership.cols = col.cluster,
          bottom.label.col = c("#b3e2cd","#fdcdac","#e5d8bd"),
          row.title.size = 6,
          row.title = "Patients",
          smooth.heat = F)

set.seed(10)
ind <- sample(x=nrow(dat),size=10,replace=FALSE)
v <- matrix(0,nrow=length(ind),ncol=6)
colnames(v) <- c("Pre LTx Workstates","less_one_perc",
                 "one_three_perc","three_five_perc",
                 "five_ten_perc","longer_ten_perc")
x <- matrix(1:ncol(v))
pre <- as.vector(dat[,"Employment"])
pre[which(pre[]=="Employed")] <- 100
pre[which(pre[]=="Non-employed")] <- 0
plot(spline(x,v[1,]),axes=FALSE,
     type="n",ylim=c(0,100),xlim=c(1,6),
     ylab="Work Percentage(%)",
                 xlab="Time Period after LTx")
for(i in 1:length(ind)){
  v[i,] <- c(pre[ind[i]],dat[ind[i],"less_one_perc"],
             dat[ind[i],"one_three_perc"],dat[ind[i],"three_five_perc"],
             dat[ind[i],"five_ten_perc"],dat[ind[i],"longer_ten_perc"])
  lines(x,v[i,],col=i*5,type="o")
  }
```

```
legend("bottomright", "Individual examples", pch=1,
       lty=1,inset=c(0,1), xpd=TRUE, horiz=TRUE,
       bty="n",cex = 1.25)


y <- c(0,20,40,60,80,100)
axis(side=1,at=x,col=1,font=3,lty=1,cex=25,
     labels=c("pre LTx","<1","1-3","3-5","5-10",">10"))
axis(side=2,at=y,col=1,font=3,lty=1,cex=25,
     labels=c("0","20","40","60","80","100"))
#title("Work percentage from a small sample(n=10)", line = 2.3)
percent <- dat[,c("less_one_perc","one_three_perc",
                  "three_five_perc","five_ten_perc","longer_ten_perc")]
colnames(percent) <-  c("Period 1 (<1 year)",
                        "Period 2 (1 to 3 years)",
                        "Period 3 (3 to 5 years)",
                        "Period 4 (5 to 10 years)",
                        "Period 5 (> 10 years)")
percent %>%
  tidyr::gather("Time", "value",1:5) %>%
  ggplot(., aes(x = Time, y = value,col=Time))+
  labs(y = "Work percentage(%)",x="Time period after LTx") +
        geom_jitter(width = 0.3, height = 0.5,size=2,na.rm=TRUE,shape=1)+
  geom_boxplot(alpha=0.2)+
        theme(line = element_blank())+
        theme(panel.grid.major = element_blank(),
                                panel.grid.minor = element_blank(),
panel.background = element_blank(),
axis.line = element_line(colour = "black"))+
  theme(legend.position="right",
                        legend.text=element_text(size=8),
                        legend.title =element_text(size=15),
                        axis.text.x = element_blank())+
        scale_colour_hue(breaks= c("Period 1 (<1 year)",
                       "Period 2 (1 to 3 years)","Period 3 (3 to 5 years)",
                       "Period 4 (5 to 10 years)","Period 5 (> 10 years)"))+
        theme(plot.title = element_text(size = 16, face = "bold"))

####### code for results: first research question
#the number of patients of working, NA and not working each time point
timep1 <- as.factor(ifelse(is.na(dat$less_one),
                      "NA", dat$less_one))
levels(timep1) <- c("Not work","Work","NA")
timep1 <- factor(timep1, levels = c("Not work",
                          "NA", "Work"))
timep2 <- as.factor(ifelse(is.na(dat$one_three),
                      "NA", dat$one_three))
levels(timep2) <- c("Not work","Work","NA")
timep2 <- factor(timep2, levels = c("Not work", "NA", "Work"))
timep3 <- as.factor(ifelse(is.na(dat$three_five),
                      "NA", dat$three_five))
levels(timep3) <- c("Not work","Work","NA")
timep3 <- factor(timep3, levels = c("Not work", "NA", "Work"))
timep4  <- as.factor(ifelse(is.na(dat$five_ten),
                       "NA", dat$five_ten))
levels(timep4) <- c("Not work","Work","NA")
```

```
timep4 <- factor(timep4, levels = c("Not work", "NA", "Work"))
timep5 <- as.factor(ifelse(is.na(dat$longer_ten),
                    "NA", dat$longer_ten))
levels(timep5) <- c("Not work","Work","NA")
timep5 <- factor(timep5, levels = c("Not work", "NA", "Work"))
likedat <- data.frame(timep1,timep2,timep3,timep4,timep5)
colnames(likedat) <- c("Period 1 (<1 year)",
                       "Period 2 (1 to 3 years)","Period 3 (3 to 5 years)",
                       "Period 4 (5 to 10 years)","Period 5 (> 10 years)")
ldatdat <- likert(likedat)
plot(ldatdat,order=F)
col1 <- as.matrix(table(dat$less_one,exclude = NULL))
col2 <- as.matrix(table(dat$one_three,exclude = NULL))
col3 <- as.matrix(table(dat$three_five,exclude = NULL))
col4 <- as.matrix(table(dat$five_ten,exclude = NULL))
col5 <- as.matrix(table(dat$longer_ten,exclude = NULL))

mat <- cbind(col1,col2,col3,col4,col5)
Total <- cbind(nrow(dat[-which(is.na(dat$less_one)),]),
               nrow(dat[-which(is.na(dat$one_three)),]),
               nrow(dat[-which(is.na(dat$three_five)),]),
               nrow(dat[-which(is.na(dat$five_ten)),]),
               nrow(dat[-which(is.na(dat$longer_ten)),]))
mat <- rbind(mat,Total)
colnames(mat) <- c("Period 1","Period 2","Period 3",
                   "Period 4","Period 5")
rownames(mat) <- c("Work","Not work","NA","Total")
res.table<-xtable(mat, caption = 'Post-LTx work status',
                  table.placement ="",label="tab:tab3")
print(res.table, scalebox=1.0, caption.placement = "top")

