Work placement Project Portfolio

Table of Content

Preface

Introduction

Chronic kidney disease (CKD) is a progressive loss of kidney function. The symptoms might include feeling generally unwell and experiencing a reduced appetite. This disease is often diagnosed through the screening of people who are at high risk of getting kidney problems such as those with diabetes or high blood pressure. Chronic kidney disease may also be identified when it leads to one of its recognized complications, such as cardiovascular disease, anemia or pericarditis [National Kidney Foundation, 2002].

Chronic Kidney disease is identified by a blood test for creatinine. A high level of creatinine in the blood indicates a low glomerular filtration rate (GFR) (the rate at which the kidney filters blood of all its impurities) and thus a low capacity of the kidney to excrete waste products. Creatinine levels may be normal in the early stages of CKD and the condition is discovered if urinalysis (the testing of a urine sample) shows that the kidney is allowing the loss of protein or red blood cells into the urine. To fully investigate the underlying cause of kidney damage, various forms of medical imaging, blood tests and often renal biopsy are employed to find out if there is a reversible cause for the kidney malfunction [National Kidney Foundation, 2002].

Recent professional guidelines classify the severity of chronic kidney disease into five stages, with Stage 1 being the mildest and usually causing few symptoms and Stage 5 being a severe illness with poor life expectancy if untreated [National Kidney Foundation, 2002]. End-stage kidney disease (ESKD), synonymous with Stage 5 CKD, is when kidney function has deteriorated to the point that life can no longer be sustained. All individuals with a GFR < 60 mL/min/1.73 m^2 for three months are classified as having CKD, irrespective of the presence or absence of kidney damage. The rationale for including these individuals is that reduction in kidney function to this level or lower represents loss of half or more of the adult level of normal kidney function, which may be associated with a number of complications [National Kidney Foundation, 2002]. All individuals with kidney

damage are classified as having CKD, irrespective of the level of GFR; the rationale for this is that GFR may be sustained at normal or increased levels despite substantial kidney damage [National Kidney Foundation, 2002]. The loss of protein in the urine is regarded as an independent marker for worsening of kidney function and cardiovascular disease. Hence, the British guidelines append the letter "P" to the stage of chronic kidney disease if there is significant protein loss [National Institute for Health and Clinical Excellence, 2008].

A person with ESKD will require permanent renal replacement therapy for life. Renal replacement therapy may be by dialysis or by kidney transplantation, with the latter being the best treatment for most people in terms of better quality of life and longer survival. Persons on renal replacement therapy have a higher risk for cancer than the general population, with persons with a transplant more at risk for cancer than persons on dialysis.

Analysis of cancer risk in people with ESKD has tended to concentrate on risk either during dialysis or after transplantation. However, since a graft has a finite life, persons switch to dialysis while waiting for a transplant. Thus there could possibly be several switches back and forth between treatments over the life-course of an ESKD patient. The first project examined cancer as a risk factor during a young (under 20 years at onset of ESKD) person's entire life-course with ESKD and determined what effect it has on a person's survival and the survival of a kidney graft. It also described the cancers that person's with ESKD had and determined if there was an association between the type of renal replacement therapy (dialysis or transplant) and getting a new cancer. Using the data from the Australia and New Zealand Dialysis and Transplant Registry which collects data on all persons in whom renal-replacement therapy was started in Australia and New Zealand from 1963, and restricting it to data for persons who were under the age of 20 years at the onset of ESKD, we examined cancer as a risk factor for a transplant failure and survival.

Past research has shown that among all patients with ESKD, the survival rate is lower than the survival rate of the general population; this lower survival rate was also seen for a subgroup of patients who were children (under 20 years) when their ESKD began (McDonald and Craig, 2004). However to the best of our knowledge no study has examined what additional effect, if any, cancer has on survival for these patients who were

children when their ESKD began. Also to the best of our knowledge no research work has provided the cause-specific incidence proportion of deaths or the cause-specific survival among patients who were children when their ESKD began. Research has also shown that patients with ESKD have a higher risk of acquiring cancer than the general population and have specified what cancers are most common among ESKD patients (Maisonneuve et al, 1999). However, to the best of our knowledge, no research work/s examines these findings for the subgroup of patients who were children when they acquired ESKD.

The two related projects that form this portfolio were undertaken between August 2008 and June 2009. Both projects involved analyzing data on all persons in the Australia and New Zealand Dialysis and Transplant Registry who were under 20 years of age when they started renal-replacement therapy (RRT).

Student's Role

My role in each project was to prepare the data for analysis, provide suggestions on how the data could be analyzed, implement the agreed analysis plan, provide interpretation of the final results and write the report.

Dr. Angela Webster provided invaluable support and advice on what type of analysis was appropriate for this study while Dr. Patrick Kelly provided invaluable support and advice on some of the technical aspects of the statistical analysis involved in the project.

Reflections on learning

The Work Place Projects did support my belief, molded from my experience in the workplace, that for any analysis, a majority of the time is taken up in cleaning data, deriving variables needed for the analysis, coding some of the data, writing SAS macros to automate some of the repetitive and similar analyses, determining the correct/appropriate methods to do the analysis and interpreting the results in the final report. In contrast very little time is spent in actually producing statistical results using statistical software. To produce results that are correct and appropriate, time must be spent on data cleaning and

management; to produce results that are meaningful, time must be spent in conceptualizing the problem and interpreting the results. While the coursework for the Masters does help one get the skills to determine the appropriate statistical tools necessary for the job and to implement the analysis using statistical software, which typically is just a narrow aspect of any project, the WPP broadened my skills to include the whole process of carrying out a statistical project.

The statistical tools/techniques used for any project must be appropriate for the analyses. Thus sometimes, one is required to use tools/techniques that go beyond the scope of one's learning in the coursework component of the Masters course. Both my projects required me to use techniques such as Relative Survival and Standard Incidence Ratios that were completely new to me, and techniques such as Survival Analysis with time-varying covariates that were not new but that went above and beyond the scope of what I learnt during the coursework.

SAS is a powerful and flexible piece of statistical software that allows one to write macros to automate some of the tasks one has to do repeatedly. For example if one has to run a Cox regression on a set of variables and then arrange the results in a table in an RTF document in a certain way and one has to do this analysis for different sets of variables, one can write a macro where one just specifies the set of variables and the macro automatically produces the results in a table. In order to use this powerful macro building feature, I had to learn how to analyze survival data using SAS on my own, as STATA was used in the coursework. Also, in order to use SAS to do the analysis I had to spend time arranging the data in a way SAS required and deriving certain variables needed by SAS which were different from those required for the analysis using STATA. Although there was a huge upfront cost in terms of time spent, the benefits in terms of time saved in using the macros multiple times, and the long-term benefits of learning another commonly used software package to do the analysis, far outweighed the costs.

Another set of skills that was greatly enhanced by these projects were communication skills. No matter how thoroughly the analyses has been conducted, no matter how important the results of the analyses, a lot may be lost if one is not able to communicate the results of the analyses in way that is understood by the reader and holds his or her interest.

One has to build a story, an interesting story; a story in which the objectives are made clear to the reader and the reader is lead to the conclusions of the objectives in an orderly way. In writing my story I have learnt to think beyond mere technicalities, see things more broadly, and be more creative in my story-telling style.

In our modern times in which short but numerous social interactions is such an important part of ones lives; times in which these types of social interactions are almost demanded of one; times in which one probably is on Face Book and Tweeter and has three or four Instant Message accounts besides having an email, SMS and a mobile phone service, it is vitally important to have a very specific type of social and inter-relational skill required to meet this demand. My experience in doing the projects has certainly enhanced this skill as I have learnt how to write emails that were short (the shortest being a one-word email) and to the point, and I learnt how to get the most out of our short but productive meetings we would have once every few weeks which were conducted in a friendly manner.

Ethical considerations

The registry complies with the Australian Commonwealth Privacy Act governing health data collection and maintains patient anonymity by releasing for analysis only de-identified data. Quality control methods for data entry and consistency are routine for all data held in ANZDATA, and corrections are also made when inaccuracies are identified when the data are used. The de-identified data was always encrypted and password-protected when not being used.

The other datasets used came from the public domain.

Location and dates

School of Public Health, University of Sydney

August 2008 – June 2009

Context

Dr. Judy Simpson asked me if I would be interested in doing the two projects in the area of End Stage Kidney Disease and working with Dr. Angela Webster and Dr. Patrick Kelly who had volunteered to be advisors, to which I answered that I would. I met with Dr. Webster and Dr. Kelly in June of 2008 and they provided me with the dataset and explained to me its content. Dr. Webster also provided me with a few ideas of what specific analyses could be done; she also provided me with a few published papers for me to get a few ideas from. I left that meeting knowing that my first step would be to get familiar with the rather complicated dataset and also knowing that this project, even though I knew it would be about the association between long-term survival and cancer, would be an evolving process. In January 2009 at one of my regular meetings with Dr. Angela Webster and Dr. Patrick Kelly we discussed ideas of what the second project should be. In the end we decided that although, in Project 1, we did paint a picture of survival for an ESKD patient who had cancer who was a child at the onset of the disease, this picture would be complete if we could contrast it with the life of similar patients who did not have cancer or did have cancer but did not have ESKD.

Student contribution

I had several meetings with Dr. Webster and Dr. Kelly in which we discussed the analysis done so far, the changes to be made to the analysis, the next steps to be taken in the analysis, and the appropriate methods needed for the analysis. The analysis itself was carried out by me.

Statistical issues

The main statistical analysis used for this Project 1 was Survival Analysis. The methods/graphs used were Kaplan Meier Survival/Failure curves, Hazard and Cumulative Hazard curves, Cox regression with time-varying covariates, and Hazard plots from Cox regressions. A slightly complicated issue was how to model some of the time-varying covariates so that the covariates adequately reflected the course over each person's span of time between the onset of ESKD and death or censor.

The statistical tools used for this Project 2 were Standardized Incidence Ratios, Standardized Mortality Ratios and Relative Survival.

Acknowledgements

I would like to thank Dr. Angela Webster and Dr. Patrick Kelly for the invaluable assistance in this project and for keeping me in track to complete it by the due date. I would also like to thank my wife Linute for all the support she has given me.

Student declaration

I declare that this project is my own work; with guidance provided by my project supervisors, Dr. Patrick Kelly and Dr. Angela Webster, and that I have not previously submitted it for academic credit.

_____

Mario I D'Souza

_____

Date

Supervisor declaration

Mario's double project is a little unusual in that the two projects use the same dataset. However, as you will see, Mario has used different statistical methods for each project. His supervisors and Professor Judy Simpson agreed that both the quantity and variety of work undertaken by Mario was sufficient to be considered as equivalent to two projects. Mario also had to comprehend a complex dataset and do a substantial amount of data manipulation.

All the work presented is Mario's own work. In order to complete the projects, he had to learn several new statistical methods - time-varying covariates for Cox models, relative survival and standardized mortality (or incidence) ratios. Mario was able to learn these new methods and complete the analyses required with little input from his supervisors.

_____

Patrick Kelly

_____

Date

The Story

To give you an idea of the course of life of a patient with ESKD from the onset of ESKD to death let us use a fictitious person named Eliss. Here is her story.

Eliss came into this world sometime in 1981. She was a happy and healthy little child. A few months after her fifth birthday her parents noticed that Eliss, who rarely made a fuss about food, was refusing to eat even her favorite dish which was pasta bolognaise. They also noticed that Eliss, who used to be so cheerful, was becoming increasingly cranky. They took her to their local pediatrician to have her checked. The pediatrician told them not to worry as this was possibly just that Eliss had reached a milestone where she was asserting her independence. Just to rule out any other possibility, he took a sample of her blood to be tested. A day later her parents received news that would change their lives forever.

The pediatrician called to tell them that Eliss had extremely high levels of creatinine in her blood and that they needed to take her immediately to the Children's Hospital at Westmead for further testing. Eliss was diagnosed with Rapidly Progressive Glomerulonephritis and immediately placed under treatment to halt the progress of the disease. Unfortunately nothing could be done and her kidney's completely failed in a few weeks and she started receiving hemodialysis dialysis. Her parents were completely devastated.

Eliss's parents attended private information sessions at the Children's Hospital at Westmead with a nephrologist and a counselor. They both assured the parents that the treatment for ESKD had progressed over the decades and it is now quite common for a child to lead a relatively full and active life. However Eliss could be affected both physically and psychologically from the disease itself and/or the medication to control the disease. The physical conditions could be extreme fatigue, inability to concentrate, weak bones, nerve damage, sleep problems, weight gain and delayed growth and development, while the psychological problems could include depression and feelings of isolation.

They were also informed about their treatment choices which were broadly hemodialysis, peritoneal dialysis and transplantation and the risks and benefits of each. Kidney failure can lead directly to more health problems, like swelling of the body, bone deformities, and growth failure but a successful kidney transplant could give Eliss the best chance to grow normally and lead a relatively normal life. The donated kidney could come from a recently deceased person or a living person, but studies suggest that the graft from a living person survives longer. A donated kidney does not last forever; studies suggest that the median survival time of a graft in all children is around 10 to 12 years, but the good news for Eliss is that the median graft survival among Asian children in particular was in fact higher. Eliss would have to go back on dialysis once the donated kidney failed and while she awaited the next donated kidney. A kidney-transplantation does have elevated risks of other diseases associated with it, in particular cancer, and cancer in turn is associated with reduced survival. However studies also suggest that children who are in the five-to-nine-year age bracket when they start ESKD have a lower risk of acquiring cancer than children who are older.

While Eliss awaited her first transplantation she could either continue to receive hemodialysis which meant that they would need to bring Eliss to the hospital three or four times a week for three to six hour sessions, or she could switch to peritoneal dialysis which could be done at home but which needed to be carried out four to five times a day evenly spread out and each exchange taking about thirty minutes. With both types of dialysis Eliss would be able to do her homework or watch TV during the procedure.

Eliss's parents decided to go in for peritoneal dialysis since her mother was a stay-at-home parent taking care of Eliss and her younger brother Nick. Eliss's mother underwent training in how to perform the dialysis procedure. Eliss would have one dialysis done just before she went to school at 8 am, the second one during her lunch break at 12 noon, the third one when she came back home from school around 4 pm and the fourth one just before she went to bed around 9 pm. During the non-school days she would keep the same time schedule. Eliss's school headmistress was most accommodating and made arrangement for a small vacant room in the school to be used as her dialysis room which was big enough to store the dialysis solution and other equipment and also have a desk and a few chairs.

Both of Eliss's parents got themselves tested at the Westmead Hospital to see if they were suitable donors. Eliss's mother's kidneys turned out not to be a good match, but fortunately for her, her father's kidneys were a good match.

Around three months after Eliss was first diagnosed with ESKD she was successfully implanted with her father's donated kidney. She was also put on immunosuppressant drugs to stop the immune system from rejecting the donated kidney. She lived the next few years of her life in as normal a way as a child with ESKD could possibly live.

Eliss was now 13 years old and like any teenager would, was worried about her excess weight and pale skin colour. Eliss knew that she had to take the immunosuppressant drugs in order for her donated kidney not to be rejected by her body, but also knew that her excess weight and pale skin was due to the drugs. The desire to lose weight overcame her and she started skipping doses. A year later her donor kidney failed and she was put back on peritoneal dialysis.

The peritoneal dialysis procedure and the disease itself was crippling her social life as she always felt very tired and thus unable to enjoy partying or hanging out with friends for too long. Besides, her tired demeanor made her seem to her friends that she was not really interested. Also, her social life had to be scheduled according to her dialysis schedule; for example if she attended any parties she had to make sure she got back home before 9 pm to have her dialysis done.

The feeling of guilt at having caused her kidney's to fail, her crippled social life and her feeling of looking ugly because she had put on weight drove her to severe depression. She started spending most of her non-school hours in her room lying in bed. Her family tried to encourage her to socialize with the rest of the family members but she would complain the drugs were making her tired and she needed to lie down. She was less and less inclined to go to school and she made all sorts of excuses to skip school. Once even when her best friend Patti came to visit her, Eliss refused to let her into her room. No amount of counseling from the school counselor was having any effect on her. Her parents grew extremely worried and decided to take her to a reputable child psychiatrist who put her on

anti-depressives. After a few weeks on anti-depressives she started feeling better and started socializing once again. From then on the regular intake of anti-depressives became an integral part of life.

When she was fifteen years old she received a donor kidney from her dad's brother. Unfortunately her body rejected the graft as soon as it was transplanted. She went back on peritoneal dialysis. A few months later her doctor gave her the good news that a Continuous Cycling Peritoneal dialysis (CCPD) machine was available for her. This machine would automatically empty her abdomen many times at night during sleep and in the morning she would just need to do one exchange with a dwell time that lasted the entire day. The switch to CCPD improved her quality of life dramatically. At school she no longer needed to spend her entire lunch break doing her dialysis and could now socialize with her friends. She could stay back after school to cheer on her house's soccer team or attend after-school drama classes, etc. In the weekends and during school holidays, she could hang out with friends the whole day. She still could not attend school camps or stay overnight at a friend's place or stay partying till late, but she had come to terms with those restrictions and felt extremely happy with the freedoms which the CCPD afforded her.

Although she was quite happy with her treatment with CCPD, the doctors advised her to take the transplant option when it comes her way as it would increase her life. At the age of 18 the transplant option did come her way in the form of a kidney graft from a young man who died in a car crash in Parramatta. Eliss and her parents received a call from Westmead Hospital telling them that they had a kidney graft that was a good match and Eliss should be brought immediately to the hospital if they wanted the transplant done. Eliss was very uncomfortable about having a dead man's kidney in her but she did assent to it. The transplant was a successful one.

As part of her treatment for ESKD, Eliss needed to have a blood test done periodically to check the progress of the disease and to watch for symptoms of other diseases. When she was 22 years old, one of these blood tests revealed that the white blood cells and platelet counts had gone unusually low. Further testing revealed that she had Leukemia. The oncologist assured Eliss and her parents that the cure rate was quite high for this type of Leukemia. She was treated with pharmaceutical medications which had horrible side-

effects like nausea, vomiting, diarrhea, poor appetite, and complete hair loss. However she was completely cured in a few months and her hair grew back again.

She was 25 years old when she was last seen at the Westmead Hospital at one of her regular visits. Her doctor noted that she registered low on the depression scale and she reported feeling happy and contented in the quality of life questionnaire she filled out. She had experienced terrible misfortunes in her life but she had now reconciled to the fact that misfortunes would come her way and was mentally prepared to handle them. On a poster stuck on the door of her room in her family house was printed the immortal words of the famous German philosopher Friedrich Nietzsche “That which does not kill us makes us stronger”.

Overall Aims

The main aims of Project 1 were to determine if having cancer (before or after the onset of ESKD) reduces survival, to specify the main types of cancers that are associated with ESKD, to determine the incidence proportion of cancer among ESKD patients, to determine the risk factors for getting a cancer after the onset of ESKD, and to determine if cancer reduces graft survival.

The main aims of Project 2 were to determine the relative survival of patients with ESKD and post-onset cancer and to compare it with the relative survival of patients with ESKD, to determine the standardized mortality ratio (SMR) of patients with ESKD and post-onset cancer and compare it with the SMR of patient with ESKD, to determine the standardized incidence ratio (SIR) of post-onset cancer across subgroups of the ESKD population, and to determine the SIR for types of cancers.

1. Project 1

Analysis of cancer as a risk factor for survival in patients who were under the age of twenty at the onset of ESKD

1.1 Objectives

There were several objectives for this project. These objectives are stated below.

1. To ascertain the degree of association between cancer and survival in patients with ESKD.

2. To provide a descriptions of cancers in patients with ESKD.

3. To ascertain the incidence of cancer during ESKD.

4. To ascertain the risk factors for new cancer during ESKD and their corresponding degrees of risk.

5. To ascertain the incidence of a kidney transplant failure.

6. To ascertain the degree of association, if any, between cancer and graft survival.

1.2 Data

This project was a prospective inception-cohort study of all children under the age of 20 years in the Australia and New Zealand Dialysis and Transplant Registry. The registry collects information about patients receiving renal-replacement therapy who have a diagnosis of chronic renal failure and for whom indefinite renal-replacement therapy is intended. These data that is obtained by the registry every six months from all the renal units in Australia and New Zealand, is complete from the first use of renal-replacement therapy in Australia and New Zealand and includes information on the cause of end-stage kidney disease (ESKD), demographic characteristics of the patients, coexisting conditions, details of dialysis treatments and renal transplantation, and cause/s of death and past and new incidents of cancer diagnoses.

The registry complies with the Australian Commonwealth Privacy Act governing health data collection and maintains patient anonymity by releasing for analysis only de-identified data. Quality control methods for data entry and consistency are routine for all data held in ANZDATA, and corrections are also made when inaccuracies are identified when the data are used. Further information about ANZDATA can be found by visiting the website (ANZDATA, 2009).

All patients in whom renal replacement therapy started in Australia or New Zealand between 1st of January 1963 and 31st December 2006 and who were registered in the ANZDATA registry and who were under the age of 20 when this therapy began are included in the study. These patients were followed until death or 31st December 2006, whichever occurred earlier. Two patients who made a temporary recovery of renal function were excluded from the dataset. There were 1930 patients in the dataset.

All the analysis was done using this registry data and these data only.

Inclusion criteria

The following is the inclusion criteria for this project:

- Residents of Australia or New Zealand
- Renal replacement therapy started in Australia or New Zealand between 1st of January 1963 and 31st December 2006
- Under the age of 20 when renal replacement therapy began

Restructuring of data and deriving variables

The analysis required the data to be restructured and variables to be described. Two data sets were created: one dataset had just one record per person and the other dataset had for those people who got a cancer one record for every cancer they got and for those who did not have a cancer one record per person.

The following is a list of the data manipulations done:

1. Certain non-cancers which were coded as cancers had to be deleted. Certain cancer related primary renal disease was in free text; these needed to be coded and brought into the primary disease field. Two or more cancers for each person could have the same exact date and this could have caused issues with the analyses and so in these cases a small amount of time was added to the later cancer/s so that the order of the cancers was maintained. A variable was created to determine whether the cancer was pre or post onset of ESKD. The number of pre and post onset of ESKD cancers had to be determined for each patient. A pre-first transplant cancer flag was derived which took the value 1 if a person did have a cancer and his or her cancer date was before the first transplant date and zero otherwise. For some of the analysis we needed to track in which period the cancer happened, i.e. did it happen in the first transplant period, or in the dialysis period straight after the first transplant period, etc. We created three variables for the first, second and third post-onset of ESKD cancer a person could possibly get and each of these variables took on the integers 1 to 10 with 1 being if the cancer fell in the period D1 (dialysis before the first transplant), 2 being if the cancer

fell in the period T1 (first transplant period), 3 being if the cancer fell in the period D2 (dialysis after first transplant), 4 being if the cancer fell in the period T2 (second transplant period), etc.

2. The variables age and age group at the onset of ESKD were derived. Age at onset of ESKD was derived by using the formula (first renal replacement therapy date – date of birth)/365.25. A categorical variable age group at onset of ESKD was created by dividing this age into five groups, i.e. < 1 years, 1-4 years, 5-9 years, 10-14 years,15-19 years. The variables age and age group at the first kidney transplant were derived. Age at first kidney transplant was derived by using the formula (first kidney transplant date – date of birth)/365.25. A categorical variable age group at first kidney transplant was created by dividing this age into five groups, i.e. < 4 years, 5-9 years, 10-14 years, 15-19 years, > 19 years.

3. The dataset had the first RRT date, transplant dates and transplant periods. Thus the dialysis dates had to be derived from this information. For example the start of the dialysis period after the first transplant failure would be the first transplant date plus the first transplant period, but only if the calculated date was not equal to the death date. The duration of each dialysis period had to be calculated. For example the duration of the first dialysis could be the first transplant date minus the first RRT date plus one; but if the person did not have a transplant then it would be the death date minus the first RRT date plus one; but if the person's first RRT date was a transplant then the person is stated to have not had a first dialysis.

