Work Placement Project Portfolio

Kevin McGeechan

Preface

Introduction

The two projects that form this portfolio were undertaken between August 2005 and June 2006. The first project involved analysing data from a laboratory experiment that was carried out by Dr Andrew Richards, a Masters of Surgery student at the University of Sydney. Dr Richards conducted a randomised controlled trial to investigate the effect of non-steroidal anti-inflammatories (NSAIDs) on prostate cancer tumour growth. The trial was carried out on 73 mice and the size of the tumours was measured on a regular basis. Dr Richards wished to determine whether the use of NSAIDs would slow the growth of these tumours.

The second project was conducted in collaboration with researchers at the School of Public Health and the George Institute. Dr Mike Jones and Dr Paul Roderick were initially interested in examining the relationship between socioeconomic status and chronic kidney disease using the data collected by the National Health and Nutrition Survey (NHANES) in the USA. These data are publicly available. Researchers at the George Institute became interested in the analysis proposed by Dr Jones and Dr Roderick and offered to provide access to datasets from surveys carried out in Thailand and Australia. The collaboration evolved into a study to examine the association between socio-economic status and chronic kidney disease and whether this association is different in different ethnic groups.

Student's role

My role in each was to prepare the data for analysis, provide advice on how the data could be analysed, implement the agreed analysis plan and report and provide interpretation of the final results. Although my role was the same in both projects my experience was quite different in the two projects. In the analysis of the prostate cancer data I was the sole person advising Dr Richards. This involved a great deal of responsibility, but also freedom on how the analysis could be approached. However, being the sole advisor meant that I had no-one directly involved in the project with whom I could discuss the more technical statistical aspects of the project. In such instances Dr Timothy Dobbins provided invaluable support and advice.

The analysis of the kidney data, in contrast, involved collaborating with six other researchers. These researchers came from diverse clinical, statistical and epidemiological backgrounds and brought to the project a wealth of knowledge on the subject and many ideas as to how such data could be analysed. The final analysis plan was developed in consultation with Dr Mike Jones with input from the other researchers. This entailed less responsibility on my part in determining the "correct" analysis, however there was a greater pressure on myself to provide clear and concise summaries of the analysis that I performed.

Reflections on learning

The WPP reinforced for me how little time in a statistical analysis is actually spent producing statistical results. The majority of time is taken up by managing and cleaning data, determining the correct methods by which to use the data for the current problem and finally interpreting the results in the final report. The production of results achieved through the implementation of statistical software is in comparison a minor task. However, to ensure that this project produces results which are meaningful, effort must be spent on the comparatively mundane task of data management, the more abstract conceptualisation of the problem and the more literal communication of results. The coursework for the Masters program tended to focus more on the narrow implementation of a particular data analysis technique, whereas the WPP has helped me to broaden my skills to incorporate the whole process of statistical consultation and analysis.

One of the most important roles of the statistician is to communicate the final results. The statistician must do this in a clear, concise way and avoid over-interpreting the results and my experience of the WPP certainly improved my communication skills. In particular, for the kidney study I regularly presented results at meetings and responded to email queries from all the other researchers involved. Dr Mike Jones' advice in this area was very useful especially with regards to keeping the results concise.

Both of the projects involved going beyond the techniques I had learned in the coursework component of the Masters course.. The prostate tumour analysis involved investigating how interval censored survival data could be analysed and how sample sizes for longitudinal studies could be calculated. The kidney project involved learning how complex surveys can be analysed using logistic regression while taking into account their sampling design.

Although both projects had relatively clear and simple research questions, how these questions were addressed demonstrated to me that statistical analysis is not neatly compartmentalised into linear models, survival analysis, longitudinal data etc as in the coursework. The prostate tumour analysis involved using methods learned in Survival Analysis and Longitudinal and Correlated Data, and the kidney analysis involved material from Clinical Biostatistics, Health Indicators and Surveys and Categorical Data Analysis.

Ethical considerations

The tumour analysis project was carried out on live animals, which ultimately resulted in their death. The conduct of this experiment was to be passed by the University of Sydney's Animal Ethics Committee to ensure the mice were cared for appropriately.

The data from the kidney study all came from population surveys that had been carried out by other agencies. Each of these studies received ethical approval regarding the conduct of the surveys.