####### code for Results
dat.choice1 <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                           | is.na(dat$Education) | is.na(dat$Relationship)
                           | is.na(dat$Living) | is.na(dat$Age)
                           | is.na(dat$BMI) | is.na(dat$X6MWD)
                           | is.na(dat$FEV1) | is.na(dat$less_one)
                           | is.na(dat$Waiting_time)),]

name_var=c("Age","FEV1","BMI","Education","Living",
           "Employment","Relationship","Sex","X6MWD",
           "Waiting_time","less_one","one_three",
           "three_five","five_ten","longer_ten")
bic <- rep(NA,10)
AUC <- rep(NA,10)
fml <- rep(NA,10)
for(i in 1:10){
  fml[i] = paste(name_var[11],"~",name_var[i],sep="")
  reg=glm(fml[i],family=binomial,data=dat.choice1)
  pred <- predict(reg,type="response")
  bic[i] <- round(BIC(reg),2)
  roc <- pROC::roc(dat.choice1$less_one ~ pred)
  AUC[i] <- round(pROC::auc(roc),2)
}
auc <- round(AUC,digits = 2)
bic <- round(bic,digits = 2)
pv <- rep(NA,10)
```

```
df1 <- matrix(0,nrow=10,ncol=3)
for(i in 1:10){
  fml[i] = paste(name_var[11],"~",name_var[i],sep="")
  reg=glm(fml[i],family=binomial,data=dat)
  pv[i]<-formatPval(summary(reg)$coefficients[2,4])
  df1[i,] <- c(round(exp(coef(reg)[2]),digits = 2),
               round(exp(confint(reg)[2,]),digits = 2))
}

df <- data.frame(df1[,1],formatCI(df1[,c(2:3)],text = "english"),pv,auc,bic)

tab_colname <- c("Odds Ratio","$95\\%$-confidence interval","$p$-value","AUC","BIC")
colnames(df) <- tab_colname
tab_rowname <- c("Age","Best FEV1(\\%)","BMI","Education(Academic)",
                 "Living(Alone)","Pre-employment(Employed)",
                 "Relationship(Single)","Sex(Male)","6MWD","Waiting time")
rownames(df) <- tab_rowname

#df[,1] <- as.numeric(df[,1])
df_tab <-xtable(df, caption = "Univariate models for work status
                within one year after LTx")
print(df_tab, scalebox=0.8,
      caption.placement = "top",
      sanitize.text.function = function(x) {x})
reg0=glm(less_one ~ Employment,family=binomial,data=dat.choice1)

reg1=glm(less_one ~ Education+ Living+Employment,family=binomial,data=dat.choice1)

reg2=glm(less_one ~ Living+Employment,family=binomial,data=dat.choice1)

reg3=glm(less_one ~ Education+Employment,family=binomial,data=dat.choice1)

tab1 <- t(data.frame(BIC(reg1),BIC(reg3),BIC(reg2),BIC(reg0)))
rownames(tab1) <- c("Pre-employment+Living+Education",
                    "Pre-employment+Education","Pre-employment+Living",
                    "Pre-employment")
colnames(tab1) <- c("BIC")
tab1_tab <- xtable(tab1,digits = 2,
                   caption = "Model selection for
                   work status within one year after LTx",label="tab:tab5")
print(tab1_tab, scalebox=0.8, caption.placement = "top")
final1  <- glm(less_one ~ Employment,family=binomial,data=dat)
tableRegression(final1,
                row.nam = "Pre-employment(Employed)",
                caption="Selected model for work status
                less than one year after LTx",
                caption.placement="top",label="tab:tab6")

dat.choice2 <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                             | is.na(dat$Education) | is.na(dat$Relationship)
                             | is.na(dat$Living) | is.na(dat$Age)
                             | is.na(dat$BMI) | is.na(dat$X6MWD)
                             | is.na(dat$FEV1) | is.na(dat$one_three)
                             | is.na(dat$Waiting_time)),]
for(i in 1:10){
  fml[i] = paste(name_var[12],"~",name_var[i],sep="")
  reg=glm(fml[i],family=binomial,data=dat.choice2)
```

```
  pred <- predict(reg,type="response")
  bic[i] <- round(BIC(reg),4)
  roc <- pROC::roc(dat.choice2$one_three ~ pred)
  AUC[i] <- round(pROC::auc(roc),4)
}
auc <- round(AUC,digits = 2)
bic <- round(bic,digits = 2)

#univariate models
  df2 <- matrix(0,nrow=10,ncol=3)
for(i in 1:10){
  fml[i] = paste(name_var[12],"~",name_var[i],sep="")
  reg=glm(fml[i],family=binomial,data=dat)
  pv[i]<-formatPval(summary(reg)$coefficients[2,4])
  df2[i,] <- c(exp(coef(reg)[2]),exp(confint(reg)[2,]))
}
rownames(df2) <- name_var[1:10]
colnames(df2)<-c("OR","lower","upper")

df <- data.frame(round(df2[,1],digits=2),
                 formatCI(df2[,c(2:3)],text = "english"),
                 pv,auc,bic)
colnames(df)<- tab_colname
rownames(df) <- tab_rowname
df_tab <- xtable(df,digits = NULL,
                 caption="Univariate models for work status one to three years after LTx")
print(df_tab, scalebox=0.8, caption.placement = "top",
      sanitize.text.function = function(x) {x})
reg1=glm(one_three ~ Education+ Employment,family=binomial,data=dat.choice2)
reg0=glm(one_three ~ Employment,family=binomial,data=dat.choice2)

tab1 <- t(data.frame(BIC(reg0),BIC(reg1)))
rownames(tab1) <- c("Pre-employment","Pre-employment+Education")
colnames(tab1) <- c("BIC")
tab1_tab <- xtable(tab1,digits = 2,
                   caption = "Model selection for work status one to three years after LTx")
print(tab1_tab, scalebox=0.8, caption.placement = "top")
final2  <- glm(one_three ~ Employment + Education,family=binomial,data=dat)
tableRegression(final2,caption="Selected model for work status one to three years after LTx",
                caption.placement="top",
                row.nam = c("Pre-employment(Employed)","Education(Academic)"),label="tab:tab9")

dat.select <- dat[-which(is.na(dat$Education)|
                           is.na(dat$three_five)),]
edu3to5_tab <- table(dat.select$Education,dat.select$three_five)
edu3to5_tab<- as.numeric(edu3to5_tab)

names(edu3to5_tab) <- c("nn","notwork","notedu","yy")

twoby1 <- ggplot(dat.select, aes(x=Education,y=three_five,col=Education))+
  geom_jitter(width = 0.3, height = 0.3)+
  theme(
    axis.title.x=element_blank(),
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank())+
  theme(legend.position="bottom",
        legend.text = element_text(size = 10),
```