4. Some duration variables for the survival analyses had to be derived:
   - Period from ESKD to death or censoring which is calculated as death date minus first RRT date plus 1
   - Period form ESKD to first cancer was calculated as first cancer date minus first RRT date

5. Era (1963-79, 1980-89, 1990-99, 2000-06) of first RRT had to be derived from the first RRT date. Race which was categorized in very narrow bands in the original data had to be collapsed into three broad categories: Caucasian, Indigenous & Pacific Islanders, and

Asians. Place of residence which was in very narrow bands in the original data had to be collapsed into three places: Australia, New Zealand, and overseas (and not NZ). 'Donor type' variable i.e. whether the donor was the sister, brother, mother, father, non-related, etc., which had narrow bands in the original data had to be collapsed into two categories: living (related or non-related) and deceased (related or non-related). The Cause of Death (from death registry) which were in narrow bands had to be collapsed into seven broad categories: cardiac, infection, vascular, treatment withdrawal, cancer, suicide/accident and other. The 'primary renal disease' (primary cause of renal disease) field which was in narrow bands had to be collapsed seven broad categories: Glomerulonephritis, reflux, congenital/urological, cystic/hereditary, interstitial nephritis, cancer related, and other/uncertain.

6. Two binary variables and two period variables were created for the first and second cancer that started while the person had his/her first donor graft in him. These variables had to be created very carefully since the first cancer could possibly have started before the first transplant, in which case the second cancer would have to be checked to see if it fell during the first graft survival period and coded as the first post-first transplant cancer; etc. Similarly the time period from the first transplant date to the first post-transplant cancer date could possibly be the date from the first transplant date to the second cancer date if the first cancer date was before the first transplant date, etc.

7. A censor variable needed to be created for the 'time to first transplant failure' analysis. This variable was a binary variable which took the value 1 if either the person had a more than one graft transplant over his or her period of observation or if he or she had just one graft then only if the there was a dialysis period after the first transplant period. Else the variable was given the value 0 but only if the number of grafts was greater than zero; if it was zero then the variable would take on a missing value. A censor variable needed to be created for the 'time to first cancer' analysis. This binary variable took on the value 1 if the first cancer date was not missing and zero otherwise.

8. To calculate the incidence of cancer in each treatment period i.e. the incidence of cancer in the dialysis period just before the first transplant (D1), the first transplant period (T1), the dialysis period just after the first transplant (D2), etc. we needed to get the

total number of first cancers in each of the treatment periods D1, T1, D2, etc. In order to calculate the incidence of the first cancer in each period we needed to calculate the number of person-years in each period. We calculated each person contribution of person years by first determining which period the person's first cancer fell into e.g. T1. Then the person's contribution to person years for that period would be from the date of his/her first cancer to the date when the first graft failed. His/her contribution to the other periods would be zero.

9. The cancers had to be described. The first step was collecting all the cancers which were in four separate fields (cancer site code and free text, cancer type code and free text), and bringing them together. The next step was to classify the cancer according to area e.g. digestive and cancer site e.g. pancreas, which was compatible to be coded using the ICD10 coding system. The next step was to give each cancer a specific ICD10 code.

2. Patient Characteristics

2.1 Description of Patient Characteristics

Table 2.1 proved some information regarding the distribution of the patients in our study over gender, race, etc, and by five age categories, the age categories being based on the age at the onset of ESKD. There were more males than females whose onset of ESKD began at a very young age; 74% of the infants and 61% of the 1 to 4 year olds who were diagnosed with ESKD were males. Almost 40% of the infants and 49% of the 1 to 4 year olds had Congenital/Urological as their primary disease category; about 50% of the 15 to 19 year olds had Glomerulonephritis as their primary disease category.

Peritoneal dialysis was the most commonly used first Renal Replacement Therapy (RRT) among the patients who were under the age of 15 at the onset of ESKD with almost 92% of the infants and 68% of the 1 to 4 year olds receiving that treatment whereas hemodialysis was more commonly used as the first RRT among patients who were 15 years and older at the onset of ESKD with 65% of the 15 to 19 year olds receiving that treatment; only about 10% of the patients received a transplant as his/her first RRT. The first kidney graft for children who were young at the onset of ESKD most often came from a living donor with almost 48% of the infants and 59% of the 1 to 4 year olds receiving it from a living donor; the first kidney graft for children who were older at the onset of ESKD most often came from a deceased donor with 58% of the 15 to 19 year olds receiving it from a deceased donor. Nearly 61% of the patients received at least one kidney graft and there were 28 patients who receiving four grafts and one patient who received five graft over the period of observation.

In the period 1963 to 1979, out of the 415 patients diagnosed with ESKD there were no infants diagnosed and only five 1 to 4 year olds. About 28% of the patients died during the period of observation and this percentage was fairly evenly distributed across the age group. The most common cause of death was cardiac failure 33% of the patients dying from it; the most common cause among patients who were young children at the onset of ESKD was infection with it accounting for 40% of the deaths among infants and 53% of the deaths among the 1 to 4 year olds; cancer was the cause of 7% of the deaths with children

whose onset of ESKD began at the ages of 10 to 19 years being most effected. 110 patients or 6% of the patients in our study were diagnosed with a cancer after the onset of ESKD out of which 99 patients were between the ages of 10 to 19 years old at the onset of ESKD.

Table 2.1 Description of Patient Characteristics

Age group at the onset of ESKD		Age Groups at the onset of ESKD					
		<1 46	**1 to 4** 147	**5 to 9** 281	**10 to 14** 484	**15 to 19** 972	**Overall** 1930
		n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Gender	Female	12 (26.1)	57 (38.8)	125 (44.5)	224 (46.3)	436 (44.9)	854 (44.2)
	Male	34 (73.9)	90 (61.2)	156 (55.5)	260 (53.7)	536 (55.1)	1076 (55.8)
Race	Caucasians	43 (93.5)	129 (87.8)	236 (84)	407 (84.1)	809 (83.2)	1624 (84.1)
	Indigenous and Pacific Islanders	2 (4.3)	7 (4.8)	31 (11)	54 (11.2)	118 (12.1)	212 (11)
	Asians	1 (2.2)	11 (7.5)	14 (5)	23 (4.8)	45 (4.6)	94 (4.9)
Primary Disease category	Glomerulonephritis	1 (2.2)	18 (12.2)	67 (23.8)	152 (31.4)	487 (50.1)	725 (37.6)
	Reflux	1 (2.2)	6 (4.1)	33 (11.7)	116 (24)	233 (24)	389 (20.2)
	Congenital/Urological	20 (43.5)	72 (49)	70 (24.9)	106 (21.9)	95 (9.8)	363 (18.8)
	Cystic/Hereditary	3 (6.5)	8 (5.4)	46 (16.4)	65 (13.4)	45 (4.6)	167 (8.7)
	Interstitial Nephritis	1 (2.2)	0	5 (1.8)	4 (0.8)	15 (1.5)	25 (1.3)
	Cancer related	2 (4.3)	4 (2.7)	3 (1.1)	1 (0.2)	2 (0.2)	12 (0.6)
	Other/Uncertain	18 (39.1)	39 (26.5)	57 (20.3)	40 (8.3)	95 (9.8)	249 (12.9)
First treatment group	Hemodialysis	4 (8.7)	11 (7.5)	53 (18.9)	175 (36.2)	633 (65.1)	876 (45.4)
	Peritoneal Dialysis	42 (91.3)	112 (76.2)	190 (67.6)	244 (50.4)	280 (28.8)	868 (45)
	Transplant	0	24 (16.3)	38 (13.5)	65 (13.4)	59 (6.1)	186 (9.6)
Country of residence at the first RRT	Australia	37 (80.4)	126 (85.7)	224 (79.7)	385 (79.5)	748 (77)	1520 (78.8)
	New Zealand	9 (19.6)	19 (12.9)	55 (19.6)	95 (19.6)	207 (21.3)	385 (19.9)
	Overseas	0	2 (1.4)	2 (0.7)	4 (0.8)	17 (1.7)	25 (1.3)
First donor	Deceased donor	10 (21.7)	40 (27.2)	123 (43.8)	253 (52.3)	563 (57.9)	989 (51.2)
	Living donor	22 (47.8)	86 (58.5)	131 (46.6)	189 (39)	264 (27.2)	692 (35.9)
Maximum number of transplants received by any patient	0	14 (30.4)	21 (14.3)	27 (9.6)	42 (8.7)	142 (14.6)	246 (12.7)
	1	30 (65.2)	102 (69.4)	188 (66.9)	287 (59.3)	560 (57.6)	1167 (60.5)
	2	2 (4.3)	19 (12.9)	57 (20.3)	110 (22.7)	205 (21.1)	393 (20.4)
	3	0	5 (3.4)	6 (2.1)	34 (7)	50 (5.1)	95 (4.9)
	4	0	0	3 (1.1)	11 (2.3)	14 (1.4)	28 (1.5)
	5	0	0	0	0	1 (0.1)	1 (0.1)
Era of first treatment	1963 to 1979	0	5 (3.4)	45 (16)	111 (22.9)	254 (26.1)	415 (21.5)
	1980 to 1989	4 (8.7)	32 (21.8)	75 (26.7)	154 (31.8)	275 (28.3)	540 (28)
	1990 to 1999	24 (52.2)	65 (44.2)	82 (29.2)	140 (28.9)	250 (25.7)	561 (29.1)
	2000 to 2006	18 (39.1)	45 (30.6)	79 (28.1)	79 (16.3)	193 (19.9)	414 (21.5)
Last known outcome	Alive Dec 2006	31 (67.4)	113 (76.9)	210 (74.7)	339 (70)	645 (66.4)	1338 (69.3)
	Dead	15 (32.6)	32 (21.8)	65 (23.1)	132 (27.3)	302 (31.1)	546 (28.3)
	Lost	0	2 (1.4)	6 (2.1)	13 (2.7)	25 (2.6)	46 (2.4)

Table 2.1 Description of Patient Characteristics (continued)

Patients with at least one pre-ESKD cancer	0	4 (2.7)	5 (1.8)	2 (0.4)	5 (0.5)	16 (0.8)
Patients with at least one post-ESKD cancer	3 (6.5)	4 (2.7)	4 (1.4)	25 (5.2)	74 (7.6)	110 (5.7)

Table 2.1(b) Description of Patient Characteristics (with number of deaths as denominator)

Age group at the onset of ESKD		Age Groups at the onset of ESKD					
		<1	**1 to 4**	**5 to 9**	**10 to 14**	**15 to 19**	**Overall**
		n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Cause of death	Cardiac	5 (33.3)	6 (18.8)	20 (30.8)	34 (25.8)	113 (37.4)	178 (32.6)
	Infection	6 (40)	17 (53.1)	8 (12.3)	19 (14.4)	61 (20.2)	111 (20.3)
	Vascular	0	1 (3.1)	12 (18.5)	21 (15.9)	30 (9.9)	64 (11.7)
	Treatment Withdrawal	3 (20)	5 (15.6)	13 (20)	17 (12.9)	25 (8.3)	63 (11.5)
	Cancer	0	1 (3.1)	2 (3.1)	13 (9.8)	23 (7.6)	39 (7.1)
	Suicide/Accident	0	1 (3.1)	2 (3.1)	8 (6.1)	13 (4.3)	24 (4.4)
	Other	1 (6.7)	1 (3.1)	8 (12.3)	20 (15.2)	37 (12.3)	67 (12.3)

2.2 Duration of dialysis and transplant periods and incidence of cancer in each period

Table 2.2 provides summary statistics about the time at each stage of treatment i.e. time on dialysis before first transplant, time on first transplant, etc. and the incidence of cancer at each stage of treatment. This analysis provides an exploratory look at a possible association between cancer and type of treatment (dialysis or transplant).

From Table 2.2 we see that persons differed considerably in terms of the duration of dialysis and transplant with the average difference in duration among them being nearly as much as the average duration itself. On average, the first and second transplants survived for seven years with 25% surviving for more than 11.7 years and 9.9 years for the first and second donor graft respectively; these are underestimations as we have not accounted for censoring here. The duration of the dialysis before the first transplant was short with an average of just 1.7 years but the duration of the later dialysis getting progressively longer with the duration of the dialysis after the third transplant being nearly six years; possibly reflecting the increasing difficulty of getting a kidney donor graft. Sixty percent of the 110 post onset of ESKD cancers started during the first transplant period and 20% occurred during the second transplant. The first transplant period, the second transplant period, and the period on dialysis just after the third transplant had incidence rates of cancer that were above five per 1000 patient years.

Table 2.2 Duration of dialysis and transplant and the incidence of cancer in each treatment period

	Duration of treatment (in years)		Incidence of cancer		
Treatment period	Mean (SD)	Median [Q1 - Q3]	Patient years	Number of cancers	Incidence (per 1000 patient years)
D1 - Dialysis before first transplant if any	1.7 (2.43)	0.9 [0.3 - 2.1]	3209	6	1.87
T1 - First transplant	7.7 (7.66)	5.3 [1.5 - 11.7]	12551	66	5.26
D2 - Dialysis just after first transplant	3.4 (3.81)	2.1 [0.9 - 4.4]	2586	9	3.48
T2 - Second transplant	6.5 (7.02)	4.2 [0.5 - 9.9]	3254	20	6.15
D3 - Dialysis just after second transplant	4.9 (5.35)	3.2 [1.1 - 6.6]	1189	5	4.21
T3 - Third transplant	5.4 (5.91)	2.8 [0.4 - 8.7]	645	2	3.10
D4 - Dialysis just after third transplant	5.9 (5.96)	3.3 [1.6 - 8.3]	353	2	5.67
T4 - Fourth transplant	5.7 (7.22)	2.4 [0.6 - 8.9]	165	0	0
D5 - Dialysis just after fourth transplant	3.4 (3.05)	2.1 [1.4 - 7.3]	34	0	0
T5 - Fifth transplant	4.9 (n/a)*	4.9 [4.9 - 4.9]	5	0	0

Note: The duration in each period and the number of cancers could be an underestimation because we have not accounted for the time to death beyond 31 December 2006.

*No SD because there is only one patient

2.3 Distribution of Cancers by Cancer Site

Table 2.3.1 provides a frequency distribution of the pre-onset-of-ESKD diagnosed cancer patients across cancer sites. Sixteen patients had a cancer before being diagnosed for ESKD out of whom 9 patients had cancer in the kidney area.

Table 2.3.1 Distribution of pre-onset-of-ESKD Diagnosed Cancer patients

	Cancer Site	ICD Codes	N=16 n(%)
Any		C00-C95	**16 (100)**
Reproductive and genitourinary			**9 (56)**
	Kidney (non-transplanted)	C64	9 (56)
Endocrine			**3 (19)**
	Adrenal gland	C74	3 (19)
Hematological			**4 (25)**
	Hodgkin disease	C81	1 (6)
	Leukemia	C91-C95	2 (13)
	Non-Hodgkin lymphoma	C82-C85	1 (6)

Table 2.3.2 provides a frequency distribution of the patients that had newly diagnosed cancers during ESKD. One hundred and ten patients had a cancer during the period of observation. Patients with cancers associated with the reproductive and genitourinary system and hematological system accounted for 67% of the patients, with the skin/connective tissue and digestive systems accounting for a further 26%. Nearly 30% of the patients had a Lymphoma related cancer.

Table 2.3.2 Distribution of post-onset-of-ESKD Diagnosed Cancer patients

	Cancer Site	ICD Codes	N=110 n(%)
Any		C00-C95	**110 (100)**
Lip and oral cavity			**4 (4)**
	Mouth	C03-C06	2 (2)
	Salivary gland	C07-C08	1 (1)
	Tonsil	C09	1 (1)
Digestive			**16 (15)**
	Anus	C21	4 (4)
	Colon	C18	3 (3)
	Esophagus	C15	1 (1)
	Gallbladder	C23-C24	4 (4)
	Liver	C22	1 (1)
	Pancreas	C25	1 (1)
	Rectum	C19-C20	1 (1)
	Stomach	C16	1 (1)
Respiratory and intra-thoracic			**1 (1)**
	Trachea, bronchus, and lung	C33-C34	1 (1)
Skin/connective tissue			**13 (12)**
	Connective and other soft tissue	C47-C49	1 (1)
	Kaposi sarcoma	C46	2 (2)
	Melanoma	C43	10 (10)
Reproductive and genitourinary			**36 (33)**
	Breast	C50	1 (1)
	Kidney (non-transplanted)	C64	7 (6)
	Kidney (transplanted)	C64	6 (6)
	Ovary	C56	1 (1)
	Penis	C60	1 (1)
	Testis	C62	3 (3)
	Uteri cervix	C53	8 (7)
	Uterus	C54	2 (2)
	Vagina	C52	2 (2)
	Vulva	C51	6 (6)
Neurological			**3 (3)**
	Brain	C71	1 (1)
	Meninges	C70	2 (2)

Table 2.3.2 Distribution of post-onset-of-ESKD Diagnosed Cancer patients (continued)

Endocrine			**6 (6)**
	Thyroid	C73	6 (6)
Hematological			**38 (35)**
	Hodgkin disease	C81	3 (3)
	Leukemia	C91-C95	3 (3)
	Non-Hodgkin lymphoma	C82-C85	15 (14)
	Lymphoproliferative disease		17 (16)
Unknown/other			**3 (3)**
	Unknown/other		3 (3)

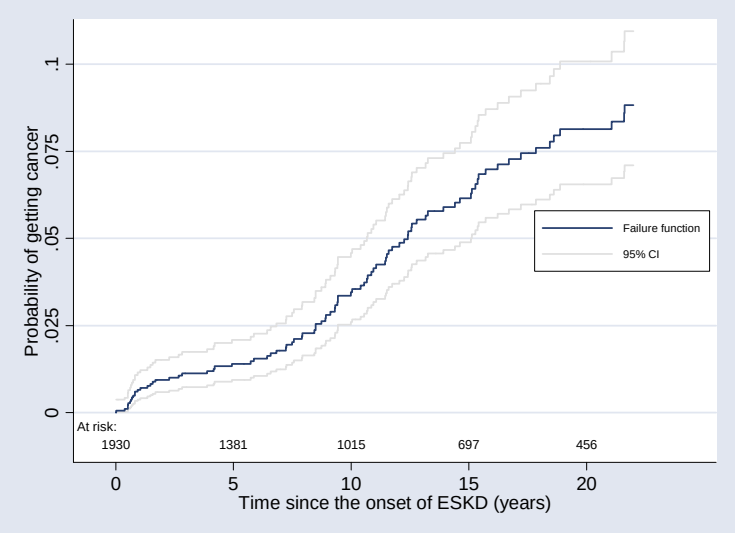
3. Time to First Cancer

3. 1 Incidence of post-onset-of-ESKD Cancer

In this section we provide the incidence proportion of cancer as a continuous function of time using the Kaplan-Meier (KM) Failure Curve for Time to First post-ESKD Cancer [Please see Appendix 1.1 for a definition and formula of the KM estimator]. Using a KM curve to calculate the incidence proportion was quite necessary here as there were, expectedly, a large number of censored observations. The total analysis time at risk for post-onset-of-ESKD cancer was 23,989 years for the 1,930 subjects we have in our data and 110 subjects were diagnosed with post-onset-of-ESKD cancer. Thus the incidence rate was 4.59 per 1000 patient years.

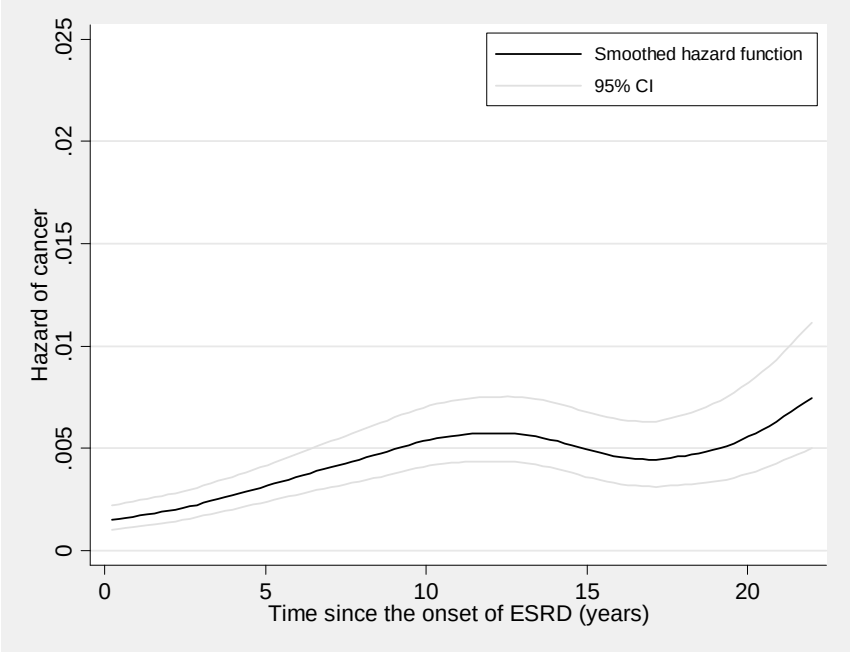
Figure 3.1.1 provides us with incidence proportion as a continuous function of time. The incidence proportion is defined here as the proportion of ESKD patients acquiring cancer. The graph shows that over a five year period from the onset of ESKD the incidence proportion was 1.25%, over a ten year period it was 3.5%, over a 15 year period it was 6.25%, and over a 20 year period it was 8%. There is a high initial hazard of getting a cancer in the first year after the onset of ESKD. This incidence proportion then increases at an increasing rate till 10-12 years and then increases at a constant rate.

Figure 3.1.1 Kaplan-Meier Failure Curve



From Figure 3.1.2 we see that the hazard of acquiring a new cancer increases till around 12 years post-onset of ESKD and then remains constant.

Figure 3.1.2 Smoothed Hazard Estimate



The Epanechnikov Kernel function is used to smooth the hazard estimate

3.2 Identifying the risk factors for post-onset-of-ESKD cancer

A Cox Proportion Hazards model was fitted with dependent variable for the model being 'time to first cancer since the onset of ESKD'. The covariates considered (deemed to be clinically important and based on the variable in the ANZDATA) for our model were gender, race, age group at onset of ESKD, treatment type, state of residence at the first renal replacement therapy, era of first treatment and primary disease category. 'Treatment type' was a time-varying variable that has the value 1 only for the time period after the person had his/her first transplant and 0 otherwise. (See Appendix 1.2 for description of the covariates).

A Proportional Hazards model was built using the model building process as prescribed by Hosmer et al (2008). The first step was to choose all covariates significant at the 25% level from a univariable analysis as well as those covariates deemed to be clinically significant. The initial multi-variable model was simplified by excluding variables that were not significant at the 5% level (starting with the covariate with the largest p-values) but testing if the excluded variable/s were important confounders (considered important if they change the coefficients of the existing variables by more than 20%). In the next step we added back, one at a time, the variables that did not meet the criteria of 25% in the univariable analysis. All tests were done using the Wald test.

The model is checked for strong violations of the proportion hazards assumption by the non-time-dependent covariates. We shall use a test based on the Schoenfeld residuals; the idea is to retrieve the residuals, fit a smooth function of time to them, and then test whether there is a relationship. We shall not make any modifications to our model if the proportion hazards assumption was not severely violated (p-value of global test ≥ 0.01) as it would just mean that we may not have modeled a weak but statistically significant interaction with time and thus have an average effect over the observed time. We choose a p-value of less than 0.01 because the number of events is large and so the tests of the PH assumption are quite sensitive to small violations in the assumption. If, however, the PH assumption was strongly violated then we shall refit the model by taking into account this interaction with

time. In the last step of our model building process we shall estimate our model and do post-hoc estimation of orthogonal contrasts.