```
        legend.title=element_blank())+
  labs(y="Post work status")+
  geom_text(x=2, y=0.5, label=edu3to5_tab["notwork"],size=5,col=1)+
  geom_text(x=2, y=2.5, label=edu3to5_tab["yy"],size=5,col=1)+
  geom_text(x=1, y=0.5, label=edu3to5_tab["nn"],size=5,col=1)+
  geom_text(x=1, y=2.5, label=edu3to5_tab["notedu"],size=5,col=1)



dat.select2 <- dat[-which(is.na(dat$Employment)|
                            is.na(dat$three_five)),]

emp3to5_tab <- table(dat.select2$Employment,
                     dat.select2$three_five)
emp3to5_tab<- as.numeric(emp3to5_tab)

names(emp3to5_tab) <- c("nn","notwork","notemp","yy")


twoby2 <- ggplot(dat.select2 ,labels=FALSE,
                 aes(x=Employment,y=three_five,col=Employment))+
  geom_jitter(width = 0.3, height = 0.3)+
  theme(
    axis.title.x=element_blank(),
    axis.title.y=element_blank(),
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank(),
  )+
    geom_text(x=2, y=0.5, label=emp3to5_tab["notwork"],size=5,col=1)+
  geom_text(x=2, y=2.5, label=emp3to5_tab["yy"],size=5,col=1)+
  geom_text(x=1, y=0.5, label=emp3to5_tab["nn"],size=5,col=1)+
  geom_text(x=1, y=2.5, label=emp3to5_tab["notemp"],size=5,col=1)+

  theme(legend.position="bottom",
        legend.text = element_text(size = 10),
        legend.title=element_blank())
table(dat.select2$Employment,dat.select$three_five)
ggarrange(twoby1,twoby2)
bic.3 <- rep(NA,8)
AUC.3 <- rep(NA,8)
fml.3 <- rep(NA,8)
pv.3 <- rep(NA,8)
dat.choice4 <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                           | is.na(dat$Education) | is.na(dat$Relationship)
                           | is.na(dat$Living) | is.na(dat$Age)
                           | is.na(dat$BMI) | is.na(dat$X6MWD)
                           | is.na(dat$FEV1) | is.na(dat$three_five)
                           | is.na(dat$Waiting_time)),]

name_var_long5=c("Age","FEV1","BMI","Living",
           "Relationship","Sex","X6MWD",
           "Waiting_time","less_one","one_three",
           "three_five","five_ten","longer_ten")
for(i in 1:8){
  fml.3[i] = paste(name_var_long5[11],"~",name_var_long5[i],sep="")
  reg=glm(fml.3[i],family=binomial,data=dat.choice4)
  pred <- predict(reg,type="response")
```

```
  bic.3[i] <- round(BIC(reg),4)
  roc.3 <- pROC::roc(dat.choice4$longer_ten ~ pred)
  AUC.3[i] <- round(pROC::auc(roc.3),4)
}
auc.3 <- round(AUC.3,digits = 2)
bic.3 <- round(bic.3,digits = 2)
#univariates models
df <- matrix(0,nrow=8,ncol=3)
for(i in 1:8){
  fml.3[i] = paste(name_var_long5[11],"~",name_var_long5[i],sep="")
  reg=glm(fml.3[i],family=binomial,data=dat)
  pv.3[i]<-formatPval(summary(reg)$coefficients[2,4])
  df[i,] <- c(exp(coef(reg)[2]),exp(confint(reg)[2,]))
}

df <- data.frame(round(df[,1],digits=2),formatCI(df[,c(2:3)],text = "english"),pv.3,auc.3,bic.3)

tab_rowname_long5 <- c("Age", "Best FEV1(\\%)","BMI",
                           "Living(Alone)",
                           "Relationship(Single)",
                           "Sex(Male)",
                           "6MWD","Waiting time"  )
colnames(df) <- tab_colname
rownames(df) <- tab_rowname_long5
df_tab <- xtable(df,digits = 2,
                 caption="Univariate models for work status three to five years after LTx",
                 label="tab:tab10")
print(df_tab, scalebox=0.8,
      caption.placement = "top",
      sanitize.text.function = function(x) {x})

dat.choice3 <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                           | is.na(dat$Education) | is.na(dat$Relationship)
                           | is.na(dat$Living) | is.na(dat$Age)
                           | is.na(dat$BMI) | is.na(dat$X6MWD)
                           | is.na(dat$FEV1) | is.na(dat$five_ten)
                           | is.na(dat$Waiting_time)),]

#dim(dat.choice3)
#table(dat$five_ten)
#table(dat.choice3$five_ten)

for(i in 1:10){
  fml[i] = paste(name_var[14],"~",name_var[i],sep="")
  reg=glm(fml[i],family=binomial,data=dat.choice3)
  pred <- predict(reg,type="response")
  bic[i] <- round(BIC(reg),4)
  roc <- pROC::roc(dat.choice3$five_ten ~ pred)
  AUC[i] <- round(pROC::auc(roc),4)
}
auc <- round(AUC,digits = 2)
bic <- round(bic,digits = 2)
#univariates models
df <- matrix(0,nrow=10,ncol=3)
for(i in 1:10){
  fml[i] = paste(name_var[14],"~",name_var[i],sep="")
  reg=glm(fml[i],family=binomial,data=dat)
```

```
  pv[i]<-formatPval(summary(reg)$coefficients[2,4])
  df[i,] <- c(exp(coef(reg)[2]),exp(confint(reg)[2,]))
}

df <- data.frame(round(df[,1],digits=2),formatCI(df[,c(2:3)],text = "english"),pv,auc,bic)
colnames(df)<- tab_colname
rownames(df) <- tab_rowname
df_tab <- xtable(df,digits = 2,
                 caption="Univariate models for work status five to ten years after LTx")
print(df_tab, scalebox=0.8,
      caption.placement = "top",
      sanitize.text.function = function(x) {x})
reg1=glm(five_ten ~ Education+ Employment,family=binomial,data=dat.choice2)
reg0=glm(five_ten ~ Employment,family=binomial,data=dat.choice2)
tab1 <- t(data.frame(BIC(reg0),BIC(reg1)))
rownames(tab1) <- c("Pre-employment","Pre-employment+Education")
colnames(tab1) <- c("BIC")
tab1_tab <- xtable(tab1,digits = 2,
                   caption = "Model selection for
                   work status five to ten years after LTx",
                   label = "tab:tab11")
print(tab1_tab, scalebox=0.8, caption.placement = "top")
final3 <- glm(five_ten ~ Employment+ Education,family=binomial,data=dat)
tableRegression(final3,
                stats = c("exp.estimate",  "ci.95", "p.value"),
                caption="Selected model for work status
                five to ten years after LTx",
                caption.placement="top",
                row.nam = c("Pre-employment(Employed)","Education(Academic)"),label = "tab:tab12")

dat.select <- dat[-which(is.na(dat$Education)|
                           is.na(dat$longer_ten)),]
edu10_tab <- table(dat.select$Education,
                     dat.select$longer_ten)
edu10_tab<- as.numeric(edu10_tab)

names(edu10_tab) <- c("nn","notwork","notedu","yy")


twoby1 <- ggplot(dat.select, aes(x=Education,y=longer_ten,col=Education))+
  #scale_color_manual(values=rhg_cols1)+
  # geom_point()+
  geom_jitter(width = 0.3, height = 0.3)+
  theme(
    axis.title.x=element_blank(),
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank())+
  theme(legend.position="bottom",
        legend.text = element_text(size = 10),legend.title=element_blank())+
  labs(y="Post work status")+
  geom_text(x=2, y=0.5, label=edu10_tab["notwork"],size=5,col=1)+
  geom_text(x=2, y=2.5, label=edu10_tab["yy"],size=5,col=1)+
  geom_text(x=1, y=0.5, label=edu10_tab["nn"],size=5,col=1)+
  geom_text(x=1, y=2.5, label=edu10_tab["notedu"],size=5,col=1)


dat.select2 <- dat[-which(is.na(dat$Employment)|
```

```
                            is.na(dat$longer_ten)),]

emp10_tab <- table(dat.select2$Employment,
                     dat.select2$longer_ten)
emp10_tab<- as.numeric(emp10_tab)

names(emp10_tab) <- c("nn","notwork","notemp","yy")

twoby2 <- ggplot(dat.select2 ,labels=FALSE,
                  aes(x=Employment,y=longer_ten,col=Employment))+
  # geom_point()+
  geom_jitter(width = 0.3, height = 0.3)+
  theme(
    axis.title.x=element_blank(),
    axis.title.y=element_blank(),
    #   axis.ticks.y=element_blank(),
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank(),
    #   axis.text.y=element_blank(),
  )+
  #  labs(y="Post work status")+
   geom_text(x=2, y=0.5, label=emp10_tab["notwork"],size=5,col=1)+
  geom_text(x=2, y=2.5, label=emp10_tab["yy"],size=5,col=1)+
  geom_text(x=1, y=0.5, label=emp10_tab["nn"],size=5,col=1)+
  geom_text(x=1, y=2.5, label=emp10_tab["notemp"],size=5,col=1)+
  theme(legend.position="bottom",
        legend.text = element_text(size = 10),
        legend.title=element_blank())
#table(dat.select2$Employment,dat.select$longer_ten)
ggarrange(twoby1,twoby2)
bic.5 <- rep(NA,8)
AUC.5 <- rep(NA,8)
fml.5 <- rep(NA,8)
pv.5 <- rep(NA,8)
dat.choice5 <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                           | is.na(dat$Education) | is.na(dat$Relationship)
                           | is.na(dat$Living) | is.na(dat$Age)
                           | is.na(dat$BMI) | is.na(dat$X6MWD)
                           | is.na(dat$FEV1) | is.na(dat$longer_ten)
                           | is.na(dat$Waiting_time)),]

name_var_long10=c("Age","FEV1","BMI","Living",
           "Relationship","Sex","X6MWD",
           "Waiting_time","less_one","one_three",
           "three_five","five_ten","longer_ten")
for(i in 1:8){
  fml.5[i] = paste(name_var_long10[13],"~",name_var_long10[i],sep="")
  reg=glm(fml.5[i],family=binomial,data=dat.choice5)
  pred <- predict(reg,type="response")
  bic.5[i] <- round(BIC(reg),4)
  roc.5 <- pROC::roc(dat.choice5$longer_ten ~ pred)
  AUC.5[i] <- round(pROC::auc(roc.5),4)
}
auc.5 <- round(AUC.5,digits = 2)
bic.5 <- round(bic.5,digits = 2)
#univariates models
df <- matrix(0,nrow=8,ncol=3)
```

```
for(i in 1:8){
  fml.5[i] = paste(name_var_long10[13],"~",name_var_long10[i],sep="")
  reg=glm(fml.5[i],family=binomial,data=dat)
  pv.5[i]<-formatPval(summary(reg)$coefficients[2,4])
  df[i,] <- c(exp(coef(reg)[2]),exp(confint(reg)[2,]))
}

df <- data.frame(round(df[,1],digits=2),formatCI(df[,c(2:3)],text = "english"),pv.5,auc.5,bic.5)

tab_rowname_long10 <- c("Age", "Best FEV1(\\%)","BMI",
                          "Living(Alone)",
                          "Relationship(Single)",
                          "Sex(Male)",
                          "6MWD","Waiting time"  )
colnames(df) <- tab_colname
rownames(df) <- tab_rowname_long10
df_tab <- xtable(df,digits = 2,
                 caption="Univariate models for work status
                 longer than ten years after LTx",
                 label="tab:tabadd")
print(df_tab, scalebox=0.8,
      caption.placement = "top",
      sanitize.text.function = function(x) {x})

####### code for the result of logistic regression
####### for all of the seprate time periods
fit_1 <- glm(less_one ~ Employment,family=binomial,data=dat)
fit_2 <- glm(one_three ~ Employment +Education,family=binomial,data=dat)
fit_4 <- glm(five_ten ~ Employment +Education,family=binomial,data=dat)

OR1 <- rbind(
              sprintf("%.2f",exp(coef(fit_1)[2])),
              sprintf("%.2f",exp(coef(fit_2)[2])),
              sprintf("%.2f",exp(coef(fit_4)[2])) )

CI1 <- rbind(formatCI(exp(confint(fit_1)[2,]),text = "english"),
             formatCI(exp(confint(fit_2)[2,]),text = "english"),
             formatCI(exp(confint(fit_4)[2,]),text = "english"))

OR2 <- rbind(
  sprintf("%.2f",exp(coef(fit_2)[3])),
  sprintf("%.2f",exp(coef(fit_4)[3])))

OR2 <- rbind("-",OR2)

CI2 <- rbind(formatCI(exp(confint(fit_2)[3,]),text = "english"),
             formatCI(exp(confint(fit_4)[3,]),text = "english"))
CI2 <- rbind("-",CI2)

p1 <- rbind(
  biostatUZH::formatPval(summary(fit_1)$coefficients[2,4]),
  biostatUZH::formatPval(summary(fit_2)$coefficients[2,4]),
  biostatUZH::formatPval(summary(fit_4)$coefficients[2,4]))
p2<- rbind(
  "-",
  biostatUZH::formatPval(summary(fit_2)$coefficients[3,4]),
  biostatUZH::formatPval(summary(fit_4)$coefficients[3,4]))
```