Table 3.2.1 shows the results of the univariable analysis. Based on the overall p-value, the covariates chosen were race, age group, era of first treatment and treatment type.

Table 3.2.1 Univariable Analysis for Time to First Cancer (after onset of ESKD)

Covariate	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	Overall P-value
Gender	Female	Male	0.93	0.64	1.35	0.690
Race	Caucasians	Indigenous and Pacific Islanders	0.37	0.12	1.16	0.229
		Asians	0.89	0.28	2.82	
Primary Disease category	Glomerulonephritis	Cystic/Hereditary	0.74	0.35	1.57	0.383
		Reflux	0.87	0.55	1.37	
		Interstitial Nephritis	1.62	0.39	6.66	
		Congenital/Urological	0.65	0.36	1.17	
		Cancer related	0.00	0.00	.	
		Other/Uncertain	0.39	0.16	0.98	
Age group at onset of ESKD	5 to 9 years	< 1 years	11.32	2.52	50.85	0.006
		1 to 4 years	2.73	0.68	10.92	
		10 to 14 years	3.03	1.05	8.71	
		15 to 19 years	4.50	1.64	12.31	
State of residence at the first RRT	Australia	New Zealand	0.96	0.58	1.57	0.976
		Overseas	0.86	0.12	6.18	
Era of first treatment	1980 to 1989	1963 to 1979	1.32	0.81	2.15	0.206
		1990 to 1999	1.55	0.91	2.65	
		2000 to 2006	2.38	0.94	6.03	
Treatment Type	Dialysis	Transplant	1.87	1.14	3.06	0.013

Table 3.2.2 shows the results of fitting a multivariable model using the covariates chosen in the univariable analysis. Instead of using the binary time varying covariate for 'treatment type' we used a categorical time varying variable which provided us more information regarding differences in transplant and dialysis periods. This covariate had the value 1 for when the person was on dialysis just before the first transplant (D1), 2 for when she was on her first transplant (T1), 3 for when she was on dialysis just after the first transplant (D2), 4 for when she was on her second transplant (T2), 5 when she was on any dialysis after the

second transplant (D3+) and 6 when she was on any transplant after the second transplant (T3+).

Looking at the overall p-values we saw that race and 'era of first treatment' were the only two covariates that did not seem to be significant at the 5% level. Confounding was checked by dropping the two covariates from the model, running the regression and seeing if that changes the coefficients of the remaining covariates by more than 20%. Our check revealed that they were not important confounders and so they were dropped permanently from our model.

We next added back, one at a time, the variables that did not meet the criteria of 25% in the univariable analysis. The variables added back one at a time were pre-cancer flag, gender, state of residence at the first RRT, and primary disease category. None of the excluded variables were significant and so they were dropped permanently from the model.

We tested whether the different dialysis periods had equal risks and similarly whether the different transplant periods had equal risks. If the transplant periods had equal risk then we would create one covariate for transplant period. D2 risk was compared to D3+ risk to see if there was a significant difference; if there was then we look at a comparison of the two combined (D2 + D3+) versus D1 and if this was also not significant then we would merge them into one. If both, dialysis periods and transplant periods were not different, then we would revert back to our binary covariate for dialysis versus transplant. From Table 3.2.3 we see that the transplant periods had equal risk and the dialysis periods had equal risks. Thus we had only one time-varying binary covariate which took the value 1 when the person was on transplant and 0 otherwise.

Table 3.2.2 Initial Multivariable Analysis

Covariate	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	Overall P-value
Race	Caucasians	Indigenous and Pacific Islanders	0.45	0.14	1.45	0.410
		Asians	0.97	0.30	3.12	
Age group at onset of ESKD	5 to 9 years	< 1 years	10.41	2.29	47.43	0.004
		1 to 4 years	2.36	0.59	9.49	
		10 to 14 years	3.20	1.11	9.22	
		15 to 19 years	4.89	1.78	13.44	
Era of first treatment	1980 to 1989	1963 to 1979	1.16	0.71	1.89	0.184
		1990 to 1999	1.58	0.91	2.73	
		2000 to 2006	2.50	0.98	6.41	
Treatment Type	D1: Dialysis just before first transplant	T1: 1st Tx	4.09	1.54	10.81	0.045
		D2: Dx just after 1st Tx	2.82	0.87	9.06	
		T2: 2nd Tx	3.96	1.32	11.89	
		D3+: All Dx after 2nd Tx	2.92	0.84	10.21	
		T3+: 3rd or higher Tx	1.35	0.24	7.56	

Table 3.2.3 Tests for the equality of the dialysis and transplant periods

Linear Test		Wald ChiSq	DF	P-Value
T1=T2=T3+:	Transplant treatments periods have equal risk	2.577	2	0.276
D2=D3+:	Dx just after 1st Tx risk = All Dx after 2nd Tx risk	0.007	1	0.933
D2+D3+ = 0:	Dx just after 1st Tx + All Dx after 2nd Tx is significantly different from Dx just before 1st Tx	3.297	1	0.069

We next fitted the model and tested the proportional hazards assumption. Table 3.2.4 shows the results of the test. We see that the global test showed non-significance and we also see that none of the time-constant covariates had a significant p-value. Thus we do not modify the model.

Table 3.2.4 Test of the Proportional Hazards Assumption

	Rho	ChiSq	DF	Prob>ChiSq
Age group (years):				
<1	-0.02239	0.06	1	0.814
1 to 4	-0.06666	0.49	1	0.4859
10 to 14	0.00017	0	1	0.9985
15 to 19	-0.05769	0.37	1	0.5429
Treatment Type	-0.26314	8.59	1	0.0034
Global test		10.86	5	0.0543

Table 3.2.5 presents the results of fitting the final model in which we see that, among the covariates considered for our model, the only two covariates that were significantly associated with time-to-first-post-onset-of-ESKD-cancer were 'age-group at the onset of ESKD' and 'type of treatment'.

In Table 3.2.6 we present post-estimation orthogonal contrast which gives us estimations of important pair-wise comparisons. The hazard of getting a cancer for the less-than-one-year-olds was 4.3 times (95% CI: 0.97-19.5) that of the one-to-four-year-olds, 11.5 times (95% CI: 2.6-51.4) that of the five-to-nine-year-olds, just under 3.8 times (95% CI: 1.1-12.4) that of the ten-to-fourteen- year-olds, and no different from that of the fifteen-to-nineteen-year-olds. The hazard for the five-to-nine-year-olds was about a third (95% CI: 0.11-0.93) the hazard of the ten-to-fourteen- year-olds and about a fifth (95% CI: 0.08-0.59) of the hazard of the fifteen-to-nineteen-year-olds.

The hazard of getting a cancer when one was on a transplant was 1.9 times (95% CI: 1.2-3.2) the hazard of getting a cancer when one was on dialysis. Thus being on a transplant almost doubled ones risk of acquiring cancer.

Table 3.2.5 Final Multivariable Analysis for Time to First Cancer (Post-Onset of ESKD)

Covariate	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	P-value	Overall P-value
Age group at onset of ESKD	5 to 9	< 1	11.47	2.56	51.43	0.012	0.005
		1 to 4	2.64	0.66	10.55	0.572	
		10 to 14	3.08	1.07	8.86	0.721	
		15 to 19	4.63	1.69	12.67	0.132	
Treatment Type	Dialysis	Transplant	1.94	1.18	3.20	0.009	0.009

Figure 3.2.1 presents the hazard of cancer by treatment type derived after estimating a Cox model. We see that the hazard for a patient during any transplant period was higher that in any dialysis period.

Figure 3.2.2 presents the hazard of cancer for the different age groups by treatment derived after estimating a Cox model. We see that the hazard of a cancer for a patient during any kidney transplant period was higher than in any dialysis period for all age groups except possibly for the five-to-nine age group for whom it seemed to be equal. We see that the less-than-one-year-olds had a much higher hazard of getting a cancer than the rest of the age-groups. Also, among the less-than-one-year-olds, the hazard of cancer during any transplant period was much higher than during any dialysis period. Taken together this implies that the less-than-one-year-olds with a transplant were at the highest risk of acquiring cancer.

Figure 3.2.1 Hazard of Cancer from Transplant or Dialysis

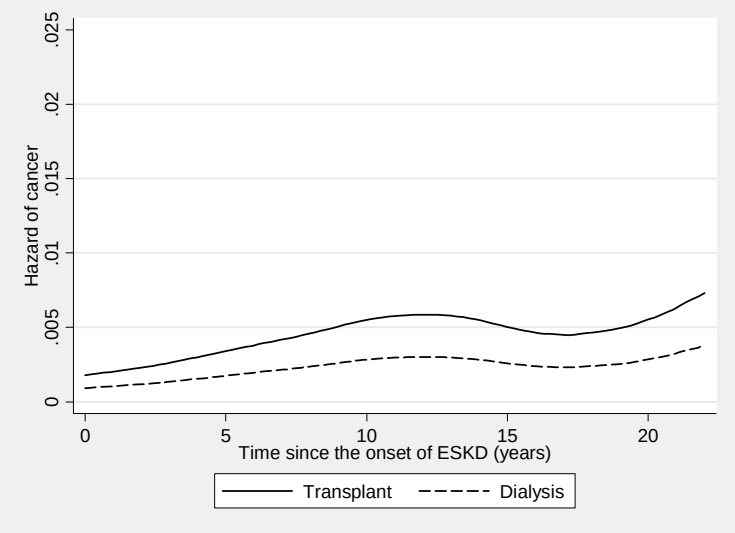
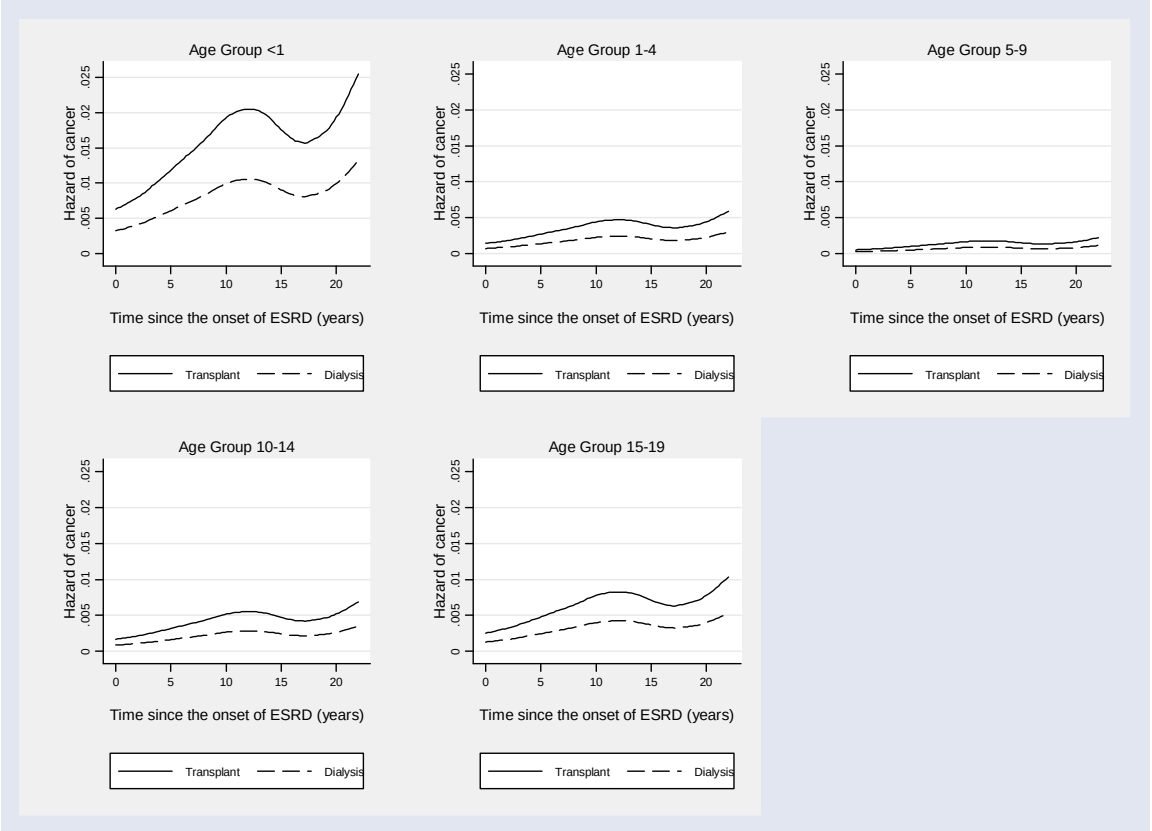


Figure 3.2.2 Hazard of Cancer for the different age groups by type of treatment



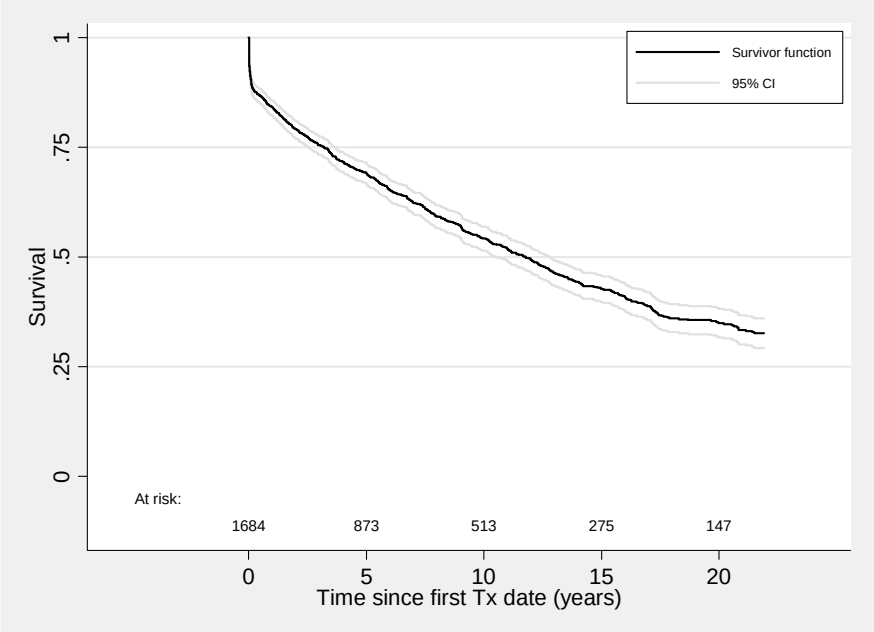
4. Time to First Kidney Transplant Failure

4.1 Incidence of First Kidney Transplant Failure

In this section we provide the incidence proportion of first kidney transplant failure as a continuous function of time using the Kaplan-Meier Failure Curve for Time to First Kidney Transplant Failure. The analysis variable here was 'time to first kidney transplant failure from the date of first transplant'. Out of the 1930 subjects in our study, 246 did not have a kidney transplant. The total analysis time for the remaining 1,684 subjects was 12,890.67 years and the number of subjects who had a transplant failure was 812. Thus the incidence rate was 63 per 1000 patient-years.

Figure 4.1.1 provides us with incidence proportion as a continuous function of time. The incidence proportion is defined here as the proportion of ESKD patients with a first kidney graft whose graft failed. Over a five year period from the date of the transplant the incidence proportion was 40%, over a ten year period it was 48%, over a 15 year period it was 60%, and over a 20 year period it was 65%. We see that there was a steep drop in survival at the very beginning of nearly 12% which includes grafts that were rejected straight after being transplanted; the incidence proportion then increases linearly over time.

Figure 4.1.1 Kaplan-Meier Survival Curve



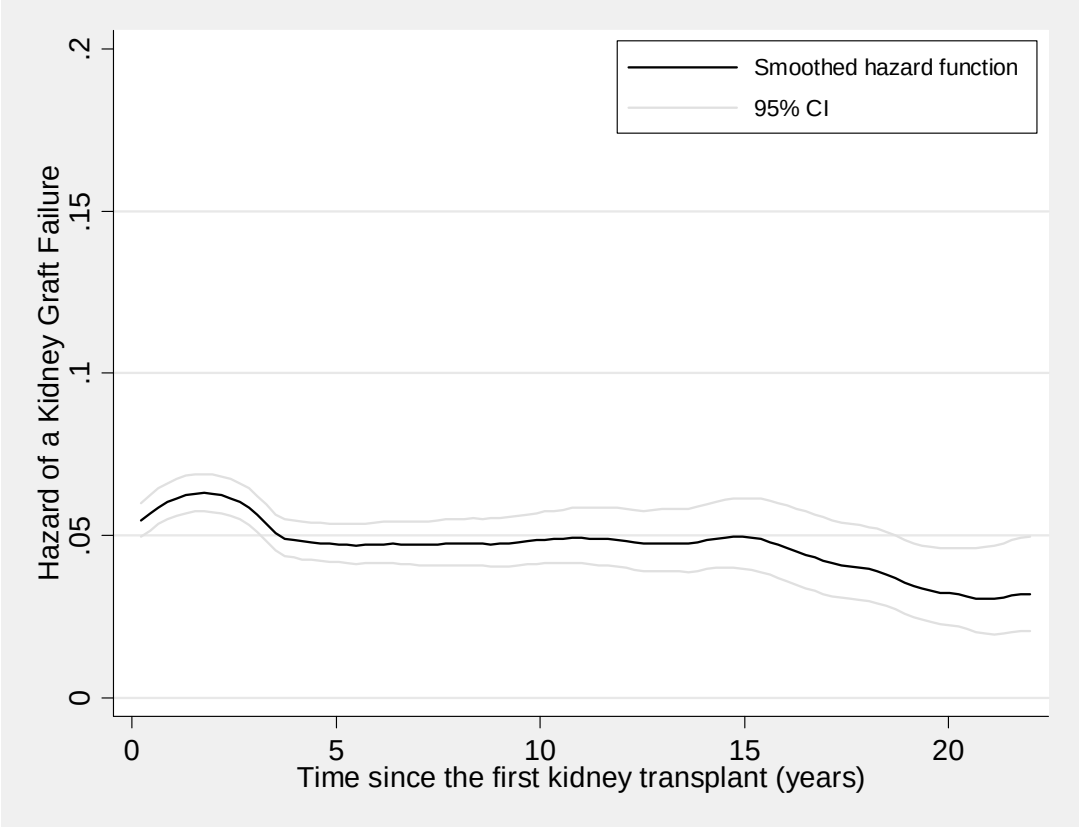
From Table 4.1.1 we see that the median survival time for a first graft was 11.77 years (10.5, 12.8). At 3.2 years 75% of the first grafts had not failed and at 28.2 years 25% of the first grafts had not failed.

Table 4.1.1 Quartiles of Survival Time

Time at Risk	Incidence Rate	No. of subjects	Survival Times (in years)			95% CI 50% Survival
			25%	50%	75%	
12891	0.063	1684	3.22	11.76	28.22	(10.54, 12.79)

In Figure 4.1.2 we see that the hazard of a transplant failure increased in the first year and then was relatively constant over the remaining life of the first kidney graft.

Figure 4.1.2 Smoothed Hazard Estimate of a First Kidney Graft Failure



The Epanechnikov Kernel function is used to smooth the hazard estimate

4.2 Identifying the risk factors for a first kidney transplant failure

A Cox Proportional Hazards model was fitted with the outcome variable being 'time to first kidney transplant failure from the date of first transplant'. The covariates considered for our model were first post-first transplant cancer, age-group at first transplant, duration of the first dialysis, gender, race, primary disease category, state of residence at the first renal replacement therapy, first donor type and era of first treatment. The 'first post-first transplant cancer' was a binary time-varying covariate which took on the value 1 from the time during the life of the first kidney graft when the person was diagnosed with cancer and zero at all other times. (See Appendix 1.3 for description of the covariates).

Since we were interested in estimating the additional effect of cancer adjusting for other covariates, we excluded the cancer variables from the variable selection process and brought them in after the preliminary main effects model containing all other covariates associated with the outcome had been fitted. We use the same model building process as described in Section 3.2.

Table 4.2.1 shows the results of the univariable analysis. Based on the overall p-value the covariates chosen were age, duration of the first dialysis, race, disease category, state of residence at the first RRT, first donor type and era of first treatment.

Table 4.2.2 shows the results of fitting a multivariable model using the covariates chosen in the univariable analysis. Looking at the overall p-values we see that 'primary disease category' and 'duration of the first dialyses' were the two covariates with the largest p-values. We next checked for confounding by dropping the two covariates from the model and running the regression and seeing if that changes the coefficients of the remaining covariates by more than 20%. Our check revealed that they were not important confounders and so they were dropped permanently from our model.

In our reduced model 'State of residence at the first renal replacement therapy' was the only non-significant covariate. We checked for confounding by dropping this covariate and running the regression and seeing if that changes the coefficients of the other remaining

covariates by more than 20%. Our check revealed that it was not an important confounder and so was dropped permanently form our model.

We next added back, one at a time, the covariates that did not meet the criteria of 25% in the univariable analysis. The covariate added back was gender. Gender was still not significant and so we removed it permanently from the model.

We next fitted the model and tested the proportional hazards assumption. Table 4.2.3 presents the results of the test. We saw that the time-constant covariate 'first donor type' strongly violated (p-value < 0.01) the proportional hazards assumption. Thus we needed to modify the model to account for the interaction between time and 'first donor type'. From the KM graph for survival stratified on the covariate (please see Appendix 1.6) we saw that the graft survival from a living donor has a different rate of decline than the rate from a deceased donor. We initially fitted both a linear and quadratic interaction with time but the quadratic interaction was not significant and so we dropped it from the model. We also saw that the hazard of era 1990-99 relative to the era 1980-89 may have changed over the observed time. Since the other eras did not strongly violate the proportion hazards assumption we made a decision to leave it the way it was and consequently got an average effect of the hazard of 1990-99 relative to 1980-89 over the observed time.

Table 4.2.4 shows the results of retesting for violations of the proportional hazards assumption after fitting the new model with the linear interaction of 'first donor type' with time. We saw that the p-value of the global test was 0.02 and this was non-significant at the 1% level of significance. Thus there was no strong evidence to suggest that the PH assumption was strongly violated.

Table 4.2.5 presents the results of fitting the final multivariable model. The significant risk factors identified for a first kidney transplant failure were age group at the first transplant, race, era of first renal replacement therapy, and first donor type i.e. living or deceased. The effect of the first donor type on the first kidney graft survival changes over the life of the graft. A first post-first transplant cancer, however, did not increase the hazard of a first kidney graft failure.