```r
p <- rbind(p1,p2)

mat_all_logistic <- cbind(rbind(OR1,OR2),rbind(CI1,CI2),p)
colnames(mat_all_logistic) <- c("Odds Ratio",
                                "$95\\%$-confidence interval",
                                "$p$-value")
rownames(mat_all_logistic) <- c("Pre-employment(<1)",
                                "Pre-employment(1-3)",
                                "Pre-employment(5-10)",
                                "","Education(3-5)",
                                "Education(5-10)")

log_Tab <- xtable(mat_all_logistic,
                  caption="Result for logistic regression
                  for all possible periods after LTx",
                  label="tab:tab13")
print(log_Tab, scalebox=0.8, caption.placement = "top",sanitize.text.function = function(x) {x})

dat.choice.a  <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                          | is.na(dat$Education) | is.na(dat$Relationship)
                          | is.na(dat$Living) | is.na(dat$Age)
                          | is.na(dat$BMI) | is.na(dat$X6MWD)
                          | is.na(dat$FEV1) | is.na(dat$less_one_perc)
                          | is.na(dat$Waiting_time)),]

dim(dat.choice.a)
name_var2 <- c("Age","FEV1","BMI","Education","Living",
               "Employment","Relationship","Sex","X6MWD",
               "Waiting_time","less_one_perc","one_three_perc",
               "three_five_perc","five_ten_perc","longer_ten_perc")
bic <- rep(NA,10)
pv <- rep(NA,10)
for(i in 1:10){
  fml[i] = paste(name_var2[11],"~",name_var2[i],sep="")
  reg=lm(fml[i], data=dat.choice.a)
  pv[i]<-formatPval(summary(reg)$coefficients[2,4])
  bic[i] <- round(BIC(reg),4)
}
bic <- round(bic,digits = 2)

df <- matrix(0,nrow=10,ncol=3)
for(i in 1:10){
  fml[i] = paste(name_var2[11],"~",name_var2[i],sep="")
  reg=lm(fml[i],data=dat)
df[i,] <- c( (coef(reg)[2]), (confint(reg)[2,]))
}
df <- data.frame(round(df[,1],digits=2),formatCI(df[,c(2:3)],text = "english"),pv,bic)
tab_colname2 <- c("Coefficient", "$95\\%$-confidence interval", "$p$-value", "BIC")
colnames(df)<- tab_colname2
rownames(df) <- c("Age","Best FEV1(\\%)","BMI","Education(Academic)",
                  "Living(Alone)","Pre-employment(Employed)",
                  "Relationship(Single)","Sex(Male)","6MWD","Waiting time")
res.table<-xtable(df,
                  caption = 'Univariate models for work percentage
                  less than one year after LTx',
                  table.placement ="",
                  label="tab:tab14")
```

```
print(res.table, scalebox=0.8,
      caption.placement = "top",
      sanitize.text.function = function(x) {x})
reg0=lm(less_one_perc ~ Employment,data=dat.choice.a)

reg1=lm(less_one_perc ~ Education+ Living+Employment,data=dat.choice.a)

reg2=lm(less_one_perc ~ Living+Employment,data=dat.choice.a)

reg3=lm(less_one_perc ~ Education+Employment,data=dat.choice.a)

tab1 <- t(data.frame(BIC(reg1),BIC(reg3),BIC(reg2),BIC(reg0)))
rownames(tab1) <- c("Pre-employment+Living+Education",
                    "Pre-employment+Education","Pre-employment+Living",
                    "Pre-employment")
colnames(tab1) <- c("BIC")
tab1 <- xtable(tab1,
               caption = "Model selection for
               work percentage less than one year after LTx",
               label="tab:tab15")
print(tab1, scalebox=0.8, caption.placement = "top")
final.1  <- lm(less_one_perc ~ Employment,data=dat)
tableRegression(final.1,intercept = F,
                caption="Selected model for work percentage
                one to three years after LTx",
                caption.placement="top",
                row.nam = "Pre-employment(Employed)")

dat.choice.b  <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                          | is.na(dat$Education) | is.na(dat$Relationship)
                          | is.na(dat$Living) | is.na(dat$Age)
                          | is.na(dat$BMI) | is.na(dat$X6MWD)
                          | is.na(dat$FEV1) | is.na(dat$one_three_perc)
                          | is.na(dat$Waiting_time)),]

for(i in 1:10){
  fml[i] = paste(name_var2[12],"~",name_var2[i],sep="")
  reg=lm(fml[i], data=dat.choice.b)
  pred <- predict(reg,type="response")
  bic[i] <- round(BIC(reg),4)
}
bic <- round(bic,digits = 2)

df <- matrix(0,nrow=10,ncol=3)
for(i in 1:10){
  fml[i] = paste(name_var2[12],"~",name_var2[i],sep="")
  reg=lm(fml[i],data=dat)
  pv[i]<-formatPval(summary(reg)$coefficients[2,4])
df[i,] <- c( (coef(reg)[2]), (confint(reg)[2,]))
}
df <- data.frame(round(df[,1],digits=2),formatCI(df[,c(2:3)],text = "english"),pv,bic)
colnames(df)<- tab_colname2
rownames(df) <- c("Age","Best FEV1(\\%)","BMI","Education(Academic)",
                  "Living(Alone)","Pre-employment(Employed)",
                  "Relationship(Single)",
                  "Sex(Male)","6MWD","Waiting time")
res.table<-xtable(df, table.placement ="",
```

```
                  caption =
                    'Univariate models for work percentage
                  one to three years after LTx')
print(res.table, scalebox=0.8,
      caption.placement = "top",
      sanitize.text.function = function(x) {x})
reg0=lm(one_three_perc ~ Employment,data=dat.choice.b)

reg1=lm(one_three_perc ~ Education+ Living+Employment,data=dat.choice.b)

reg2=lm(one_three_perc ~ Living+Employment,data=dat.choice.b)

reg3=lm(one_three_perc ~ Education+Employment,data=dat.choice.b)

tab1 <- t(data.frame(BIC(reg3),BIC(reg1),BIC(reg2),BIC(reg0)))
rownames(tab1) <- c("Pre-employment+Education",
                    "Pre-employment+Living+Education",
                    "Pre-employment+Living","Pre-employment")
colnames(tab1) <- c("BIC")
tab1 <- xtable(tab1,
               caption =
                 "Model selection for work percentage
               one to three years after LTx")
print(tab1, scalebox=0.8, caption.placement = "top")
final.2  <- lm(one_three_perc ~ Employment,data=dat)
tableRegression(final.2,intercept = F,
                caption=
                  "Selected model for work percentage one to three years after LTx",
                caption.placement="top",row.nam = "Pre-employment(Employed)",
                label="tab:tab19")