Table 4.2.1 Univariable Analysis for Time to First Kidney Graft Failure

Covariate	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	Overall P-value
Post-transplant cancer	No	Yes	1.26	0.78	2.03	0.342
Age group at the 1st Tx Date	10 to 14	<5	0.55	0.38	0.79	0.001
		5 to 9	0.85	0.66	1.09	
		15 to 19	1.09	0.92	1.30	
		>19	0.85	0.68	1.06	
Duration of first dialysis	0 days	Q1: Less than six months	1.79	1.33	2.41	<0.001
		Q2: Six months to one year	1.87	1.38	2.53	
		Q3: One year to two years	1.87	1.38	2.54	
		Q4: Greater than two years	1.55	1.14	2.11	
Gender	Female	Male	1.00	0.87	1.14	0.964
Race	Caucasians	Indigenous and Pacific Islanders	1.87	1.49	2.35	<0.001
		Asians	0.66	0.42	1.04	
Primary Disease category	Glomerulonephritis	Cystic/Hereditary	0.76	0.58	0.99	0.035
		Reflux	0.97	0.82	1.16	
		Interstitial Nephritis	0.67	0.32	1.42	
		Congenital/Urological	0.72	0.58	0.88	
		Cancer related	0.00	0.00	2.25E 114	
		Other/Uncertain	0.92	0.72	1.18	
State of residence at the first RRT	Australia	New Zealand	1.17	0.98	1.39	0.186
		Overseas	0.83	0.43	1.61	
First donor	Deceased donor	Living donor	0.51	0.43	0.59	<0.001
Era of first treatment	1980 to 1989	1963 to 1979	1.08	0.92	1.27	<0.001
		1990 to 1999	0.61	0.51	0.73	
		2000 to 2006	0.22	0.14	0.33	

Table 4.2.2 Initial Multivariable Analysis

Covariate	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	Overall P-value
Age group at the 1st Tx Date	10 to 14	<5	0.81	0.56	1.19	0.066
		5 to 9	0.98	0.76	1.27	
		15 to 19	0.97	0.81	1.16	
		>19	0.72	0.57	0.92	
Duration of first dialysis	0 days	Q1: Less than six months	1.11	0.81	1.53	0.572
		Q2: Six months to one year	1.09	0.78	1.51	
		Q3: One year to two years	1.09	0.78	1.51	
		Q4: Greater than two years	0.93	0.66	1.32	
Race	Caucasians	Indigenous and Pacific Islanders	1.95	1.52	2.49	<0.001
		Asians	0.86	0.54	1.38	
Primary Disease category	Glomerulonephritis	Cystic/Hereditary	0.83	0.63	1.09	0.708
		Reflux	0.94	0.79	1.12	
		Interstitial Nephritis	0.92	0.43	1.95	
		Congenital/Urological	0.89	0.71	1.11	
		Cancer related	0.00	0.00	3.72E118	
		Other/Uncertain	1.09	0.85	1.40	
State of residence at the first RRT	Australia	New Zealand	1.15	0.95	1.38	0.346
		Overseas	1.12	0.52	2.40	
First donor	Deceased donor	Living donor	0.57	0.47	0.68	<0.001
Era of first treatment	1980 to 1989	1963 to 1979	0.98	0.83	1.17	<0.001
		1990 to 1999	0.71	0.59	0.86	
		2000 to 2006	0.27	0.17	0.42	

Table 4.2.3 Test of the Proportional Hazards Assumption

	rho	chi2	df	Prob>chi2
Age group at first Transplant				
<5	-0.055	2.54	1	0.111
5 to 9	0.008	0.05	1	0.825
15 to 19	-0.056	2.54	1	0.111
>19	-0.080	5.02	1	0.025
Race				
Ind. & PI	0.002	0.00	1	0.960
Asians	0.031	0.81	1	0.367
Era of first Treatment				
1963-79	0.083	5.73	1	0.017
1990-99	0.098	8.57	1	0.003
2000-06	0.018	0.28	1	0.597
First donor type				
Living	0.121	11.42	1	0.001
Post-tx cancer	-0.060	3.00	1	0.083
Global test		46.40	11	<0.001

Table 4.2.4 Test of the Proportional Hazards Assumption for model with the donor type and time interaction term

	rho	chi2	df	Prob>chi2
Age group at first Transplant				
<5	-0.047	1.83	1	0.18
5 to 9	0.016	0.20	1	0.65
15 to 19	-0.044	1.58	1	0.21
>19	-0.077	4.74	1	0.03
Race				
Ind. & PI	0.003	0.01	1	0.93
Asians	0.031	0.82	1	0.37
Era of first Treatment				
1963-79	0.075	4.70	1	0.03
1990-99	0.109	10.42	1	0.00
2000-06	0.033	0.95	1	0.33
First donor type				
Living	-0.004	0.01	1	0.90
Donor type & time Interaction term	-0.041	1.33	1	0.25
Post-tx cancer	-0.059	2.92	1	0.09
Global test		0.69	12	0.02

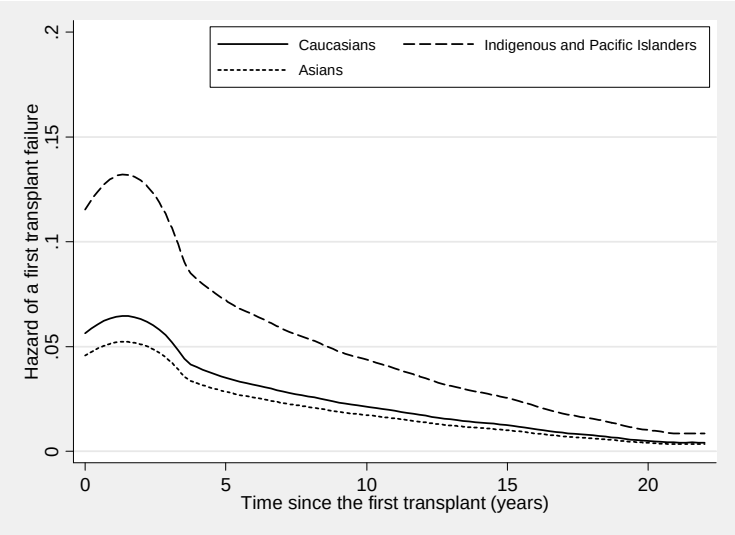
Table 4.2.5 Final Multivariable Analysis for Time to First Kidney Graft Failure

	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	Estimate	P-value	Overall P-value
Post-transplant cancer	No	Yes	1.48	0.91	2.40	0.39	0.115	0.115
Age group at the 1st Tx Date	10 to 14	<5	0.83	0.57	1.20	-0.09	0.548	0.012
		5 to 9	1.01	0.78	1.31	0.11	0.246	
		15 to 19	1.01	0.85	1.20	0.11	0.094	
		>19	0.71	0.57	0.89	-0.24	0.006	
Race	Caucasians	Indigenous and Pacific Islanders	2.05	1.62	2.59	0.55	<0.001	<0.001
		Asians	0.81	0.51	1.28	-0.38	0.017	
Era of first treatment	1980 to 1989	1963 to 1979	1.04	0.88	1.22	0.40	<0.001	<0.001
		1990 to 1999	0.71	0.59	0.86	0.03	0.758	
		2000 to 2006	0.31	0.20	0.49	-0.79	<0.001	
First donor	Deceased donor	Living donor	0.39	0.31	0.49	-0.48	<0.001	<0.001
Donor Type times Time	0	0	1.07	1.04	1.10	0.07	<0.001	<0.001

Caucasians had a hazard of a first transplant failure that was half (95% CI, 0.4-0.6) the hazard of the Indigenous and Pacific Islanders. Caucasians and Asians had hazards that were not significantly different. The less-than-five-year-olds, five-to-nine-year-olds, ten-to-fourteen- year-olds, and the fifteen-to-nineteen-year-olds had hazards that were not significantly different from each others but were significantly greater than the hazard of the greater-than-nineteen-year-olds. The renal replacement treatment (RRT) begun in earlier eras (1963-79 and 1980-89 were not significantly different from each other but were higher than the hazards in the later eras (19990-99 and 2000-06) and the hazards grew progressively less over time. A person whose graft was from a living donor had a hazard that was initially just 0.4 times (95% CI, 0.3-0.5) the hazard of a person whose graft was from a deceased donor; this hazard, however, increased by 7% each year of graft survival relative to the hazard associated with a deceased donor.

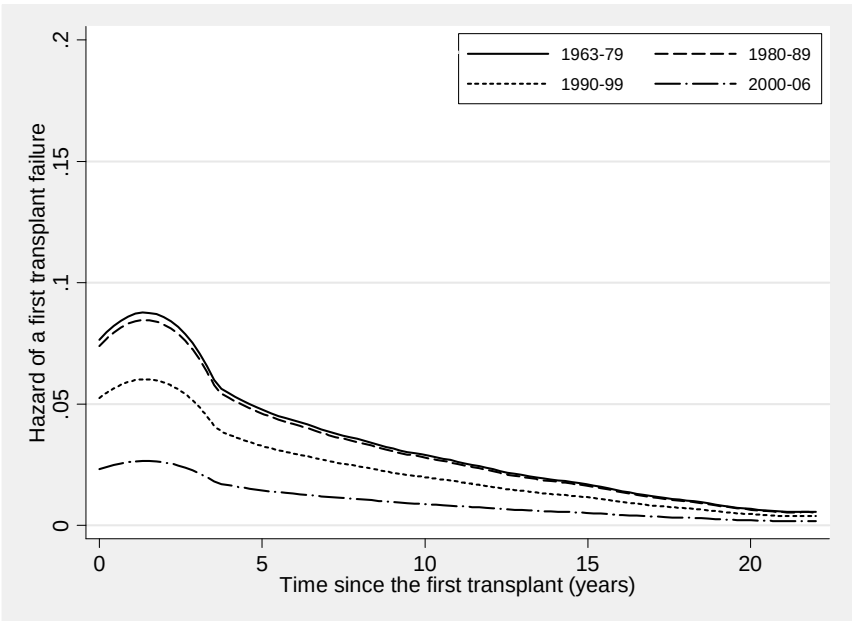
In Figure 4.2.1 we clearly see that the hazard among 'indigenous and pacific islanders' was much higher than the hazard among Asians and Caucasians.

Figure 4.2.1 Hazard of a First Kidney Transplant Failure by Race



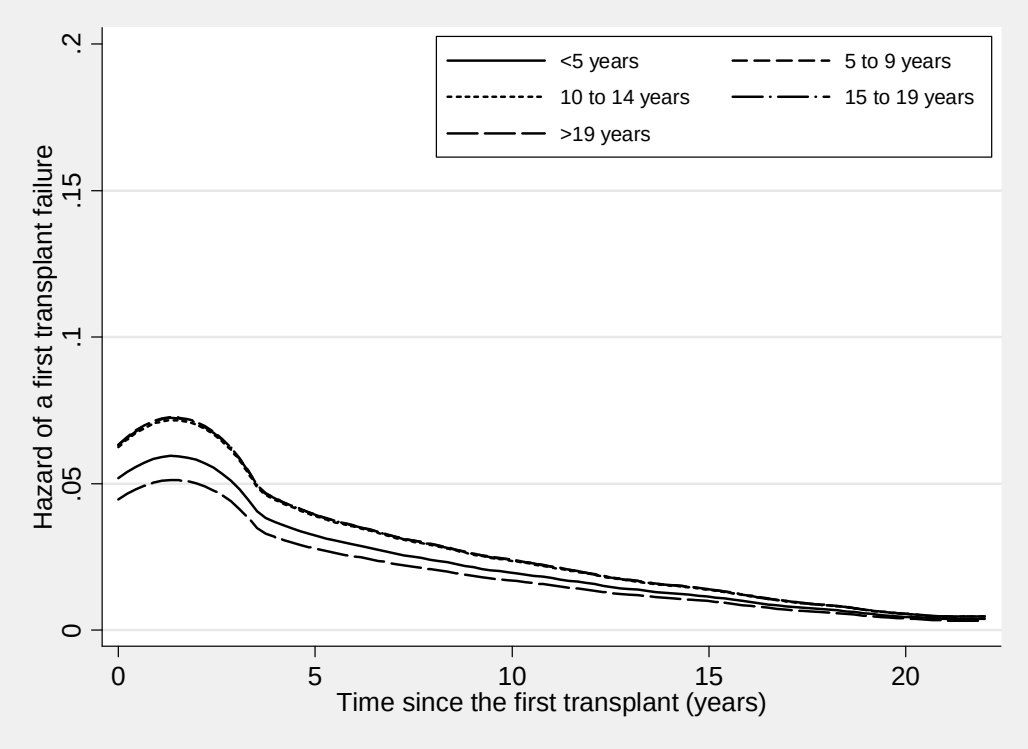
In Figure 4.2.2 we see that in the eras 1963-79 and 1980-89 the hazard of a first transplant failure seemed to be equal. This hazard became lower in the era 1990-99 and even lower in the era 2000-06.

Figure 4.2.2 Hazard of a First Kidney Transplant Failure by Treatment Decade



In Figure 4.2.3 we see that the five-to-nine-year-olds, ten-to-fourteen- year-olds, and the fifteen-to-nineteen-year-olds had the same hazard for a first kidney transplant failure which was higher than the hazard of the less-than-five-year-olds and the greater-than-nineteen-year-olds.

Figure 4.2.3 Hazard of a First Kidney Transplant Failure by Age Group at First Transplant



5. Time to Death

5.1 Mortality Rates

We begin the process of analyzing mortality by doing an exploratory analysis on mortality rates. Since this was just an exploratory analysis no formal analyses was provided.

In the calculation of mortality rates in Table 5.1, for persons who had a post-onset of ESKD cancer we, we only added to the person years at risk of death, his time after the cancer was diagnosed; the time before the cancer was not included in the analysis as that time did not increase his/her risk of dying because of a post-cancer. In Appendix 1.8 we provide a table where the time before the cancer was diagnosed was added to the sum of the person-years at risk of death to the group which did not have a cancer.

From Table 5.1 we can see that the mortality rates among persons with post-onset of ESKD cancer was more than double the mortality rates among persons with no post-onset of ESKD cancer across almost all patient characteristics. Among persons with no cancer post-onset, there was no difference in the mortality rates between males and females; however among persons with cancer post-onset there seems to be a difference with males having a higher mortality rate than females. The mortality rate among the less-than-five-year-olds who had cancer was very high compared to the other age groups. Among patients who did not have a post-onset cancer, the persons who had a transplant as their first renal replacement therapy (RRT) had a higher risk of death than patients who had a dialysis as their first RRT, whereas among persons who did have a cancer post-onset, persons with peritoneal dialysis as their first RRT had a higher risk of death than persons who had either hemodialysis or a transplant as their first RRT. Among persons who did not have a cancer post-onset and received a kidney transplant, those who received their first transplant from a living donor had a lower risk of death than those who received theirs from a deceased donor.

Table 5.1 Mortality Rate

Patient characteristic		No post-ESKD cancer			Post-ESKD cancer		
		Patient-years at risk	Number of deaths	Mortality rate (per 1000 patient years)	Patient-years at risk	Number of deaths	Mortality rate (per 1000 patient years)
Gender	Female	9874	233	23.6	345	19	55.1
	Male	12758	268	21	298	26	87.2
Age group	< 1	246	15	61	11	0	0
	1 to 4	1270	30	23.6	6	2	333.3
	5 to 9	3190	63	19.7	11	2	181.8
	10 to 14	6163	122	19.8	107	10	93.5
	15 to 19	11762	271	23	507	31	61.1
Race	Caucasians	20007	414	20.7	579	44	76
	Indigenous and Pacific Islanders	1858	70	37.7	20	1	50
	Asians	767	17	22.2	44	0	0
First treatment group	Peritoneal Dialysis	10344	227	21.9	177	19	107.3
	Hemodialysis	10539	258	24.5	434	24	55.3
	Transplant	1749	16	9.1	32	2	62.5
Era of first treatment	1963 to 1979	6628	227	34.2	290	25	86.2
	1980 to 1989	8936	169	18.9	210	13	61.9
	1990 to 1999	5696	80	14	129	6	46.5
	2000 to 2006	1371	25	18.2	13	1	76.9
State of residence at the first RRT	Australia	18267	387	21.2	485	37	76.3
	New Zealand	4105	108	26.3	135	8	59.3
	Overseas	260	6	23.1	23	0	0
First donor	Deceased donor	14529	305	21	389	30	77.1
	Living donor	7341	74	10.1	253	14	55.3

Table 5.1 Mortality Rate (continued)

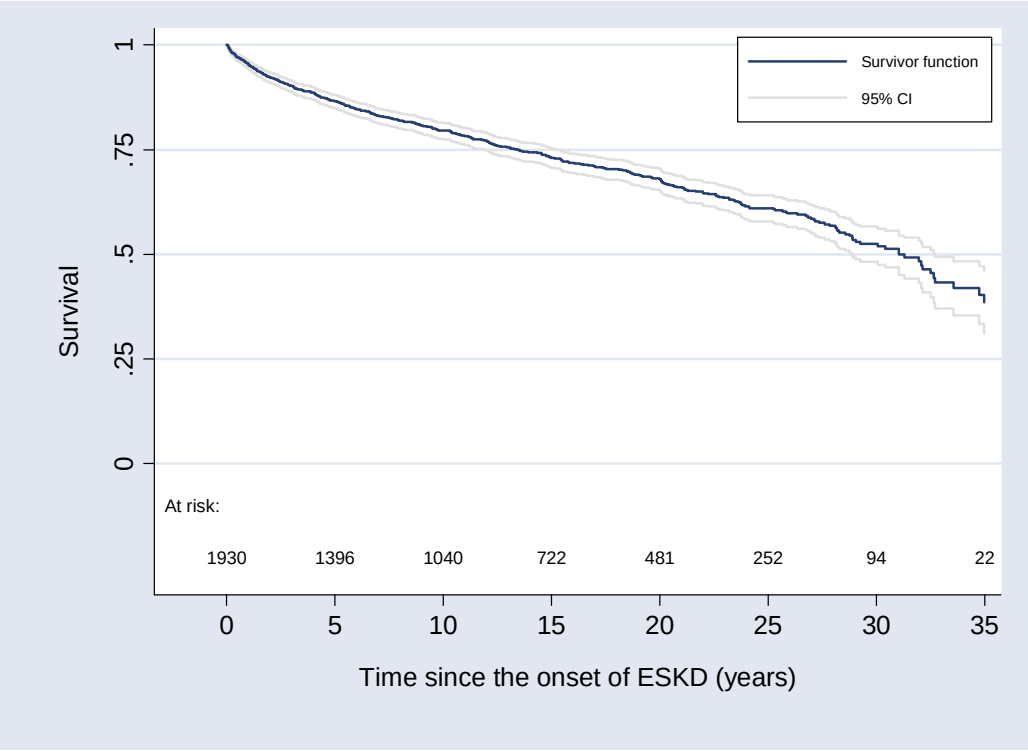
Primary Disease category	Glomerulonephritis	8636	207	24	331	23	69.5
	Cystic/Hereditary	1953	41	21	41	2	48.8
	Reflux	5396	101	18.7	211	10	47.4
	Interstitial Nephritis	240	9	37.5	7	1	142.9
	Congenital/Urological	3959	68	17.2	46	6	130.4
	Cancer related	69	3	43.5	.	0	.
	Other/Uncertain	2378	72	30.3	6	3	500

5.2 Incidence of Death among ESKD Patients

In this section we provide the incidence proportion of death as a continuous function of time using the Kaplan-Meier Failure Curve for Time to Death. The analysis variable here was 'time to death from the onset of ESKD'. Using a KM curve to calculate the incidence proportion was quite necessary here as there were, expectedly, a large number of censored observations. The total analysis time at risk for death was 24,631.12 years for the 1,930 subjects we have in our data and 546 patients died (the time includes the time before cancer for patients who did get a post-onset cancer). Thus the incidence rate was 22.17 per 1000 patient years.

Figure 5.2.1 provides us with incidence proportion as a continuous function of time. The incidence proportion was defined here as the proportion who die. The graph shows that over a five year period from the onset of ESKD the incidence proportion was 15%, over a ten year period it was 20%, over a 15 year period it was 30%, over a 20 year period it was 35%, over a 30 year period it was 50%, and over a 35 year period it was 60%. Thus the incidence proportion increases at a constant rate over time.

Figure 5.2.1 Kaplan-Meier Survival Curve



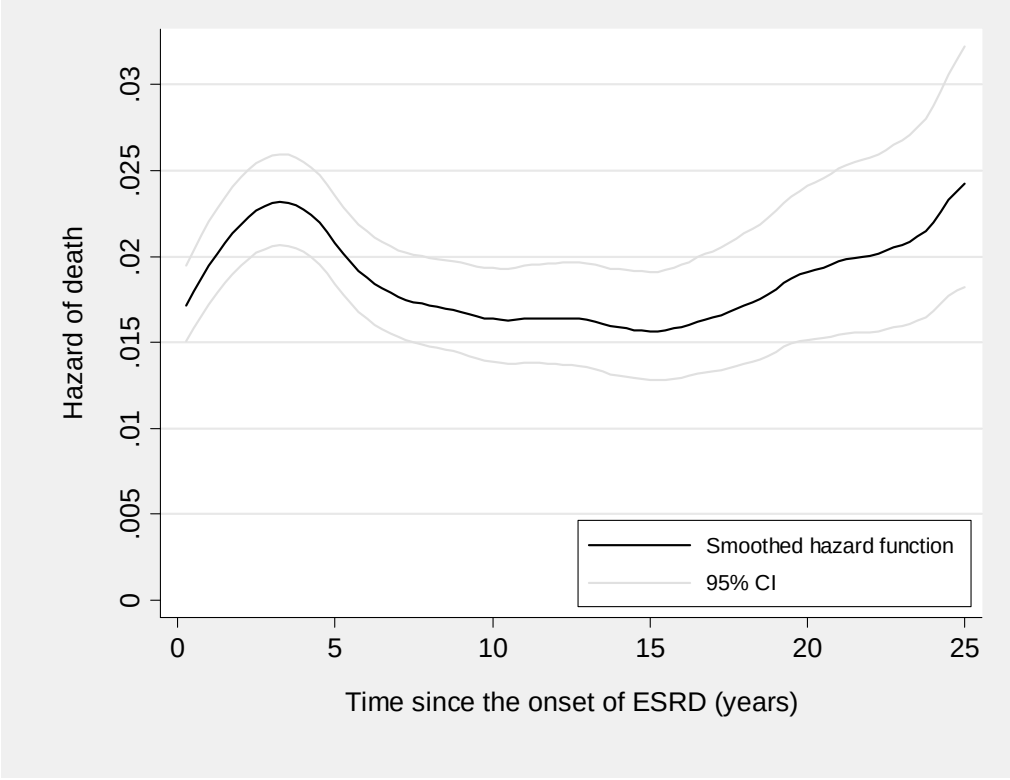
From Table 5.2.1 we see that the median survival time was 31 years (95% CI, 28.9-32.7) and at 13 years 75% of the patients were still alive.

Table 5.2.1 Quartiles of Survival Time

			Survival Times			
Time at Risk	Incidence Rate	No. of subjects	25%	50%	75%	95% CI 50% Survival
24631.12	0.0222	1930	13.42	31.04		(28.89, 32.70)

In Figure 5.2.2 we see that the hazard of death was relatively constant.

Figure 5.2.2 Smoothed Hazard Estimate Graph



The Epanechnikov Kernel function is used to smooth the hazard estimate

5.3 Identifying Cancer as a Risk Factor for Survival

A Cox Proportional Hazards model was fitted with the outcome variable being 'time to death'. The covariates considered for our model were pre-onset of ESKD cancer flag, first post-onset of ESKD cancer, gender, race, age group at onset of ESKD, treatment type (time-varying and categorical), era of first treatment, and primary disease category. The 'first post-onset of ESKD cancer' was a binary time-varying covariate which took on the value of one from the time during ESKD when the person was diagnosed with his/her first post-onset cancer, and zero otherwise. For the univariable analysis we used the binary form of the time-varying covariate 'treatment type' which took on the value one after a person has his/her first kidney transplant, and zero otherwise. (See Appendix 1.4 for description of the covariates).

Since we were interested in estimating the additional effect of cancer adjusting for other covariates, we excluded the cancer variable from the variable selection process and brought them in after the preliminary main effects model containing all other variables associated with the outcome had been fitted. We used the same model building process as described in Section 3.2.

Table 5.3.1 shows the results of the univariable analysis. Based on the overall p-value, the only covariate not chosen for the multivariable analysis was gender.

Table 5.3.2 shows the results of fitting a multivariable model using the covariates chosen in the univariable analysis. Instead of using the binary time varying form of 'treatment type' we used the same categorical time varying form used in the 'time to first post-onset of ESKD cancer' analysis in Section 3.2. Looking at the overall p-values we see that all the covariates were significant at the 5% level and so we did not exclude any covariate from our model.