dat.choice.c <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                          | is.na(dat$Education) | is.na(dat$Relationship)
                          | is.na(dat$Living) | is.na(dat$Age)
                          | is.na(dat$BMI) | is.na(dat$X6MWD)
                          | is.na(dat$FEV1) | is.na(dat$three_five_perc)
                          | is.na(dat$Waiting_time)),]

for(i in 1:10){
  fml[i] = paste(name_var2[13],"~",name_var2[i],sep="")
  reg=lm(fml[i], data=dat.choice.c)
  pred <- predict(reg,type="response")
  bic[i] <- round(BIC(reg),4)
}
bic <- sprintf("%.2f",bic)

df <- matrix(0,nrow=10,ncol=3)
for(i in 1:10){
  fml[i] = paste(name_var2[13],"~",name_var2[i],sep="")
  reg=lm(fml[i],data=dat)
  pv[i]<-formatPval(summary(reg)$coefficients[2,4])
df[i,] <- c( (coef(reg)[2]), (confint(reg)[2,]))
}
df <- cbind(sprintf("%.2f",df[,1]),formatCI(df[,c(2:3)],text = "english"),pv,bic)
colnames(df)<- tab_colname2
rownames(df) <- c("Age","Best FEV1(\\%)","BMI","Education(Academic)",
                  "Living(Alone)","Pre-employment(Employed)",
```

```
                  "Relationship(Single)",
                  "Sex(Male)","6MWD","Waiting time")
res.table<-xtable(df, table.placement ="",
                  caption =
                    'Univariate models for work percentage
                  three to five years after LTx')
print(res.table,
      scalebox=0.8,
      caption.placement = "top",sanitize.text.function = function(x) {x})
reg0=lm(three_five_perc ~ Employment,data=dat.choice.c)

reg1=lm(three_five_perc ~ Education+ Living+Employment,data=dat.choice.c)

reg2=lm(three_five_perc ~ Living+Employment,data=dat.choice.c)

reg3=lm(three_five_perc ~ Education+Employment,data=dat.choice.c)

tab1 <- t(data.frame(BIC(reg1),BIC(reg3),BIC(reg2),BIC(reg0)))
rownames(tab1) <- c("Pre-employment+Living+Education",
                    "Pre-employment+Education",
                    "Pre-employment+Living","Pre-employment")
colnames(tab1) <- c("BIC")
tab1 <- xtable(tab1,caption =
                 "Model selection for work percentage three to five years after LTx")
print(tab1, scalebox=0.8, caption.placement = "top")
final.3  <- lm(three_five_perc ~ Employment,data=dat)
tableRegression(final.3,intercept = F,caption=
                  "Selected model for work status three to five years after LTx",
                caption.placement="top",row.nam = "Pre-employment(Employed)")

dat.choice.d <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                          | is.na(dat$Education) | is.na(dat$Relationship)
                          | is.na(dat$Living) | is.na(dat$Age)
                          | is.na(dat$BMI) | is.na(dat$X6MWD)
                          | is.na(dat$FEV1) | is.na(dat$five_ten_perc)
                          | is.na(dat$Waiting_time)),]

for(i in 1:10){
  fml[i] = paste(name_var2[14],"~",name_var2[i],sep="")
  reg=lm(fml[i], data=dat.choice.d)
  pred <- predict(reg,type="response")
  bic[i] <- round(BIC(reg),2)
}

df <- matrix(0,nrow=10,ncol=3)
for(i in 1:10){
  fml[i] = paste(name_var2[14],"~",name_var2[i],sep="")
  reg=lm(fml[i],data=dat)
  pv[i]<-formatPval(summary(reg)$coefficients[2,4])
df[i,] <- c( (coef(reg)[2]), (confint(reg)[2,]))
}
df <- data.frame(round(df[,1],digits=2),formatCI(df[,c(2:3)],text = "english"),pv,bic)
colnames(df)<-tab_colname2
rownames(df) <- c("Age","Best FEV1(\\%)","BMI","Education(Academic)",
                  "Living(Alone)","Pre-employment(Employed)","Relationship(Single)",
                  "Sex(Male)","6MWD","Waiting time")
res.table<-xtable(df, caption =
```

```
                    'Univariate models for work percentage
                  five to ten years after LTx',
                  table.placement ="")
print(res.table, scalebox=0.8,
      caption.placement = "top",
      sanitize.text.function = function(x) {x})
reg0=lm(five_ten_perc ~ Employment,data=dat.choice.d)

reg1=lm(five_ten_perc ~ Education+ Living+Employment,data=dat.choice.d)

reg2=lm(five_ten_perc ~ Living+Employment,data=dat.choice.d)

reg3=lm(five_ten_perc ~ Education+Employment,data=dat.choice.d)

tab1 <- t(data.frame(BIC(reg1),BIC(reg3),BIC(reg0),BIC(reg2)))
rownames(tab1) <- c("Pre-employment+Living+Education",
                    "Pre-employment+Education","Pre-employment",
                    "Pre-employment+Living")
colnames(tab1) <- c("BIC")
tab1 <- xtable(tab1,caption =
                 "Model selection for work percentage
               five to ten years after LTx")
print(tab1, scalebox=0.8, caption.placement = "top")
final.4  <- lm(five_ten_perc ~ Employment,data=dat)
tableRegression(final.4,intercept = F,caption=
                  "Selected model for work status five to ten years after LTx",
                caption.placement="top",row.nam = "Pre-employment(Employed)")

dat.choice.e <- dat[-which( is.na(dat$Employment) | is.na(dat$Sex)
                          | is.na(dat$Education) | is.na(dat$Relationship)
                          | is.na(dat$Living) | is.na(dat$Age)
                          | is.na(dat$BMI) | is.na(dat$X6MWD)
                          | is.na(dat$FEV1) | is.na(dat$longer_ten_perc)
                          | is.na(dat$Waiting_time)),]

for(i in 1:10){
  fml[i] = paste(name_var2[15],"~",name_var2[i],sep="")
  reg=lm(fml[i], data=dat.choice.e)
  pred <- predict(reg,type="response")
  bic[i] <- round(BIC(reg),2)
}
df <- matrix(0,nrow=10,ncol=3)
for(i in 1:10){
  fml[i] = paste(name_var2[15],"~",name_var2[i],sep="")
  reg=lm(fml[i],data=dat)
  pv[i]<-formatPval(summary(reg)$coefficients[2,4])
df[i,] <- c( (coef(reg)[2]), (confint(reg)[2,]))
}
df <- data.frame(round(df[,1],digits=2),formatCI(df[,c(2:3)],text = "english"),pv,bic)
colnames(df)<-tab_colname2
rownames(df) <- c("Age","Best FEV1(\\%)","BMI","Education(Academic)",
                  "Living(Alone)","Pre-employment(Employed)",
                  "Relationship(Single)","Sex(Male)","6MWD","Waiting time")
res.table<-xtable(df, caption =
                    'Univariate models for work percentage longer than ten years after LTx',
                  table.placement ="")
print(res.table, scalebox=0.8,
```