'Gender' was added into the model and the new model was estimated but 'gender' was still not significant and so was excluded.

Table 5.3.1 Univariable Analysis for Time to Death (from onset of ESKD)

Covariate	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	Overall P-value
Patients with at least one pre-ESKD cancer	No	Yes	2.15	0.89	5.19	0.089
Post-onset of ESKD cancer	No	Yes	3.81	2.76	5.25	<0.001
Gender	Female	Male	0.91	0.77	1.08	0.271
Race	Caucasians	Indigenous and Pacific Islanders	1.72	1.34	2.21	<0.001
		Asians	0.93	0.57	1.51	
Primary Disease category	Glomerulonephritis	Cystic/Hereditary	0.88	0.63	1.22	0.005
		Reflux	0.79	0.63	0.99	
		Interstitial Nephritis	1.66	0.88	3.12	
		Congenital/Urological	0.75	0.58	0.98	
		Cancer related	1.65	0.53	5.15	
		Other/Uncertain	1.26	0.97	1.64	
Age group	5 to 9 years	< 1 years	2.42	1.38	4.25	0.022
		1 to 4 years	1.20	0.78	1.83	
		10 to 14 years	1.00	0.75	1.35	
		15 to 19 years	1.14	0.87	1.49	
Era of first treatment	1980 to 1989	1963 to 1979	1.69	1.38	2.05	<0.001
		1990 to 1999	0.64	0.49	0.83	
		2000 to 2006	0.57	0.37	0.87	
Treatment Type	Dialysis	Transplant	0.20	0.17	0.25	<0.001

Table 5.3.2 Initial Multivariable Analysis

Covariate	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	Overall P-value
Race	Caucasians	Indigenous and Pacific Islanders	1.54	1.18	2.01	0.006
		Asians	1.17	0.71	1.92	
Primary Disease category	Glomerulonephritis	Cystic/Hereditary	1.18	0.84	1.65	0.034
		Reflux	0.83	0.66	1.04	
		Interstitial Nephritis	1.89	1.00	3.58	
		Congenital/Urological	0.92	0.70	1.21	
		Cancer related	2.18	0.67	7.09	
		Other/Uncertain	1.23	0.94	1.62	
Age group at onset of ESKD	5 to 9 years	< 1 years	3.49	1.94	6.30	<0.001
		1 to 4 years	1.60	1.04	2.48	
		10 to 14 years	0.88	0.65	1.19	
		15 to 19 years	0.93	0.70	1.22	
Era of first treatment	1980 to 1989	1963 to 1979	1.87	1.52	2.30	<0.001
		1990 to 1999	0.66	0.51	0.87	
		2000 to 2006	0.59	0.38	0.92	
Treatment Type	D1: Dialysis just before first transplant	T1: 1st Tx	0.42	0.31	0.58	<0.001
		D2: Dx just after 1st Tx	3.08	2.26	4.20	
		T2: 2^{nd} Tx	0.77	0.50	1.18	
		D3+: All Dx after 2nd Tx	3.00	2.07	4.35	
		T3+: 3rd or higher Tx	0.64	0.32	1.30	

We tested whether the different dialysis periods had equal risks and similarly whether the different transplant periods had equal risks. If the transplant periods have equal risk then we shall create one covariate for transplant period. D2 risk was compared to D3+ risk to see if there was a significant difference; if there was then we look at a comparison of the two combined versus D1 and if this was also not significant then merge them into one. If both, dialysis periods and transplant periods were not different, then we shall revert back to our binary covariate for dialysis versus transplant.

From Table 5.3.3 we see that T2 was not significantly different from T3+ and so can be combined. We also see that D2 was not significantly different from D3+ and so they can be combined as well. We also see that D2+ was significantly different from D1 and T2+ was significantly different from T1. Thus we shall not combine any further.

Table 5.3.3 Test if the risks in the treatment periods were equal

		Wald ChiSq	DF	P-Value
T1 = T2 = T3+:	Tx periods have equal risk	10.888	2	0.004
T2 = T3+:	2nd Tx = 3rd+ Tx	0.264	1	0.608
D2 = D3+:	Dx just after 1st Tx risk = All Dx after 2nd Tx risk	0.035	1	0.853

		Hazard Ratio	HR LCL	HR UCL	P-Value
T1 vs D1:	1st Tx vs Dialysis just before 1st Tx	0.42	0.31	0.58	<0.001
D2+ vs D1:	All Dx after 1st Tx vs Dx just before 1st tx	3.06	2.26	4.16	<0.001
T2+ vs D1:	2nd Tx+ vs Dx just before 1st Tx	0.75	0.50	1.13	0.164

Linear Test		Wald ChiSq	DF	P-Value
T1 vs T2+:	1st Tx = 2nd+ Tx	10.700	1	0.001

We next include the cancer covariates in our chosen model, fit the model and test the proportional hazards assumption. Table 5.3.4 presents the results of the test of the PH assumption. We see that none of the covariates violate the proportional hazards assumption.

Table 5.3.4 Test of the Proportional Hazards Assumption

	rho	chi2	df	Prob>chi2
Pre-ESKD cancer	-0.002	0.00	1	0.963
Post-onset of ESKD Cancer	-0.051	1.54	1	0.215
Race				
Indigenous & PIs	0.032	0.54	1	0.463
Asians	-0.042	1.07	1	0.300
Primary Disease				
Cystic/Hereditary	-0.002	0.00	1	0.961
Reflux	0.014	0.11	1	0.736
Interstitial Nephritis	0.032	0.59	1	0.443
Cong/Urological	0.069	2.54	1	0.111
Cancer related	-0.008	0.04	1	0.835
Other/Uncertain	-0.076	3.16	1	0.075
Age-group				
<1 years	-0.046	1.17	1	0.280
1 to 4 years	-0.058	1.83	1	0.177
10 to 14 years	-0.029	0.47	1	0.493
15 to 19 years	-0.007	0.02	1	0.877
Era of first treat				
1963 to 1979	-0.117	7.58	1	0.006
1990 to 1999	-0.046	1.22	1	0.270
2000 to 2006	-0.017	0.16	1	0.686
Treatment Periods				
T1: First Tx	-0.022	0.32	1	0.571
D2+: All Dx post 1st Tx	-0.048	1.62	1	0.203
T2+: 2nd or higher Tx	-0.013	0.11	1	0.744
Global test		27.09	20	0.133

Table 5.3.5 presents the results of fitting the final multivariable model. The significant risk factors (at the 5% level) identified for death were pre-onset cancer, post-onset cancer, race, age group at onset of ESKD, era of first renal replacement therapy and treatment type. Primary disease category was significant at the 10% level.

Table 5.3.5 Final Multivariable Analysis for Time to Death (from onset of ESKD)

Covariate	Reference Level	Level	Hazard Ratio	HR LCL	HR UCL	P-value	Overall P-value
Patients with at least one pre-ESKD cancer	No	Yes	3.83	1.25	11.79	0.019	0.019
Post-onset of ESKD cancer	No	Yes	4.41	3.16	6.15	<0.001	<0.001
Race	Caucasians	Indigenous and Pacific Islanders	1.62	1.24	2.11	0.008	0.002
		Asians	1.02	0.62	1.69	0.399	
Primary Disease category	Glomerulonephritis	Cystic/Hereditary	1.20	0.86	1.68	0.780	0.055
		Reflux	0.84	0.67	1.06	0.040	
		Interstitial Nephritis	1.87	0.99	3.56	0.094	
		Congenital/Urological	0.97	0.74	1.28	0.302	
		Cancer related	1.07	0.26	4.43	0.911	
		Other/Uncertain	1.30	0.98	1.71	0.399	
Age group at onset of ESKD	5 to 9 years	< 1 years	3.33	1.84	6.01	<0.001	<0.001
		1 to 4 years	1.60	1.03	2.47	0.227	
		10 to 14 years	0.85	0.63	1.15	<0.001	
		15 to 19 years	0.88	0.67	1.16	<0.001	
Era of first treatment	1980 to 1989	1963 to 1979	1.90	1.54	2.33	<0.001	<0.001
		1990 to 1999	0.62	0.47	0.81	<0.001	
		2000 to 2006	0.55	0.35	0.85	0.003	
Treatment Type	D1: Dialysis just before first transplant	T1: First transplant	0.41	0.30	0.57	<0.001	<0.001
		D2+: All dialysis after first transplant	3.01	2.22	4.08	<0.001	<0.001
		T2+: Second or higher transplant	0.76	0.51	1.15	0.194	0.194

Patients with at least one pre-onset cancer had a hazard of death that was 3.8 times (95% CI: 1.3-11.8) the hazard of patients with no pre-onset cancer. Patients with at least one post-onset-of-ESKD cancer had a hazard of death that was 4.4 times (95% CI: 3.2-6.2) the hazard of patients with no post-onset-of-ESKD cancer.

Caucasians had a hazard of death that was 0.6 times (95% CI: 0.5-0.8) the hazard of Indigenous and Pacific Islanders. Caucasians and Asians had hazards that were not significantly different from each others.

None of the primary disease categories had hazards that were significantly different from each others.

The hazard of death for the '<1 years' age group was 3.3 times that of the '5 to 9 years' age group (95% CI: 1.8-6.0). The hazard for the '1 to 4 years' age group was 60% higher than that of the '5 to 9 years' age group (95% CI: 1.0-2.5). The hazard of death for the '5 to 9 years' age group was not significantly different from that of the older age groups.

The hazard of death for RRT begun in earlier eras was significantly higher than the hazard for RRT begun in later eras. The hazard of death for persons with the onset of RRT in the period '1963-79' was nearly double that of the persons with an onset in the period '1980-89' (HR 1.9; 95% CI: 1.5-2.3). The hazard in the period '1990-99' and '2000-06' were 38% and 45% respectively lower than in the period '1980-89'.

The hazard of death for patients with their first donor graft in them was 0.4 times (95% CI: 0.3-0.6) the hazard for patients who were on dialysis and have not had a transplant. The hazard of death for patients on dialysis at any time after the first graft failure was 3 times (95% CI: 2.2-4) the hazard for patients who were on dialysis but had not had a transplant. The hazard of death for patients with their second donor graft in them was not significantly different from the hazard for patients who were on dialysis and have not had a transplant.

1. Project 2

The risk of cancer and survival relative to that in the general population in patients who were under the age of 20 years at the onset of ESKD

1.1 Objectives

There were several objectives for this project. These objectives are stated below.

1. To determine the relative survival of patients with ESKD.

   Patients who have ESKD could die of causes that were not related to their disease. However the interest was in determining what proportion of the population with ESKD who die after a certain period of follow-up, died because he or she acquired ESKD. The information we get was probability of dying before t periods of follow-up from causes directly associated with the disease/s or the incidence proportion of deaths caused by the disease/s in the t periods of follow-up.

2. To determine the Standardized Mortality Ratio (SMR) of patients with ESKD.

   Here too we were interested in determining the cause-specific mortality over a certain period of time; the difference here is that as opposed to cause-specific incidence proportion in relative survival, here we get how many times greater is the mortality among patients in our population compared to the mortality in the general population. We count the number of deaths in a certain period in our observed sample of patients with ESKD and divide that by the amount of deaths we would expect. We shall determine SMR for the all the ESKD patients as well as subgroups of ESKD patients based on age at onset of disease, gender, era of onset of the disease, and whether he/she had a post-onset-of-ESKD cancer. The information we get was how much more likely a person was to die of ESKD or ESKD and cancer than he/she was to die from any other cause.

3. To determine the Standardized Incidence Ratio (SIR) of cancer for patients with ESKD.

   Here we were interested in determining cause-specific morbidity. Here we count the number of post-onset cancers in a certain period in our observed sample of patients with ESKD and divide that by the number of post-onset cancers we would expect. We shall determine SIR for all ESKD patients as well as subgroups of patients based on age at onset of disease, gender, era of onset of the disease, and cancer site. The information we get was how much more likely a person with ESKD was to get a post-onset cancer compared to the general population.

1.2 Data

All patients in whom renal replacement therapy (RRT) started in Australia or New Zealand and who were registered in the ANZDATA registry and who were under the age of 20 when this therapy began were included in the study. For relative survival and standardized mortality ratios, those patients whose RRT started between January 1 1980 and December 31 1996 were included in the study; there were 1,495 patients in the dataset out of which 67 or 4.5 had cancer during ESKD. For standardized incidence ratios of cancer, those patients whose RRT began in Australia between January 1 1982 and December 31 2005 and those patients whose RRT began in New Zealand between January 1 1980 and December 31 2004 were included in the study.

The dataset for the population information used to determine Relative Survival and Standardized Mortality Rates consisted of population, deaths, rate of deaths and probability of death by age, sex, year (1980 to 2007) and country (New Zealand and Australia). These data came from the Australian Bureau of Statistics (ABS, 2009). This dataset was merged with the ESKD data to calculate the two measures.

The cancer incidence data for Australia came from the Australian Institute of Health and Welfare which is Australia's national agency for health and welfare statistics and information (AIHW, 2009). The dataset consists of the numbers of cancers by year (1982 to 2005), gender, age group and cancer site. The cancer incidence data for New Zealand came from New Zealand Cancer Registry (NZCR, 2009). The dataset consists of the number of cancers by year (1980 to 2004), gender, age group and cancer site. These datasets was merged with the population dataset for Australia and New Zealand to calculate the incidence in the general population in Australia and New Zealand and further merged with the ESKD data to calculate Standardized Incidence Ratios for cancers for the Australian and New Zealand population.

The inclusion criteria for this project were:

- Residents of Australia or New Zealand

- Renal replacement therapy started in Australia or New Zealand between 1st of January 1980 and 31st December 2006 for the Relative Survival and Standardized Mortality Ratio analyses
- Renal replacement therapy started in Australia between 1982 and 2005 and in New Zealand between 1980 and 2004 for the Standardized Incidence Ratio analysis
- Under the age of 20 when renal replacement therapy began

2. Relative Survival

In this section we analyze relative survival in patients with ESKD. For details regarding the calculation of relative survival please see Appendix 2.1.

There 1,495 patients whose onset of ESKD was between January 1 1980 and December 31 2006 and out of these patients 290 died on or before December 31 2006.

As can be seen in Table 2.1 there were 1495 patients at the start of the first year of follow-up of which 52 patients dropped out either because they were censored or they were lost to follow-up. Thus the effective number at risk during the first year of follow-up was 1469 (1495 – 52/2). Of the patients who were at risk, 48 died, and so the interval-specific observed survival was 0.967 {(1469 – 48)/1469). The interval-specific expected survival is just the average of the expected survival of all patients in their first year of follow-up given their age at that time, the calendar year at that time, their country and gender; i.e. say a male patient in Australia who is 25 years old in 1985 in which his first year of follow-up began has an expected probability of survival that is the probability of survival of all 25-year old males in Australia in 1985. The interval-specific relative survival is the interval-specific observed survival divided by the interval-specific expected survival; for the first year of follow-up this would be 0.967/0.999 equal to 0.968. For the first year of follow-up the cumulative observed survival and the cumulative expected survival are just their interval-specific counterparts. From the second period of follow-up onwards, the cumulative observed survival and the cumulative expected survival are just their interval-specific counterparts for the period of follow-up multiplied by their respective values in the previous period. For example for the second period of follow-up the cumulative observed survival 0.9447 is just 0.9766 multiplied by 0.967. The cumulative observed survival at any period of follow-up *t* tells us the probability of surviving *t* periods of follow-up. The cumulative relative survival at any period *t* tells us how many times less likely a patient is to survive *t* periods of follow-up compared to a person from the general population who is similar in terms of age of the patient at the t^{th} period, gender, calendar year at t^{th} period and country. One minus the cumulative observed survival can be interpreted as the incidence proportion of deaths among ESKD patients in t periods of follow-up from any cause, while one minus the cumulative relative survival can be interpreted as the incidence proportion of

deaths among ESKD patients in *t* periods of follow-up from causes directly related to ESKD.

At the first year of follow-up, patients with ESKD were 3% less likely to survive than persons in the general population. In terms of incidence proportion, all 3% of the ESKD patients who died, died from causes directly related to ESKD. As the years of follow-up increases patients with ESKD become less and less likely to survive than persons in the general population. At the 20^{th} year of follow-up, patients with ESKD were 25% less likely to survive than persons in the general population. In terms of incidence proportion, in the 20 years of follow-up, of the 27% of deaths among ESKD patients from all causes, 25% were directly related to ESKD and thus only 2% were from other causes. Figure 2.1.1 shows this steady decline in cumulative relative survival and because our sample size was large the confidence interval band was tight around it.

Figure 2.1.2 displays the cumulative relative survival of patients with ESKD split by the age group they were in at the onset of ESKD. We see that the cumulative relative survival of patients who were infants at the onset of ESKD had a lower initial level than the other age groups i.e. the cumulative relative survival at the first year of follow-up was lower than that of the other age groups; the rate of decline in the first seven years was steeper than the other age groups i.e. their chance of survival relative to the general population grew worse over the seven years of years of follow-up at a faster rate than that of the other age groups; after seven years the relative survival did not decline further. The cumulative relative survival of patients who were 1 to 4 years old at the onset of ESKD was high (nearly one) at the first year of follow-up, but declined quite rapidly in the first four years of follow-up to around 0.85; after four years the decline was gradual. The cumulative relative survival of the older age groups had initial levels that were close to one, but declined gradually over the follow-up years.

Figure 2.1.3 shows that the cumulative relative survival of females was lower than that of males for all the follow-up years. Thus during the period of observation females had a higher probability of dying from causes directly related to ESKD than males. The mortality rate of females was lower than that of males in the general population but in the ESKD

population this rate was the same (as seen from the results of Project 1); thus the mortality rate for females from causes directly related to ESKD was higher.

Table 2.1 Life Table Estimates of Patient Survival for Patients with ESKD

Interval	L	D	W	Effective number at risk	Interval-specific observed survival	Cumulative observed survival	Interval-specific expected survival	Cumulative expected survival	Interval-specific relative survival	Cumulative relative survival	Lower 95% CI for CR	Upper 95% CI for CR
0.0 - 1.0	1495	48	52	1469	0.96732	0.96732	0.99925	0.99925	0.96805	0.96805	0.95759	0.97601
1.0 - 2.0	1395	32	52	1369	0.97663	0.94471	0.99938	0.99863	0.97723	0.94601	0.93291	0.95666
2.0 - 3.0	1311	20	54	1284	0.98442	0.93000	0.99938	0.99801	0.98504	0.93185	0.91730	0.94398
3.0 - 4.0	1237	17	68	1203	0.98587	0.91686	0.99937	0.99738	0.98649	0.91926	0.90351	0.93261
4.0 - 5.0	1152	20	60	1122	0.98217	0.90051	0.99937	0.99675	0.98280	0.90345	0.88627	0.91823
5.0 - 6.0	1072	18	52	1046	0.98279	0.88502	0.99936	0.99611	0.98342	0.88847	0.87002	0.90453
6.0 - 7.0	1002	14	48	978	0.98569	0.87235	0.99933	0.99544	0.98635	0.87634	0.85689	0.89341
7.0 - 8.0	940	10	48	916	0.98908	0.86282	0.99930	0.99475	0.98977	0.86738	0.84717	0.88519
8.0 - 9.0	882	10	51	857	0.98832	0.85275	0.99928	0.99404	0.98903	0.85787	0.83685	0.87648
9.0 - 10.0	821	10	48	797	0.98745	0.84205	0.99928	0.99332	0.98816	0.84771	0.82583	0.86717
10.0 - 11.0	763	11	47	740	0.98513	0.82953	0.99927	0.99260	0.98585	0.83571	0.81281	0.85618
11.0 - 12.0	705	12	48	681	0.98238	0.81491	0.99925	0.99185	0.98312	0.82161	0.79751	0.84324
12.0 - 13.0	645	10	47	622	0.98391	0.80180	0.99922	0.99107	0.98468	0.80902	0.78384	0.83171
13.0 - 14.0	588	6	41	568	0.98943	0.79332	0.99919	0.99027	0.99022	0.80111	0.77519	0.82452
14.0 - 15.0	541	6	57	513	0.98829	0.78403	0.99917	0.98945	0.98911	0.79239	0.76561	0.81663
15.0 - 16.0	478	8	40	458	0.98253	0.77034	0.99915	0.98861	0.98337	0.77921	0.75108	0.80474
16.0 - 17.0	430	3	45	408	0.99264	0.76467	0.99911	0.98773	0.99352	0.77417	0.74542	0.80028
17.0 - 18.0	382	3	23	371	0.99190	0.75847	0.99909	0.98683	0.99280	0.76860	0.73912	0.79539
18.0 - 19.0	356	5	40	336	0.98512	0.74719	0.99908	0.98592	0.98603	0.75786	0.72699	0.78594
19.0 - 20.0	311	5	47	288	0.98261	0.73419	0.99908	0.98501	0.98352	0.74537	0.71274	0.77508

Figure 2.1.1 Cumulative Relative Survival of Patients with ESKD

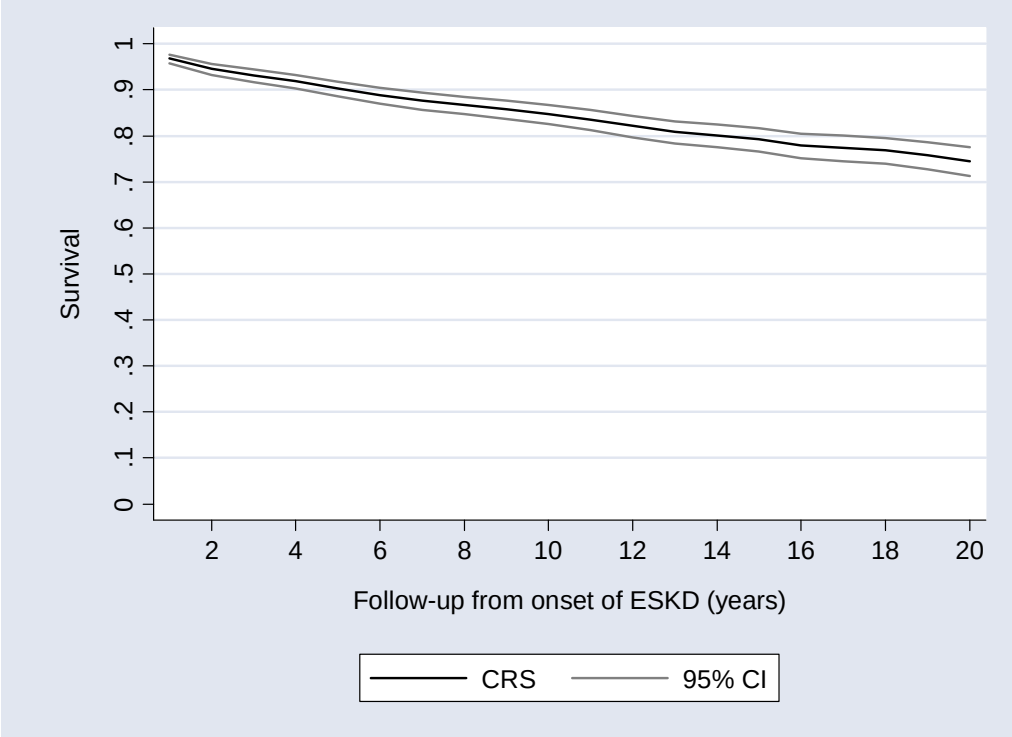


Figure 2.1.2 Cumulative Relative Survival of Patients with ESKD by Age Group

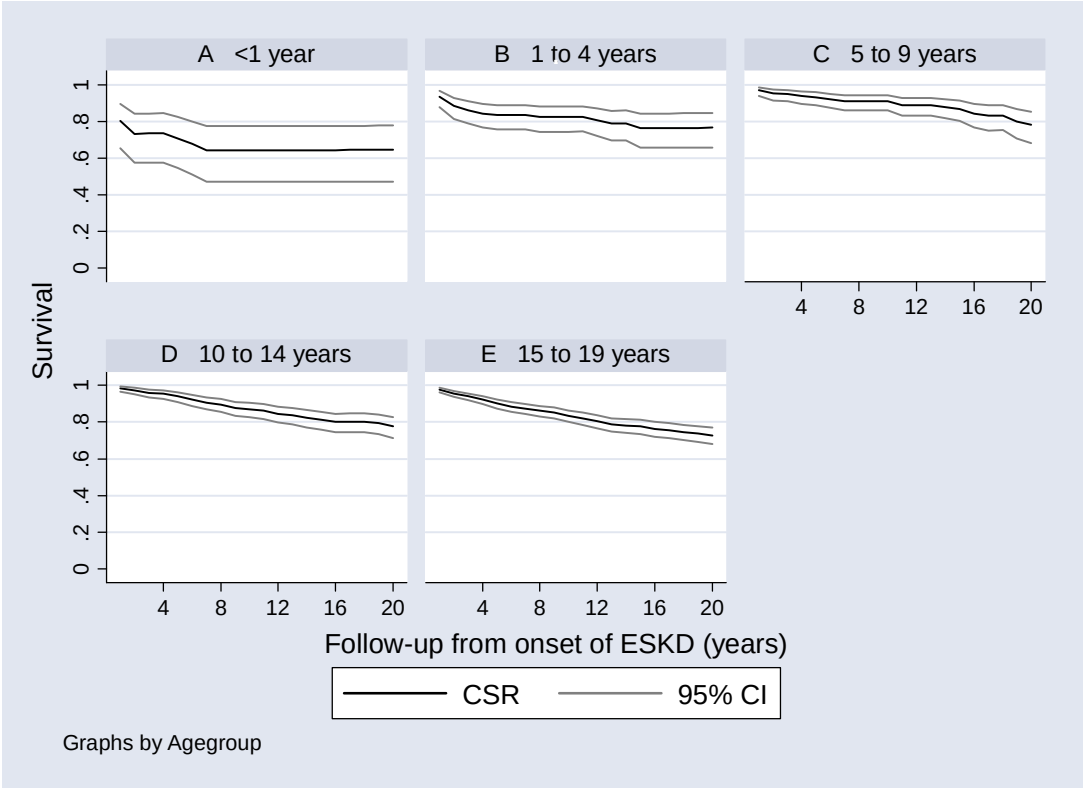


Figure 2.1.2 Cumulative Relative Survival of Patients with ESKD by Age Group (continued)