```
      caption.placement = "top",
         sanitize.text.function = function(x) {x})
reg0=lm(longer_ten_perc ~ Employment,data=dat.choice.e)

reg1=lm(longer_ten_perc ~ Education+ Living+Employment,data=dat.choice.e)

reg2=lm(longer_ten_perc ~ Living+Employment,data=dat.choice.e)

reg3=lm(longer_ten_perc ~ Education+Employment,data=dat.choice.e)

tab1 <- t(data.frame(BIC(reg2),BIC(reg1),BIC(reg0),BIC(reg3)))
rownames(tab1) <- c("Pre-employment+Living","Pre-employment+Living+Education",
                    "Pre-employment","Pre-employment+Education")
colnames(tab1) <- c("BIC")
tab1 <- xtable(tab1,caption =
                 "Model selection for work percentage longer than ten years after LTx")
print(tab1, scalebox=0.8, caption.placement = "top")


final.5  <- lm(longer_ten_perc ~ Employment,data=dat)
tableRegression(final.5,intercept = F,caption=
                  "Selected model for work status longer than ten years after LTx",
                caption.placement="top",row.nam = "Pre-employment(Employed)")

####### code for the result of logistic regression
####### for all of the seprate time periods
fit1 <- lm(less_one_perc ~ Employment,data=dat)
fit2 <- lm(one_three_perc ~ Employment,data=dat)
fit3 <- lm(three_five_perc ~ Employment,data=dat)
fit4 <- lm(five_ten_perc ~ Employment,data=dat)
fit5 <- lm(longer_ten_perc ~ Employment,data=dat)

coeff1 <- rbind(
              (coef(fit1)[2]),
              (coef(fit2)[2]),
              (coef(fit3)[2]),
              (coef(fit4)[2]),
              (coef(fit5)[2]))

coeff1 <- sprintf("%.2f", coeff1)
Conf1 <- rbind(formatCI((confint(fit1)[2,]),text = "english"),
             formatCI((confint(fit2)[2,]),text = "english"),
             formatCI((confint(fit3)[2,]),text = "english"),
             formatCI((confint(fit4)[2,]),text = "english"),
             formatCI((confint(fit5)[2,]),text = "english"))


p <- rbind(
  biostatUZH::formatPval(summary(fit1)$coefficients[2,4]),
  biostatUZH::formatPval(summary(fit2)$coefficients[2,4]),
  biostatUZH::formatPval(summary(fit3)$coefficients[2,4]),
  biostatUZH::formatPval(summary(fit4)$coefficients[2,4]),
  biostatUZH::formatPval(summary(fit5)$coefficients[2,4]))



mat_all_linear <- cbind(coeff1,Conf1,p)
```

```
colnames(mat_all_linear) <- c("Coefficients",
                              "$95\\%$-confidence interval",
                              "$p$-value")
rownames(mat_all_linear) <- c("Pre-employment(<1)",
                              "Pre-employment(1-3)",
                              "Pre-employment(3-5)",
                              "Pre-employment(5-10)",
                              "Pre-employment(>10)")

linear_Tab <- xtable(mat_all_linear,
                     caption="Result for linear regression  for periods after LTx",
                     label="tab:tab29")
print(linear_Tab,
      scalebox=0.8,
      caption.placement = "top",sanitize.text.function = function(x) {x})



### code for LDA
sum(is.na(duration))#4 NA
dat.dura <- cbind(dat,duration)
patient_longerten <- dat.dura[which(duration>10),c("ind","duration")]
#28 patients who end at the time period 5.
patient_longerfive <- dat.dura[which((duration>5)&(duration<10)),c("ind","duration")]
#19 patients who end at the time period 4.
patient_longerthree <- dat.dura[which((duration>3)&(duration<5)),c("ind","duration")]
#15 patients who end at the time period 3
patient_longerone <- dat.dura[which((duration>1)&(duration<3)),c("ind","duration")]
#23 patients who end at the time period 2
patient_withinone <- dat.dura[which((duration<1)&(duration>0)),c("ind","duration")]
#6 patients who end at the time period 2
time_periods <- c(nrow(patient_longerten),
        nrow(patient_longerfive),
        nrow(patient_longerthree),
        nrow(patient_longerone),
        nrow(patient_withinone))
names(time_periods) <- c(5,4,3,2,1)
tb_bar <- barplot(time_periods,ylim=c(0,35),
                  ylab  = "Number of patients",
                  xlab="Number of measurements")
text(x = tb_bar, y = time_periods, label = time_periods, pos = 3, cex = 0.8, col = "black")

## code for LAD for work status

dat <- longformat
dat$ind <- factor(dat$ind)
dat0 <- dat[-which(is.na(dat$Education)|
                     is.na(dat$Workstatus)|
                     is.na(dat$Employment)),]

reg0 <- glm(Workstatus ~ log(Time) + Employment +Education,data=dat0,family="binomial")
reg1 <- glm(Workstatus ~ (Time) + Employment +Education,data=dat0,family="binomial")
reg2 <- glm(Workstatus ~ I(Time^2) + Employment +Education,data=dat0,family="binomial")
bic1 <- BIC(reg0)
bic2 <- BIC(reg1)
bic3 <- BIC(reg2)
Mwith2 <- c(bic1,bic2,bic3)
```

```
reg.0 <- glm(Workstatus ~ log(Time) + Employment ,data=dat0,family="binomial")
reg.1 <- glm(Workstatus ~ (Time) + Employment ,data=dat0,family="binomial")
reg.2 <- glm(Workstatus ~ I(Time^2) + Employment ,data=dat0,family="binomial")
bic4 <- BIC(reg.0)
bic5 <- BIC(reg.1)
bic6 <- BIC(reg.2)
Mwith1 <- c(bic4,bic5,bic6)

bicdf.log <- data.frame(Mwith1,Mwith2)
colnames(bicdf.log) <- c("Pre-employment","Education + Pre-employment")
rownames(bicdf.log) <- c("log(Time)","Time","$(\\mathrm{Time})^2$")
bicdf.log <- xtable(bicdf.log,caption=
                     "Time transformation for work status after LTx based on BIC",label="tab:tab30")
print(bicdf.log,
      caption.placement="top",
      sanitize.text.function = function(x) {x})



dat.long <- longformat
fit_GLMM <- glmer(Workstatus ~ Employment + Education  + (1|ind) + CLAD + log(Time) ,
                data=dat.long, family="binomial")