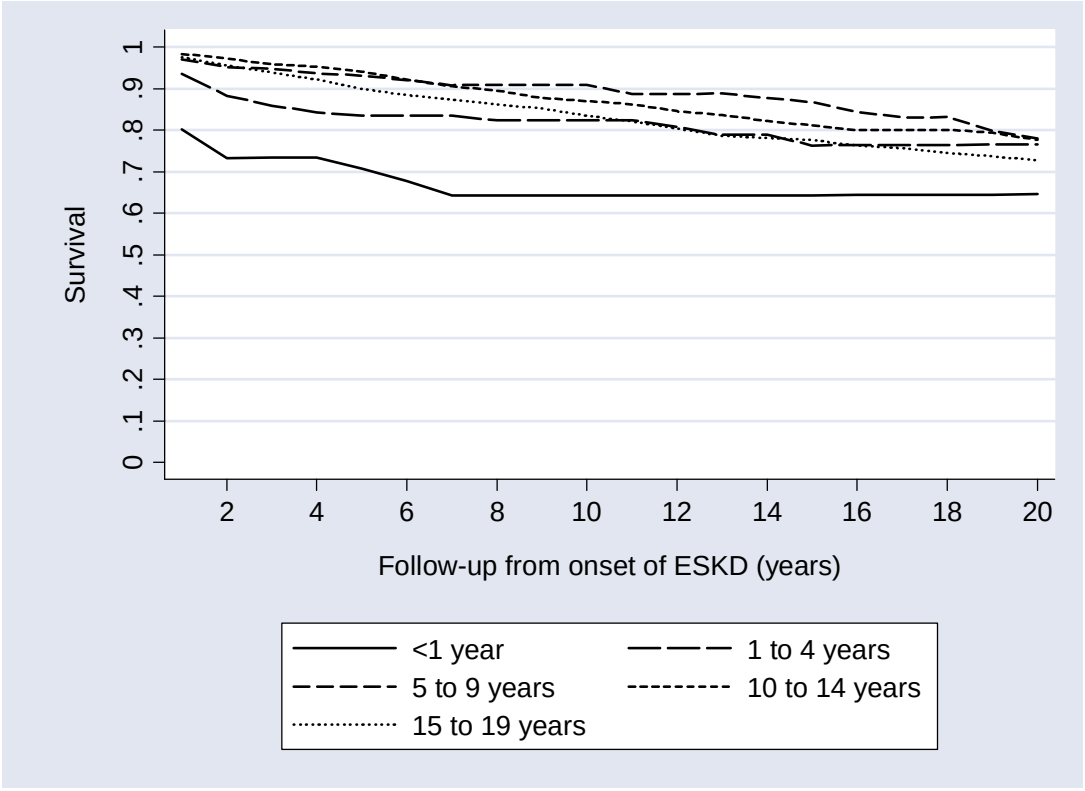
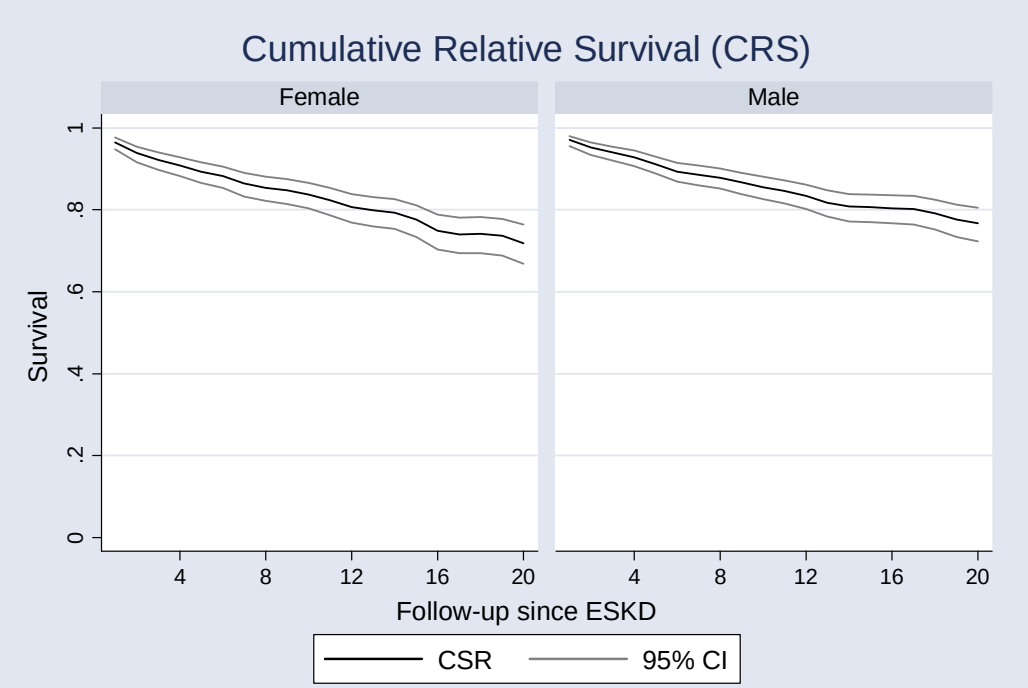
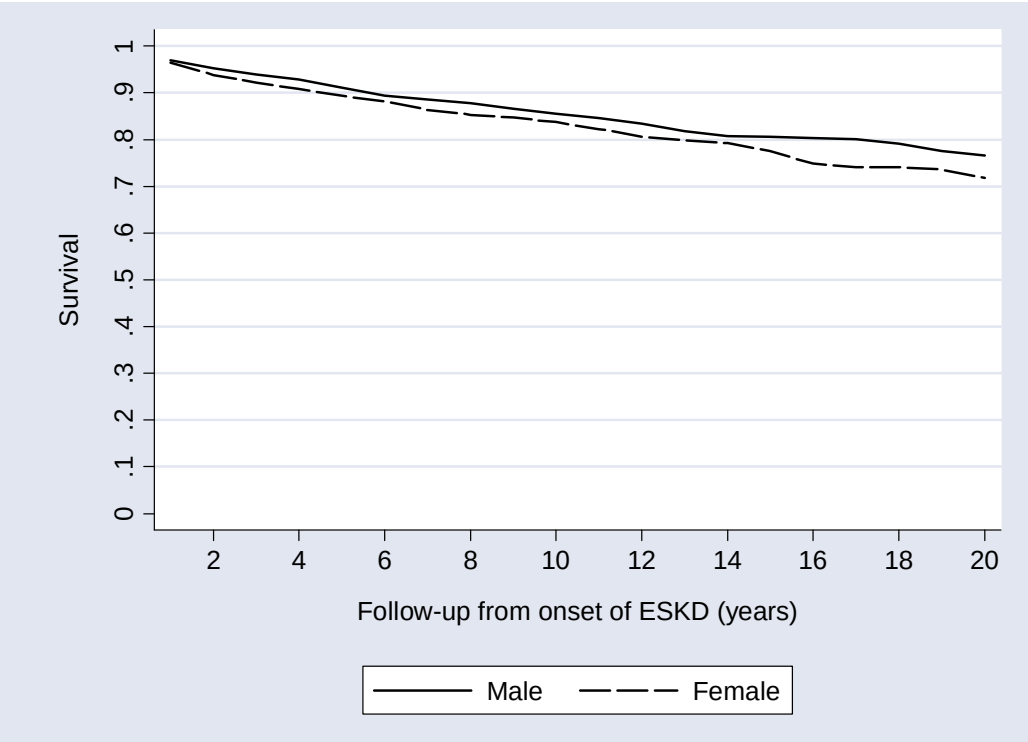


Figure 2.1.3 Cumulative Relative Survival of Patients with ESKD by Gender





3. Standardized Mortality Ratios

There were 1,495 patients whose onset of ESKD was between January 1 1980 and December 31 2006 and out of these patients 290 died on or before December 31 2006. In that same period 67 patients with ESKD also had cancer and out of these 20 patients died on or before December 31 2006.

The SMR for all patients with ESKD was calculated in the following way. The column 'observed deaths' was derived from the ESKD dataset for the period January 1 1980 to December 31 2006. The 'expected deaths' and the 'standardized mortality ratio' were calculated in the following way. Each person's follow-up period in the ESKD dataset from the onset of ESKD to the time either he/she dies or was censored was split by age group and calendar year and then sorted by country (New Zealand ad Australia), gender, age group and calendar year. A separate population dataset provides the death rates (per 100,000 person-years) among patients in strata by country, gender, age group and calendar year; for example the dataset has for males in Australia who were infants in the year 1980, a rate of x deaths per 100,000 person-years. This population dataset was sorted by country, gender, age group and calendar year and merged with our split ESKD dataset. We then count the number of patients in each stratum and then multiply that number by the rate of deaths per person-year to come up with the expected number of deaths in each stratum. The expected deaths in each of these strata were then added together to come up with the 'expected deaths'. Standardized Mortality Ratio (SMR) was derived by dividing the 'observed deaths' with the 'expected deaths'.

The SMR for a subgroup, say '1 to 4 years', was derived by taking a subset from the ESKD database whose onset of ESKD began when they were between one and four years old. The steps to calculate the SMR for that subgroup then follows the same process as was used to calculate the SMR for all patients with ESKD.

The results of Table 3.1 are shown graphically in Figure 3.1. From this table and figure we see that the SMR of patients with ESKD and a post-onset cancer (SMR, 76.1; CI, 49.1-118) was much higher than the SMR of patients with effectively just ESKD (SMR, 23.7; CI, 21.1-26.6). Thus the risk of dying for patients with a post-onset cancer and ESKD from

causes directly related to cancer and/or ESKD was much higher than the risk of dying for patients with ESKD from causes directly related to ESKD.

Table 3.1 Standardized Mortality Ratios

Population	Level	Observed Deaths	Expected Deaths	Standard Mortality Ratio (with CIs)
All patients with ESKD		290	12.25	23.68 (21.11, 26.57)
All patients with ESKD	Age at onset of ESKD			
	Infant	15	0.31	48.89 (29.47, 81.10)
	1 to 4 years	27	0.66	40.62 (27.86, 59.24)
	5 to 9 years	34	1.14	29.86 (21.33, 41.79)
	10 to 14 years	65	3.09	21.06 (16.51, 26.85)
	15 to 19 years	149	7.05	21.13 (18.00, 24.82)
All patients with ESKD	Gender			
	Male	149	9.27	16.08 (13.69, 18.88)
	Female	141	2.98	47.35 (40.14, 55.85)
All patients with ESKD	Era at onset of ESKD			
	1980-89	180	7.95	22.66 (19.58, 26.22)
	1990-99	84	3.66	22.92 (18.51, 28.39)
	2000-06	26	0.64	40.79 (27.78, 59.91)

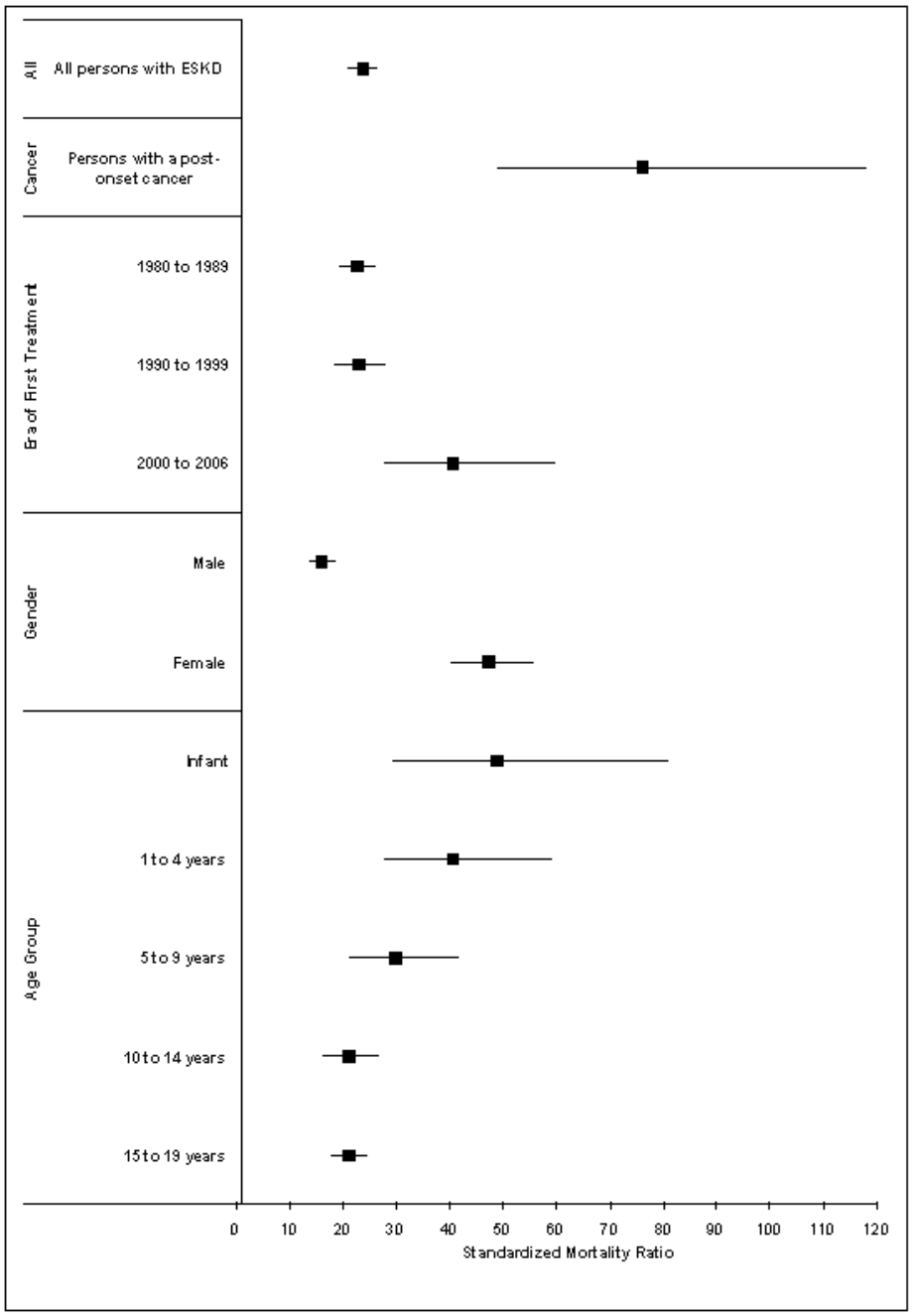
The SMRs of infants and 1 to 4 year olds were much higher than that of the older age groups. The risk of dying from causes directly related to ESKD was much higher among children whose onset of ESKD began at an age below five years than among children whose onset of ESKD began when they were five years and older.

The SMR of females (SMR, 47.4; CI, 40.1-55.9) was significantly higher than that of males (SMR, 16.1; CI, 13.7-18.9). Female mortality in the general population was significantly lower than male mortality. However among the ESKD patients the mortality rates of males and females was the same (as seen in Project 1). Thus the SMR of females was higher than the SMR of females.

The SMR for the 1980s (SMR, 22.7; CI, 19.6-26.2) and 1990s (SMR, 22.9; CI, 18.5-28.4) were not significantly different (the CIs overlap). The SMR of the era 2000-06 appears higher but could be in part just a data artifact; persons were censored in December 31 2006 and so the age distribution of children in this era may be biased towards their being more younger children (relative to the other eras), whom, as we have already seen, have higher

SMRs. Another possible reason was that the diagnosis of ESKD was much better in the era 2000-06. Thus the worst cases were being diagnosed with ESKD before they died and so worst cases were in the ESKD database. The mortality of the worst cases was now being captured by the ESKD database and this in part has made the SMR in this era appear as though it had increased.

Figure 3.1 Standardized Mortality Ratios



4. Standardized Incidence Ratios

There were 1,060 patients whose onset of ESKD began in Australia between January 1 1982 and December 31 2005, and out of these patients 54 got a post-onset-of-ESKD cancer on or before December 31 2005. There were 275 patients whose onset of ESKD began in New Zealand between January 1 1980 and December 31 2004, and out of these patients six got a post-onset-of-ESKD cancer.

The SIR of cancer for all patients with ESKD was derived in the following way. The 'observed number of cancers' was just the sum of the cancers in the ESKD database for Australia and New Zealand in their respective periods of observation. The 'expected number of cancers' column was derived in the following way. Each person's follow-up period in the ESKD dataset from the onset of ESKD to the time he/she gets cancer or was censored was split by age group and calendar year and then sorted by country (New Zealand ad Australia), gender, age group and calendar year. A separate population dataset provides the rates of cancer (per 100,000 person-years) among patients in strata by country, gender, age group and calendar year; for example the dataset has for males in Australia who were infants in the year 1982, a rate of x cancers per 100,000 persons-years. This population dataset was sorted by country, gender, age group and calendar year and merged with our split ESKD dataset. We then count the number of patients in each stratum and then multiply that by the rate of cancers per person-year to come up with the expected number of cancers for each stratum. The expected number in each stratum was then added up to give us the 'expected number of cancers'. The Standardized Incidence Ratio (SIR) was derived by dividing the 'observed number of cancer' with the 'expected number of cancers'.

The SIR for a subgroup, say '0 to 4 years', was derived by taking a subset from the ESKD database whose onset of ESKD began when they were between zero and four years old. The steps to calculate the SIR for that subgroup then follows the same process as was used to calculate the SIR for all patients with ESKD.

The results from Table 4.1 are displayed graphically in Figure 4.1.1. The incidence of cancer among patients with ESKD was around 9 times (SIR, 9.21; CI, 7.15-11.86) greater

than the incidence in the general population. The SIR for persons who were under the age of five when they began ESKD (SIR, 26.24; CI, 12.51-55.03) was significantly higher than the SIR of patients who were in the '10 to 14 years' age group (SIR, 6.86; CI, 3.8-13.4) and the '15 to 19 years' age group (SIR, 8; CI, 6.5-12.4) at the onset of ESKD. The SIRs of the '5 to 9 years', 10 to 14 years' and 15 to 19 years' age groups were not significantly different from each other. The SIRs of males and females were the same. The SIR of persons whose onset of ESKD began in the 1980s (SIR, 6.2; CI, 4.3, 9) was significantly lower than the SIR of persons whose ESKD began in the 1990s and 2000s. One possible reason for this seemingly counter-intuitive situation was that with the advances in treatment persons whose ESKD began in the later eras, were living longer with the disease, and so had more chance of acquiring cancer and cancers among ESKD patients occur at a higher rate relative to the general population.

Table 4.1 Standardized Incidence Ratios for Cancer

	Observed No. of Cancers	Expected No. of Cancers	Standardized Incidence Ratios (with CIs)
All	60	6.51	9.21 (7.15, 11.86)
Age at onset of ESKD			
0 to 4 years	7	0.27	26.24 (12.51, 55.03)
5 to 9 years	5	0.53	9.39 (3.91, 22.55)
10 to 14 years	11	1.60	6.86 (3.80, 12.39)
15 to 19 years	37	4.11	8.00 (6.52, 12.42)
Gender			
Male	31	3.43	9.03 (6.35, 12.84)
Female	29	3.08	9.41 (6.54, 13.55)
Era at onset of ESKD			
1980-89	28	4.52	6.20 (4.28, 8.96)
1990-99	27	1.80	14.99 (10.28, 21.85)
2000-05	5	0.19	25.79 (10.73, 61.96)

Figure 4.1 Standardized Incidence Ratios for Cancer

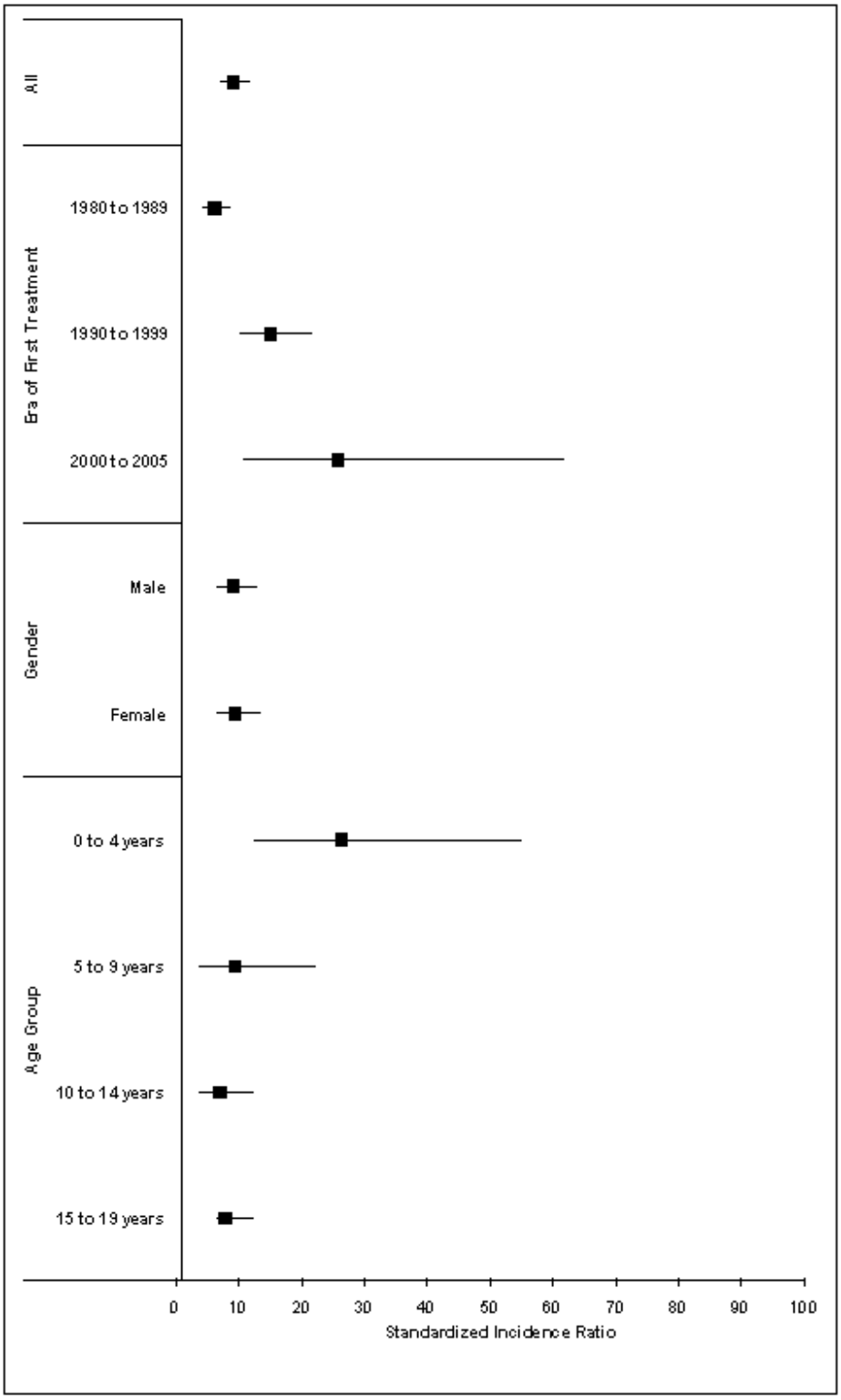


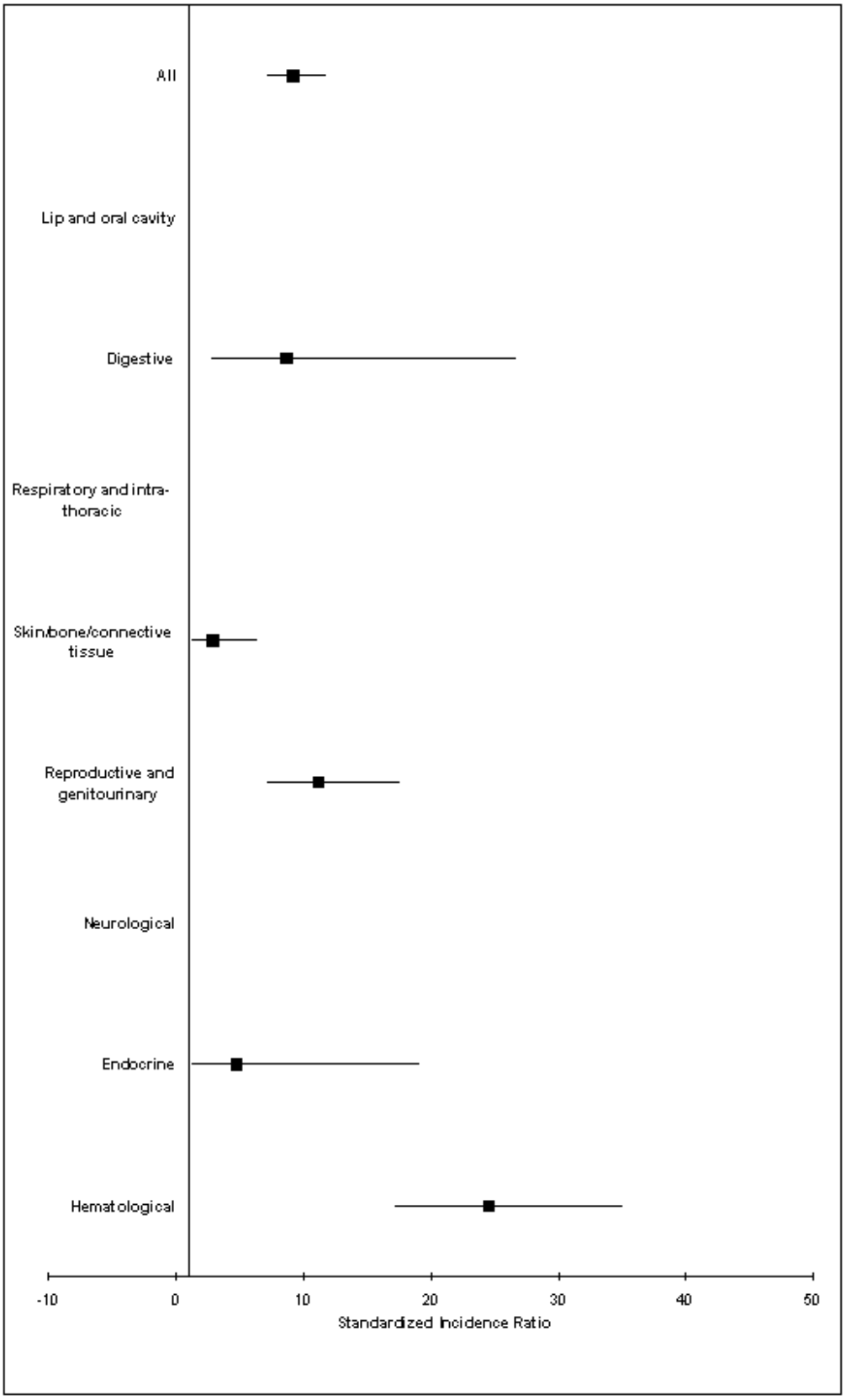
Figure 4.2 displays graphically the results from Table 4.2 which presents SIR by cancer site. Here each person's follow-up period in the ESKD dataset from the onset of ESKD to the time when either he/she gets a cancer in that specific site or was censored, was split by age group and calendar year and then sorted by country (New Zealand ad Australia), gender, age group and calendar year. The population dataset provides the rates of cancer in specific sites (per 100,000 person-years) among patients in strata by country, gender, age group and calendar year; for example the dataset has for males in Australia who were infants in the year 1982, a rate of x cancers in the 'lip and oral cavity' per 100,000 persons-years. The SIRs were derived in an identical manner as was described before.

Table 4.2 shows that the hematological cancers were the dominating cancers accounting for half of all cancers. The incidence of hematological cancers (SIR 24.6; CI, 17.2-35.2) among persons with ESKD was almost 25 times higher than the incidence in the general population. The cancers associated with the reproductive and genitourinary systems were the second most common cancers accounting for 19 of the 60 cancers. The incidence of this type of cancer (SIR, 11.2; CI, 7.2-17.6) was around 11 times higher than the incidence in the general population. In our sample there were no cancers associated with the 'lip and oral cavity', 'respiratory and intra-thoracic', and 'neurological' systems.

Table 4.2 Standardized Incidence Ratios of Cancer by Cancer Site

	ICD Codes	Observed No. of Cancer	Expected No. of Cancers	Standardized Incidence Ratios (with CIs)
All	C00 - C95	60	6.51	9.21 (7.15, 11.86)
Lip and oral cavity	C00 - C14	0	0.23	
Digestive	C15 - C25	3	0.35	8.64 (2.79, 26.79)
Respiratory and Intra-thoracic	C30 - C38	0	0.11	
Skin/bone/connective tissue	C40 - C48	6	2.10	2.86 (1.28, 6.36
Reproductive and genitourinary	C50 - C67	19	1.70	11.21 (7.15, 17.57)
Neurological	C69 - C72	0	0.40	
Endocrine	C73 - C75	2	0.42	4.78 (1.20, 19.11)
Hematological	C81 - C95	30	1.22	24.59 (17.19, 35.16)

Figure 4.2 Standardized Incidence Ratios of Cancer by Cancer Sites



Discussion

The main aims of Project 1 were to determine if having cancer (before or after the onset of ESKD) reduces survival, to specify the main types of cancers that were associated with the patients in our study, to determine the incidence proportion of cancer among ESKD patients, to determine the risk factors for getting a cancer after the onset of ESKD, and to determine if cancer reduces graft survival. In our analysis we saw that cancer substantially increases the risk of dying with patients with cancer either before or after the onset of ESKD being four times more likely (HR 3.83; CI, 1.25-11.79) to die at any given point than patients with no cancer. Among the cancers which began before the onset of ESKD, the cancers associated with the kidneys accounted for 9 out of 16 or 56% of the cancers. Among the cancers which began after the onset of ESKD, the cancers associated with the hematological system accounted for 38 out of the 110 or 35% of the cancers and the cancers associated with the reproductive and genitourinary systems accounted for a further 36 out of 110 or 33% of the cancers. In the first five years of follow-up since the onset of ESKD 1.25% of the patients got cancer, in the first 10 years 3.5% of the patients got cancer, in the first 15 years 6.25% of the patients got cancer and in the first 20 years 8% of the patients got cancer. The main risk factor for getting a cancer was age at the onset of ESKD with the patients who were infants at a substantially greater risk of acquiring cancer than the other age groups and patients who were between 5 and 9 years old at a substantially lower risk than the other age groups. Another risk factor for acquiring cancer was the type of renal replacement therapy used with patients with a donor kidney graft twice as likely to get cancer as patients on dialysis (HR, 1.94; CI, 1.18-3.20). Lastly we see that having a cancer either before the first kidney transplant or during the life of the first kidney graft was not associated with graft survival; we can generalize this result to state that cancer was not associated with graft survival.

The main aims of Project 2 were to determine the relative survival of patients with ESKD and post-onset cancer and to compare it with the relative survival of patients with ESKD, to determine the standardized mortality ratio (SMR) of patients with ESKD and post-onset cancer and compare it with the SMR of patient with ESKD, to determine the standardized incidence ratio (SIR) of post-onset cancer across subgroups of the ESKD population, and to

determine the SIR for types of cancers. Among patients with ESKD (1,495 patients), in the 20 years of follow-up 25% of the patients died of causes directly related to ESKD; whereas among patients with ESKD and cancer (67 patients), in the 20 years of follow-up the incidence proportion of deaths was much worse with 62% of the patients dying from causes directly related to ESKD and/or cancer. For both groups, very few patients died from other causes and so nearly all the deaths were caused by ESKD and/or cancer. The mortality rate among the ESKD-cancer patients was 76 times the mortality rate of the general population (SMR, 76.14; CI, 49.12-118.02) whereas the mortality rate among the ESKD patients was 24 times the mortality rate of the general population (SMR, 23.68; CI, 21.11-26.57). Thus the mortality rate among the ESKD-cancer patients was three times the mortality rate among patients with ESKD. The cancer rate among ESKD patients was nine times the cancer rate in the general population (SIR, 9.21; CI, 7.15-11.86). ESKD patients under the age of five at the onset of ESKD were at very high risk of acquiring cancer (SIR, 26.24; CI, 12.51-55.03). Lastly the patients in our population were at a considerable risk of acquiring cancer associated with the hematological system (SIR, 24.59; CI, 17.19-35.16).

The study had some limitations. The population of ESKD patients in our database in the later eras were a truer reflection of the population of ESKD patients that the population of ESKD patients in the earlier eras as with the improvement in the diagnosis and treatment of ESKD [Herwig-Ulf Meier-Kreiesche et al, 2001] a lot more of the patients with rapidly progressive kidney disease were being captured in our database than would not have been in the earlier eras (they would have died before being entered into the database); thus cancer incidence and survival would be overestimated. An implicit assumption made in calculating the relative survival in ESKD-cancer patients is that on average the negative effect of cancer on survival does not get worse with time but remains constant. This may not be an unreasonable assumption as some peoples' cancer does get worse while others go into remission and so on average the negative effect of cancer on survival may remain constant.

References

Australia and New Zealand Dialysis and Transplant Registry (2009), http://www.anzdata.org.au/, accessed July 4 2009

Australian Bureau of Statistics (2009), http://www.abs.gov.au/, accessed July 4 2009

Australian Institute of Health and Welfare (2009) *Cancer Incidence Data Cubes*, http://www.aihw.gov.au/, accessed July 4 2009

Hosmer DW, Lemeshow S and May S (2008) *Applied Survival Analysis,* 2nd ed. Wiley, New York

Maisonneuve P, Agodoa L, Gellert R, Stewart J, Buccianti G et al (1999) *Cancer in patients on dialysis for end-stage renal disease: an international collaborative study.* Lancet; 354:93-99

McDonald S and Craig J (2004) *Long-Term Survival of children with End-Stage Renal Disease;* N Engl J Med; 350:2657-62

National Institute for Health and Clinical Excellence (2008) *Clinical guidelines 73: Chronic kidney disease*

National Kidney Foundation (2002) *K/DOQI clinical practice guidelines for chronic kidney disease*

New Zealand Cancer Registry (2009) *Cancer: New Registrations and Deaths 2004,* http://www.nzhis.govt.nz, accessed July 4 2009

SEER*Stat (2009) http://seer.cancer.gov/seerstat, accessed July 4 2009

StatsDirect(2009) http://www.statsdirect.com, accessed July 4 2009

Herwig-Ulf Meier-Kreiesche, Akinlolu O. OJO, Freidrich K. Port, Julie A. Arndorfer, Diane M. Cibrik, and Bruce Kaplan (2001) *Survival Improvement among Patients with End-Stage Renal Disease: Trends over Time for Transplant Recipients and Wait-Listed Patients;* J Am Soc Nephrol 12: 1293–1296, 2001

Dickman et al. *Regression models for relative survival.* Statistics in Medicine/2004; 23:51-64.

Acronyms

ANZDATA	Australia and New Zealand Dialysis and Transplant Registry
CI	Confidence Interval
CRS	Cumulative Relative Survival
CKD	Chronic Kidney Disease
ESKD	End Stage Kidney Disease
GFR	Glomeruler Filtration Rate
ICD9	International Classification of Disease 9th Version
ICD10	International Classification of Disease 10th Version
KM	Kaplan-Meier
PH	Proportional Hazard
RRT	Renal Replacement Therapy
SIR	Standardized Incidence Ratio
SMR	Standardized Mortality Ratio

Appendices

Appendix 1.1 Definitions of Survival, KM estimator, and Hazard

Survival function at any point *t* reports the probability of surviving beyond time *t*.

The Kaplan-Meier estimator of Survival is the nonparametric maximum likelihood estimate of *S*(*t*). It is a product of the form

$$\hat{S}(t) = \prod_{t_i<t} \frac{n_i - d_i}{n_i}.$$

In the absence of censoring, n_i is the number of survivors just prior to time t_i whereas with censoring, n_i is the number of survivors minus the number of losses (censored cases). The 'at risk' population is only those surviving cases that were still being observed (have not yet been censored)

Hazard rate is the instantaneous rate of failure. It is the limiting probability that the failure event occurs in a given interval, conditional upon the subject having survived to the beginning of that interval, divided by the width of the interval.

Cumulative Hazard at a point *t* is the sum total of the hazards up to point *t*.

The Nelson-Aalen estimator of Cumulative Hazard at time *t* is defined as the sum over all distinct failure times less than or equal to *t* of the ratio of the failures over number at risk at each distinct failure time.

The Cox proportional hazard model can be written as:

$h(t) = h_0(t)*\exp(b_1*x_1 + ... + b_n*x_n)$ where *h(t)* denotes the resultant hazard, given the values of the *n* covariates for the respective case (x_1, x_2, ..., x_n) and the respective survival time (*t*). The term $h_0(t)$ is called the *baseline hazard*; it is the hazard for the respective individual when all independent variable values are equal to zero. This model is made linear by dividing both sides of the equation by $h_0(t)$ and then taking the natural logarithm of both sides: $\log[h(t)/h_0(t)] = b_1*x_1 + ... + b_n*x_n$

Appendix 1.2 Covariates – Time to First post-onset of ESKD Cancer

Covariates	Description
Pre-ESKD Cancer flag	This binary covariate takes on the value 1 if the subject had a cancer pre-ESKD
Gender	Gender
Race	Race
Age Group at the onset of ESKD	Age group at the onset of ESKD
Treatment Type 1	This binary time-varying covariate takes on the value when the subject has a kidney transplant
Treatment Type 2	This group of binary time-varying covariates was made up of 5 covariates C2: 1 when the subject was on his/her 1st Tx, and 0 otherwise C3: 1 when the subject was on Dx just after 1st Tx, and 0 otherwise C4: 1 when the subject was on his/her 2nd Tx, and 0 otherwise C5: 1 when the subject was on all Dx after 2nd Tx, and 0 otherwise C6: 1 when the subject was on his/her 3rd + Tx, and 0 otherwise
Treatment Type 3	This group of binary time-varying covariates was made up of 3 covariates C2: 1 when the subject was on his/her 1st Tx, and 0 otherwise C3: 1 when the subject was on all Dx just after 1st Tx, and 0 otherwise C4: 1 when the subject was on his/her 2nd + Tx, and 0 otherwise
State of Residence at the first RRT	State of residence at the first RRT
Decade of First Treatment	Era of first treatment
Primary Disease Category	Primary disease category

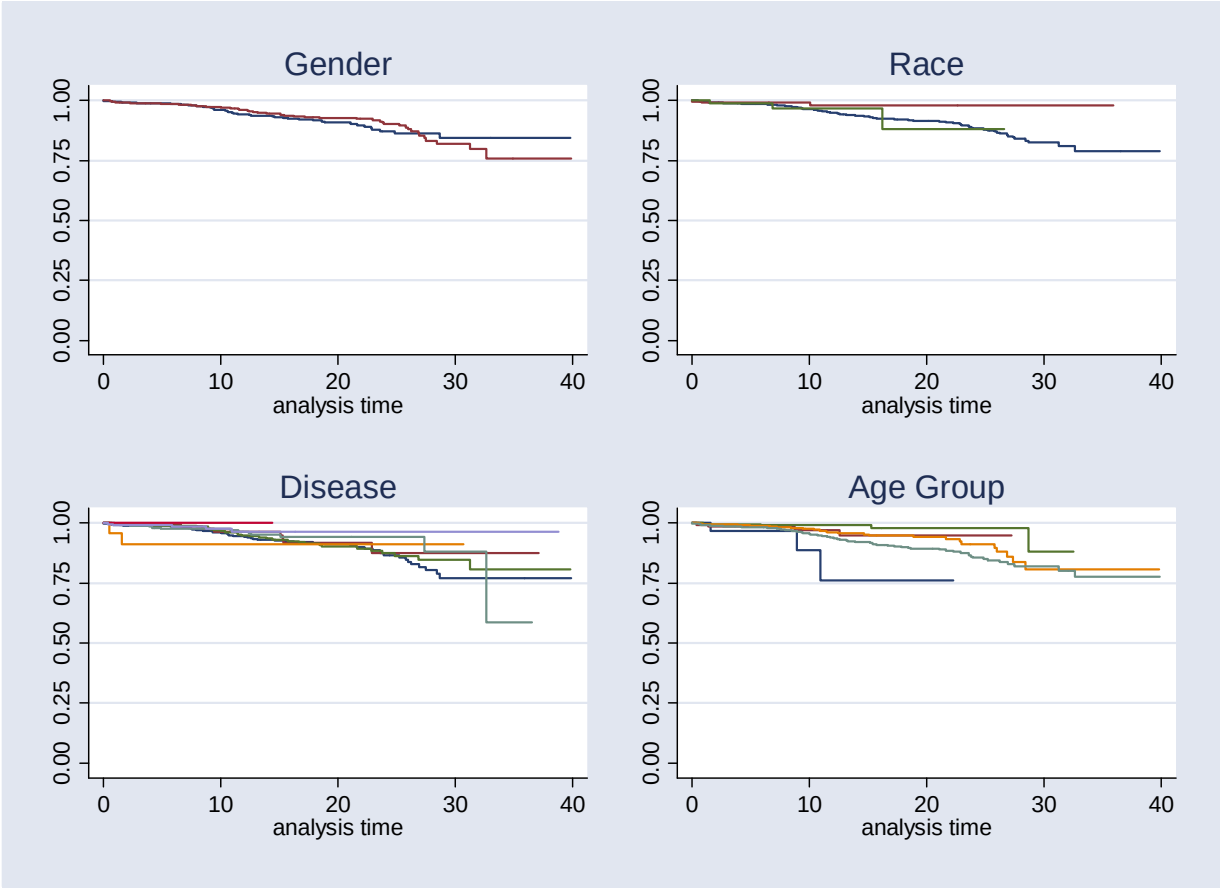
Appendix 1.3 Covariates – Time to First Transplant Failure

Covariates	Description
First post-transplant cancer	This binary time-varying covariate takes on the value 1 when the subject has his/her first post-transplant cancer
Age Group at First Transplant	Age group at first transplant
Duration of First Dialysis	Total duration of the first dialysis was broken up into four quartiles with each level of this categorical covariates corresponding to a quartile.
Gender	Gender
Race	Race
Primary Disease Category	Primary disease category
State of Residence at the first RRT	State of residence at the first RRT
First Donor Type	Status of the donor for the first transplant i.e. deceased or living
Era of First Treatment	Decade of first treatment

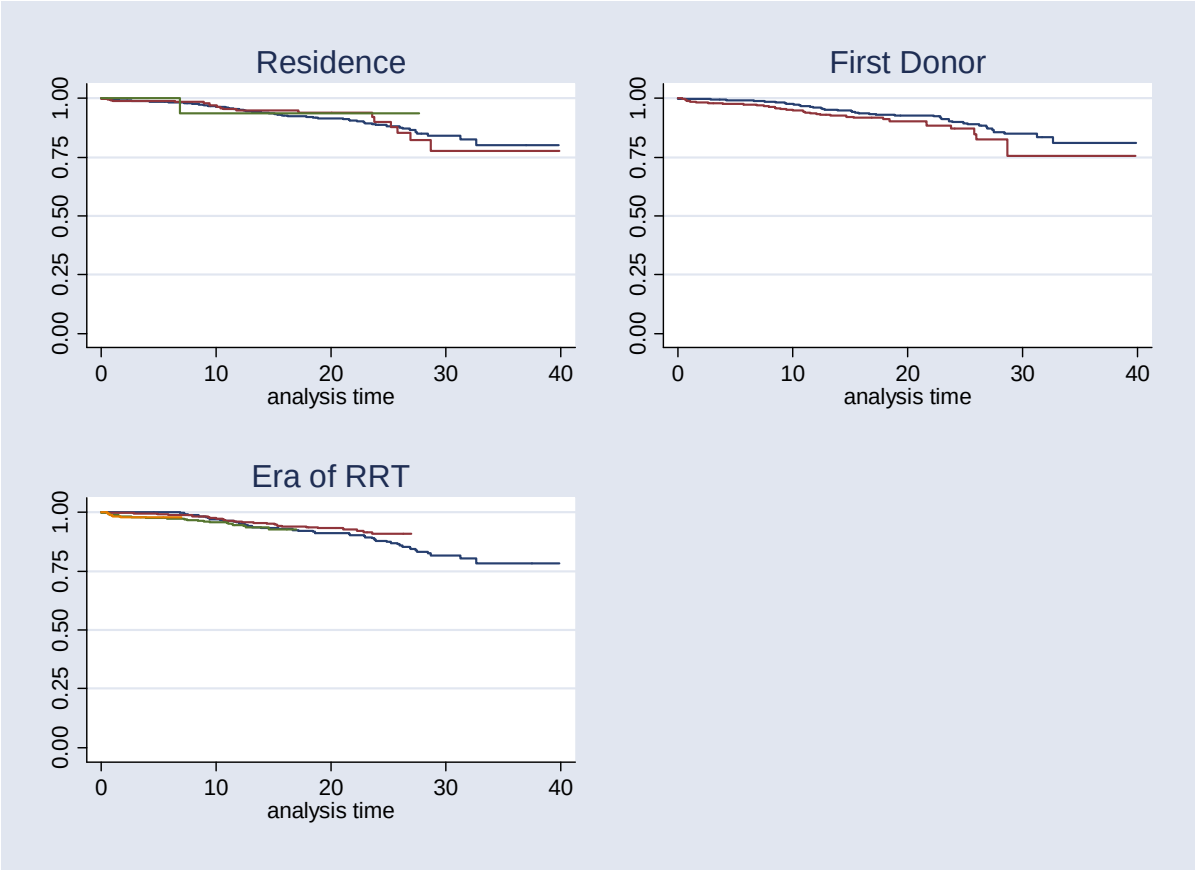
Appendix 1.4 Covariates – Time to Death

Covariates	Description
Pre-ESKD Cancer flag	This binary covariate takes on the value 1 if the subject had a cancer pre-ESKD
First post-ESKD cancer	This binary time-varying covariate takes on the value 1 when the subject has his/her first post-ESKD cancer
Gender	Gender
Race	Race
Age Group at the onset of ESKD	Age group at the onset of ESKD
Treatment Type 1	This binary time-varying covariate takes on the value when the subject has a kidney transplant
Treatment Type 2	This group of binary time-varying covariates was made up of 5 covariates C2: 1 when the subject was on his/her 1st Tx, and 0 otherwise C3: 1 when the subject was on Dx just after 1st Tx, and 0 otherwise C4: 1 when the subject was on his/her 2nd Tx, and 0 otherwise C5: 1 when the subject was on all Dx after 2nd Tx, and 0 otherwise C6: 1 when the subject was on his/her 3rd + Tx, and 0 otherwise
Treatment Type 3	This group of binary time-varying covariates was made up of 3 covariates C2: 1 when the subject was on his/her 1st Tx, and 0 otherwise C3: 1 when the subject was on all Dx just after 1st Tx, and 0 otherwise C4: 1 when the subject was on his/her 2nd + Tx, and 0 otherwise
Era of First Treatment	Decade of first treatment
Primary Disease Category	Primary disease category