OR_GLMM <- exp(summary(fit_GLMM)$coefficients[,1])[-1]
test <- confint(fit_GLMM)
ci_GLMM <-formatCI(exp(test),text = "english")[-c(1:2)]
p_GLMM <- formatPval(summary(fit_GLMM)$coefficients[,4])[-1]

GLMM_colname <- c("Odds Ratio","$95\\%$-confidence interval","$p$-value")
GLMM_rowname <- c("Pre-employment","Education","CLAD","log(Time)")
df.GLMM <- data.frame(round(OR_GLMM,digits=2),ci_GLMM,p_GLMM)
colnames(df.GLMM)<- GLMM_colname
rownames(df.GLMM) <- GLMM_rowname
df_tab <- xtable(df.GLMM,
caption=
"Result for Generalized Linear Mixed-effects Model
for Post-LTx Work Status",
label="tab:tab32")
print(df_tab, scalebox=0.8, caption.placement = "top",sanitize.text.function = function(x) {x})
## code for LAD for work percentage
dat0 <- dat.long[-which(is.na(dat.long$Education)|
                          is.na(dat.long$Percentage)|
                          is.na(dat.long$Employment)),]
reg0 <- lm(Percentage ~ log(Time) + Employment +Education,data=dat0)
reg1 <- lm(Percentage ~ (Time) + Employment +Education,data=dat0)
reg2 <- lm(Percentage ~  I(Time^2) + Employment +Education,data=dat0)
bic1 <- BIC(reg0)
bic2 <- BIC(reg1)
bic3 <- BIC(reg2)
reg.0 <- lm(Percentage ~ log(Time) + Employment ,data=dat0)
reg.1 <- lm(Percentage ~ (Time) + Employment ,data=dat0)
reg.2 <- lm(Percentage ~  I(Time^2) + Employment ,data=dat0)
bic4 <- BIC(reg.0)
bic5 <- BIC(reg.1)
bic6 <- BIC(reg.2)
Mwith1 <- c(bic4,bic5,bic6)
Mwith2 <- c(bic1,bic2,bic3)
```

```
bicdf.linear <- data.frame(Mwith1,Mwith2)
colnames(bicdf.linear) <- c("Pre-employment","Education + Pre-employment")
rownames(bicdf.linear) <- c("log(Time)","Time","$(\\mathrm{Time})^2$")
bicdf.linear <- xtable(bicdf.linear,
                       caption="Time transformation for work percentage after LTx",label="tab:tab29")
print(bicdf.linear,
      caption.placement="top",
      sanitize.text.function = function(x) {x})
dat.sel <- dat[-which(
  is.na(dat$Education)|is.na(dat$Percentage)|
    is.na(dat$Employment)|is.na(dat$CLAD)|
    is.na(dat$Cancer)|is.na(dat$Kidney_Dialysis)),]
reg1 <- lm(Percentage ~ log(Time) + Employment+Education
             +CLAD
             +Cancer
             +Kidney_Dialysis,
             data=dat.sel)
r1 <- c(1,1,1)
reg2 <- lm(Percentage ~ log(Time) + Employment+Education
            # +CLAD
             +Cancer
             +Kidney_Dialysis,
             data=dat.sel)
r2 <- c(0,1,1)
reg3 <- glm(Percentage ~ log(Time) + Employment+Education
             +CLAD
            # +Cancer
             +Kidney_Dialysis,
             data=dat.sel)
r3 <- c(1,0,1)
reg4 <- lm(Percentage ~ log(Time) + Employment+Education
             +CLAD
             +Cancer
             #+Kidney_Dialysis
             ,data=dat.sel)
r4 <- c(1,1,0)

reg5 <- lm(Percentage ~ log(Time) + Employment+Education
           #   +CLAD
            # +Cancer
             +Kidney_Dialysis,
             data=dat.sel,family="binomial")
r5 <- c(0,0,1)

reg6<- lm(Percentage ~ log(Time) + Employment+Education
             +CLAD
            # +Cancer
            # +Kidney_Dialysis
           ,
             data=dat.sel)
r6 <- c(1,0,0)

reg7<- lm(Percentage ~ log(Time) + Employment+Education
            # +CLAD
             +Cancer
             #+Kidney_Dialysis
           ,
```

```
              data=dat.sel)
r7 <- c(0,1,0)
df <- data.frame(r1,r2,r3,r4,r5,r6,r7)
bicdf.linear <- c(BIC(reg1),BIC(reg2),BIC(reg3),BIC(reg4),BIC(reg5),BIC(reg6),
                  BIC(reg7))
df <- rbind(df,bicdf.linear)
rownames(df) <- c("CLAD","Cancer","Kidney-Dialysis","BIC")
df <- t(df)
df <- df[order(df[,"BIC"]),]
rownames(df) <- c("Model 1","Model 2","Model 3",
                  "Model 4","Model 5","Model 6","Model 7")
df.table<-xtable(df,  digits=c(0,0,0,0,2), table.placement ="",
                 caption = 'Model selection for work percentage based on Time-dependent factors')
print(df.table, scalebox=0.8, caption.placement = "top")



fit_LMM <-lme4::lmer(Percentage ~ Employment  + Education +
(1|ind) + Kidney_Dialysis +log(Time) ,
data=dat.long)

coef_fit <- lmerTest::lmer(fit_LMM,data=dat.long)
coef_LMM <- (summary(coef_fit)$coefficients[,1])[-1]
test <- confint(coef_fit)
ci_LMM <-formatCI((test),text = "english")[-c(1:3)]
p_LMM <- formatPval(summary(coef_fit)$coefficients[,5])[-1]
LMM_colname <- c("Coefficients","$95\\%$-confidence interval","$p$-value")
LMM_rowname <- c("Pre-employment","Education","Kidney-Dialysis","log(Time)")
df.LMM <- data.frame(round(coef_LMM,digits=2),ci_LMM,p_LMM)
colnames(df.LMM)<- LMM_colname
rownames(df.LMM) <- LMM_rowname
df_tab <- xtable(df.LMM,
                 caption=
                   "Result for Linear Mixed-effects
                 Model for Post-LTx Work percentages",
                 label="tab:tab35")
print(df_tab, scalebox=0.8, caption.placement = "top",sanitize.text.function = function(x) {x})
```