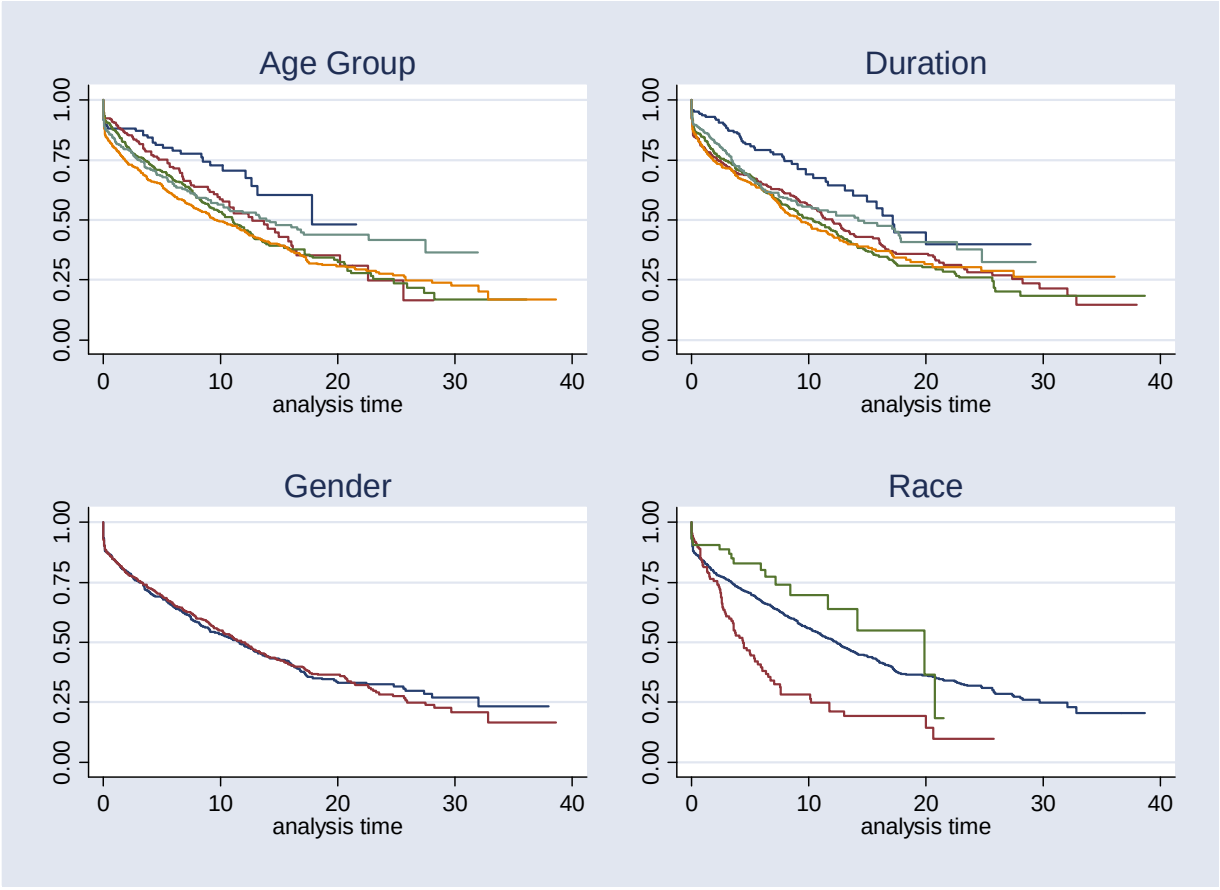
Appendix 1.5 KM Graphs – Time to First post-onset of ESKD Cancer



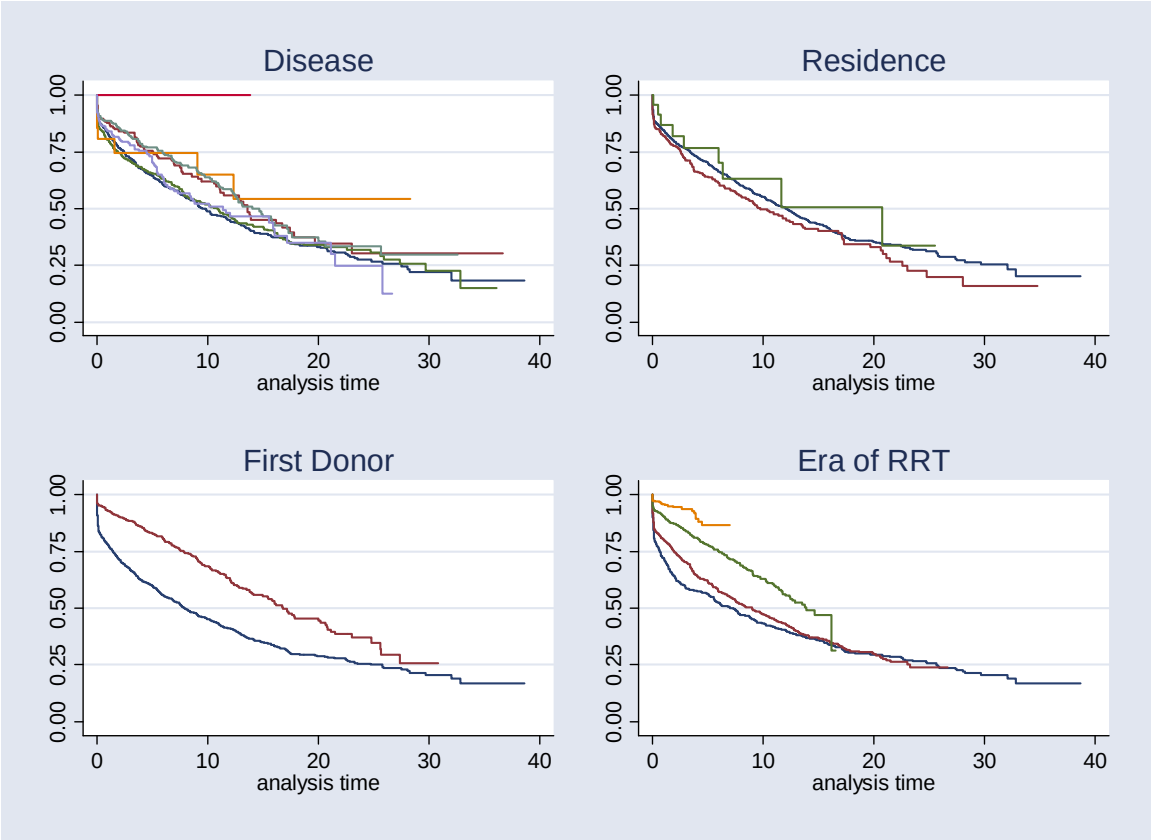
Appendix 1.5 KM Graphs – Time to First post-onset of ESKD Cancer (continued)



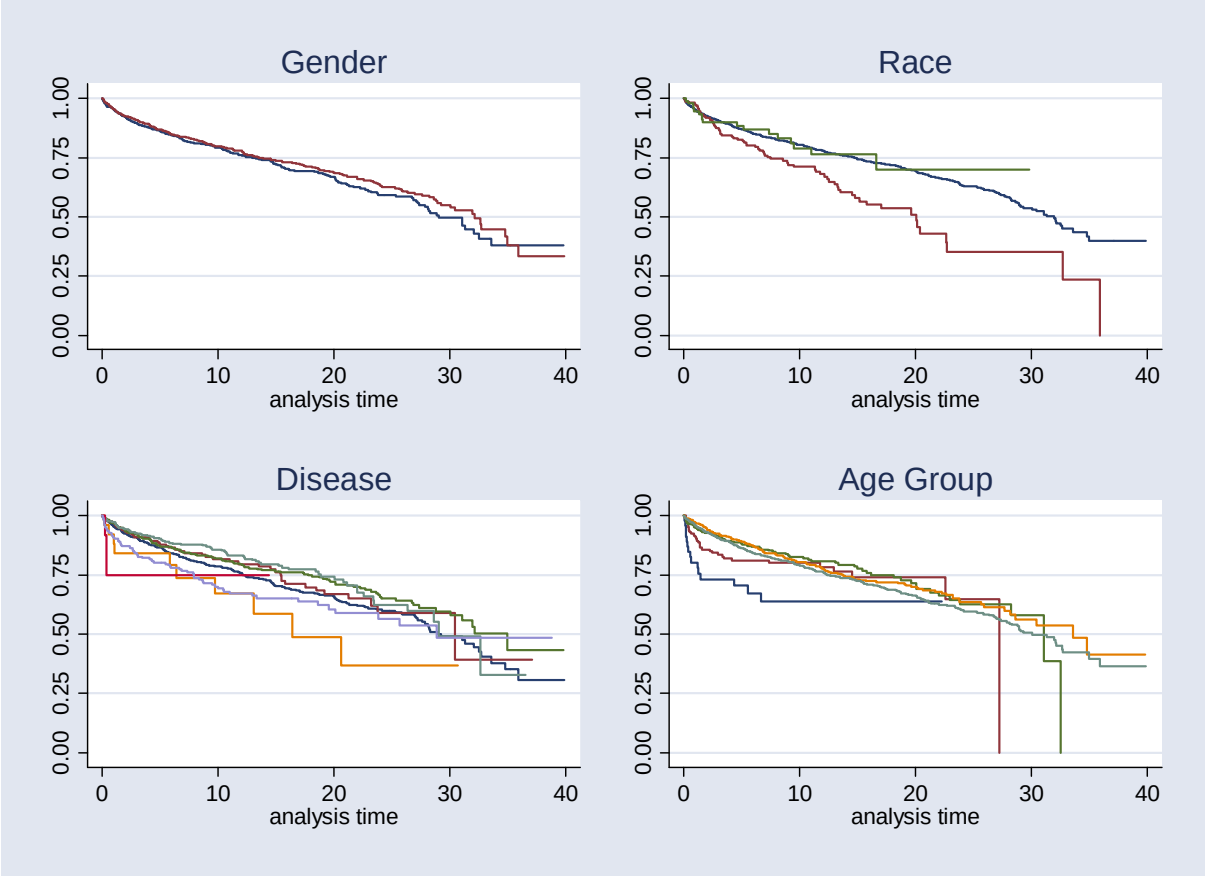
Appendix 1.6 KM Graphs – Time to First Transplant Failure



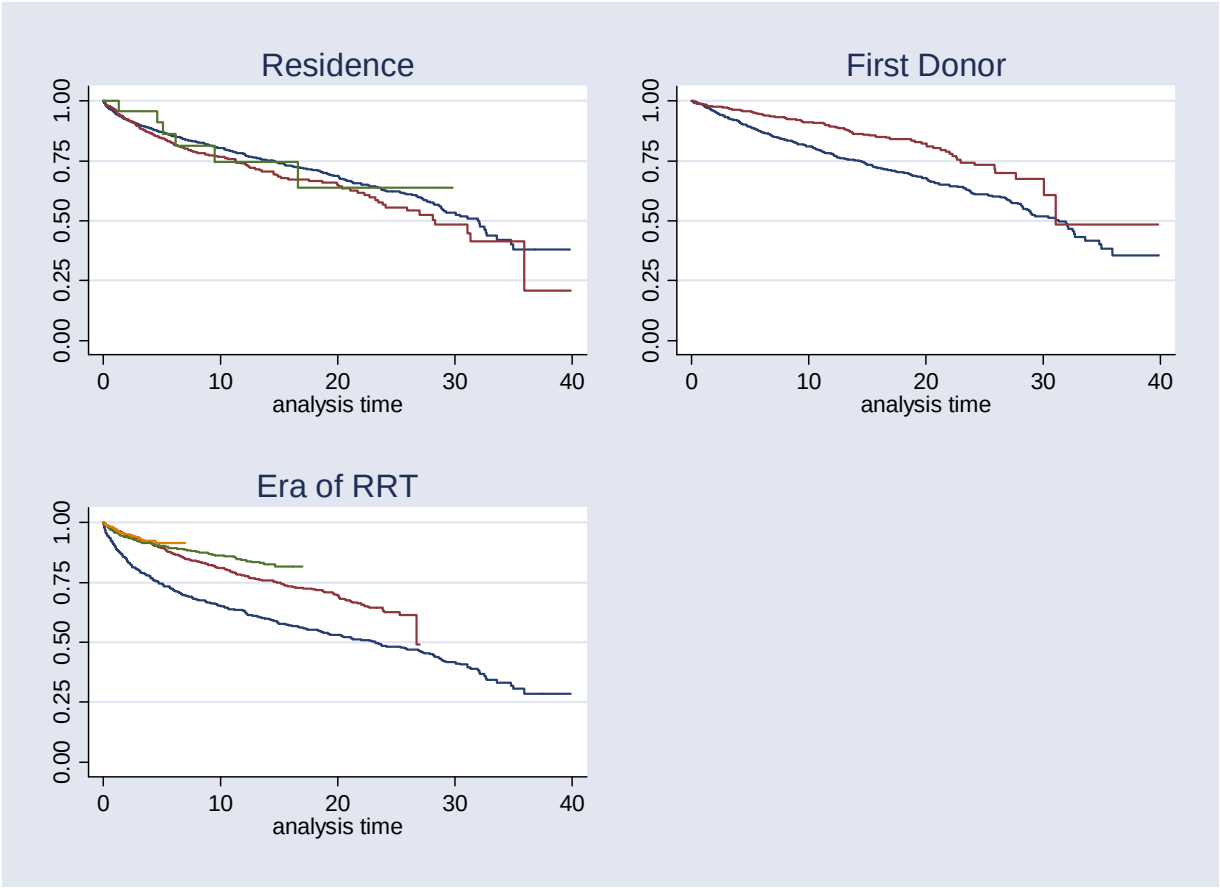
Appendix 1.6 KM Graphs – Time to First Transplant Failure (continued)



Appendix 1.7 KM Graphs – Time to Death



Appendix 1.7 KM Graphs – Time to Death (Continued)



Appendix 1.8 Mortality Rates

Patient characteristic		No post-ESRD cancer			Post-ESRD cancer		
		Patient-years at risk	Number of deaths	Mortality rate (per 1000 patient years)	Patient-years at risk	Number of deaths	Mortality rate (per 1000 patient years)
Gender	Female	10440	233	22.3	345	19	55.1
	Male	13548	268	19.8	298	26	87.2
Age group	< 1	268	15	56	11	0	0
	1 to 4 years	1293	30	23.2	6	2	333.3
	5 to 9 years	3241	63	19.4	11	2	181.8
	10 to 14 years	6514	122	18.7	107	10	93.5
	15 to 19 years	12672	271	21.4	507	31	61.1
Race	Caucasians	21328	414	19.4	579	44	76
	Indigenous and Pacific Islanders	1869	70	37.5	20	1	50
	Asians	791	17	21.5	44	0	0
First treatment group	Peritoneal Dialysis	10781	227	21.1	177	19	107.3
	Hemodialysis	11410	258	22.6	434	24	55.3
	Transplant	1797	16	8.9	32	2	62.5
Era of first treatment	1963 to 1979	7420	227	30.6	290	25	86.2
	1980 to 1989	9338	169	18.1	210	13	61.9
	1990 to 1999	5853	80	13.7	129	6	46.5
	2000 to 2006	1377	25	18.2	13	1	76.9
State of residence at the first RRT	Australia	19366	387	20	485	37	76.3
	New Zealand	4355	108	24.8	135	8	59.3
	Overseas	267	6	22.5	23	0	0
First donor	Deceased donor	15455	305	19.7	389	30	77.1
	Living donor	7749	74	9.5	253	14	55.3
Primary Disease category	Glomerulonephritis	9330	207	22.2	331	23	69.5
	Cystic/Hereditary	2047	41	20	41	2	48.8
	Reflux	5795	101	17.4	211	10	47.4
	Interstitial Nephritis	242	9	37.2	7	1	142.9
	Congenital/Urological	4100	68	16.6	46	6	130.4
	Cancer related	.	3	.	.	0	.
	Other/Uncertain	2405	72	29.9	6	3	500

Appendix 2.1 Calculation of Relative Survival

Relative Survival was defined as the observed (all-cause) survival in the patient group divided by the expected survival of a comparable group from the general population. It was usual to estimate the expected survival proportion from nationwide population life tables, stratified by age, sex and calendar year. Relative survival represents an estimate of the effect of the disease alone on survival. For example, if the observed survival for a group of cancer patients was 50% at 5 years and the expected survival was 80% for the general population, then the relative survival would be 50/80, or 63%. So, among the 50% of cancer patients who die within 5 years, it was expected that 37% (1−0.63) will die from their cancer and 13% from other causes.

Cumulative Relative Survival of each follow-up year (life-table interval) was just the Cumulative Observed Survival in each follow-up year divided by the Cumulative Expected Survival in the same follow-up year. We follow each person in our dataset for a maximum of 10 years.

Calculation of Cumulative Observed Survival

The following were taken to arrive at the cumulative observed survival in each interval:

- We split each person's observation time into multiple intervals, one for each follow-up year ensuring that each observation had the correct value for attained age and attained calendar year. For example, if the onset of ESKD in a 15 year old girl happened in 1980 and she died in the fifth year of follow-up, then she would be contributing five observations where in the fifth observation her attained age at the start of the interval would be 19 years and attained year at the start of the interval would be 1984.
- We note if the person died or withdrew in the last interval. In our example the person died in the fifth interval.
- We aggregate the number who were living in the start of each of the 10 intervals (n_x), the number who died in the interval (d_x) and the number who withdrew in the interval (w_x)
- We then calculate the adjusted number at risk in any interval as the number living at the start of the interval minus half the number who withdrew in the interval

i.e. $n_x' = n_x - .5*w_x$

- We then calculate the interval specific observed survival as $p_x = (n_x' - d_x)/n_x'$
- The cumulative observed survival (l_x) for the first interval would just be the interval specific observed survival for the first interval (p_x). For subsequent intervals, l_x for the interval is the p_x for the interval times the l_x of the previous interval.

Calculation of Cumulative Expected Survival

The following were taken to arrive at the cumulative expected survival in each interval:

- We split each person's observation time into multiple intervals, one for each follow-up year ensuring that each observation had the correct value for attained age and attained calendar year. We also note the person's gender and country (New Zealand or Australia) in which he of she first received renal replacement therapy.
- We have an external file which has expected probabilities for the population for each stratum of country (Australia or New Zealand), gender, age and year (from 1980 to 2007).
- We merge our split file with this external file on country, gender, attained age and attained year to get the expected probability for each interval of persons' observation time
- For each of the ten intervals we take the average of the expected probabilities to give us the interval-specific expected survival (this is using the Ederer II method to estimate expected survival)
- The cumulative expected survival for the first interval would just be the interval specific expected survival for the first interval. For subsequent intervals, the cumulative expected survival for the interval is the interval-specific expected survival for the interval times the cumulative expected survival of the previous interval

Below is an example of the table from a paper by Paul Dickman (Dickman, 2004) on using SAS to calculated Relative Survival.

Int.	N	D	W	Interval-specific observed survival	Cumulative observed survival	Interval-specific expected survival	Cumulative expected survival	Interval-specific relative survival	Cumulative relative survival
0 - 1	75	4	0	0.94667	0.94667	0.99697	0.99697	0.94954	0.94954
1 - 2	71	8	0	0.88732	0.84000	0.99682	0.99381	0.89015	0.84524
2 - 3	63	1	1	0.98400	0.82656	0.99649	0.99032	0.98747	0.83464
3 - 4	61	3	0	0.95082	0.78591	0.99625	0.98660	0.95440	0.79658
4 - 5	58	3	0	0.94828	0.74526	0.99601	0.98266	0.95208	0.75841
5 - 6	55	2	0	0.96364	0.71816	0.99562	0.97836	0.96787	0.73404
6 - 7	53	0	0	1.00000	0.71816	0.99532	0.97378	1.00470	0.73749
7 - 8	53	0	0	1.00000	0.71816	0.99491	0.96882	1.00512	0.74127
8 - 9	53	1	0	0.98113	0.70461	0.99453	0.96352	0.98653	0.73128
9 - 10	52	2	0	0.96154	0.67751	0.99418	0.95792	0.96717	0.70727

We calculate four tables of Relative Survival:

- Relative Survival for the entire population in our study
- Relative Survival for the entire population in our study broken down by age at the onset of ESKD (0-4 and 5-19) and era of onset (1980-89 and 1990-2006)

Appendix 2.2 Formula for Standardized Incidence/Mortality Ratios

Below are definitions taken from the website StatsDirect(2009) http://www.statsdirect.com, accessed July 4 2009.

Indirect standardization was used to calculate the expected mortality rate for the index population, given age specific mortality rates from a reference population.

The confidence intervals are calculated by the exact Poisson method. Below are the equations for Standard Mortality Rates (SMR), the number of expected deaths (e), the lower confidence limit of SMR (LL), and the upper confidence limit of SMR.

$$SMR = 100 \times \frac{d}{e}$$

$$e = \sum_{i=1}^{k} n_i R_i$$

$$LL = \frac{\left(\chi^2_{\frac{\alpha}{2},2d} \right) \times 0.5}{e}$$

$$UL = \frac{\left(\chi^2_{1-\frac{\alpha}{2},2(d+1)} \right) \times 0.5}{e}$$

where $\chi^2_{\nu,\alpha}$ is the (100*α)th chi-square centile with ν degrees of freedom, d is the number of observed deaths, n_i is the person-time for the ith study group stratum and R_i is the reference population rate for the ith stratum.

For SMR the event is death and for SIR of cancer the event is cancer.