Project A

Project title

Analysis of tumour growth data using survival curves and mixed models

Location and dates

School of Public Health, University of Sydney
August 2005 – November 2006

Context

Dr Andrew Richards was conducting an experiment to examine the effects of non-steroidal anti-inflammatories on the growth of advanced prostate tumours amongst laboratory mice. Dr Richards sought my advice on how he could analyse the data collected. He had considered using survival analysis methods but at the time did not have the skills required to complete this analysis. I also advised Dr Richards that as he had collected the size of the tumours at different time points, alternative methods to survival analysis could be used.

Student contribution

I met with Dr Richards on several occasion to discuss the data he had collected, how I thought it could be analysed and to provide interpretation of the final results. I carried out all analysis myself after extensive consideration of the appropriate methods to use. Dr Timothy Dobbins provided assistance in clarifying certain statistical issues.

Statistical issues

The data collected could be analysed in a number of different ways. The statistical issues involved deciding which of the methods would be appropriate. The possible methods included the survival analysis of right, as well as interval, censored data, general estimating equation methods (GEE) and linear mixed models.

Also, I further extended my knowledge of longitudinal data analysis by investigating how the sample size for such studies could be calculated. This had not been covered in the coursework.

Acknowledgements

I would like to thank Dr Andrew Richards for allowing me to analyse his data, and Dr Timothy Dobbins and Dr Judy Simpson for providing advice as the analysis progressed.

Student declaration

I declare that this project is my own work, with guidance provided by my project supervisor, Dr Timothy Dobbins, and that I have not previously submitted it for academic credit.

_____

Kevin McGeechan

_____

Date

Supervisor declaration

Kevin performed the majority of this project independently. He was instrumental in meeting with Dr Richards, proposing and performing the appropriate analyses and communicating the results to Dr Richards. I provided minimal supervision in assisting Kevin communicate and interpret the models used in the project.

_____

Dr Timothy Dobbins

_____

Date

Analysis of tumour growth data using survival curves and mixed models

Background

Prostate cancer is one of the most frequently diagnosed cancers amongst men in Australia (AIHW, 2006) and the current treatment for men with advanced prostate cancer is androgen withdrawal. Androgen withdrawal is carried out by either chemically, or surgically, castrating the patient. This delays the growth of the tumour, however after a certain time the tumour inevitably begins to grow again. This stage is known as androgen independence. There have been studies conducted recently on mice to determine whether different therapeutic strategies could extend this time to tumour re-growth (or time to androgen independence) (Eigl, 2005. Nicholson, 2004). One such study was carried out by Dr Andrew Richards, a Master of Surgery student at the University of Sydney.

The research project's objective was to investigate whether certain drugs could inhibit the growth of advanced prostate cancer tumours which had been grafted onto a group of 73 mice. Dr Richards' description of the project was as follows -

> Thymus deficient male nude mice of similar ages were acquired through The University of Sydney Laboratory Animal Services (LAS). All 73 mice were xenografted with the LNCaP human prostate cancer cell line on roughly the same day (+/-2). Animals were marked and randomised to 12 boxes. *(There were six mice in 11 boxes, and seven mice in one box.)*

Ethics approval was obtained from University of Sydney's Animal Ethics Committee (AEC approval no: K26/3-2005/3/4105).

Once tumour reached an index volume ($263mm^3$) as determined by twice weekly calliper measurement, animals were castrated to effect androgen withdrawal (the standard treatment for advanced prostate cancer) and implanted with a subcutaneous pump delivering 1 month of constantly infused vehicle control (Group 1), pure celecoxib (Group 2), celecoxib from tablets

(Group 3), ibuprofen (Group 4) or indomethacin (Group 5). *(All mice within each box were assigned to the same group.)*

We know that, after initial inhibition or shrinkage of tumours with castration, they will all begin to grow again and become androgen independent. Once the tumour grows to greater than 105 mm^3 above its castration size, we call it androgen independent.

Our hypothesis is that these anti-inflammatory drugs will prolong the time to androgen independence.

In addition to the above description Dr Richards also stated that for some mice the time to androgen independence may not be available as they either did not reach androgen independence before the experiment finished or because the mice had to be euthanased for reasons unrelated to the treatment before they reached androgen independence. Table 1 provides the number of mice in each treatment group and the number who reached the index volume of 263mm^3.

Table 1: Number of mice in each treatment group

Treatment group	Number of mice xenografted with tumour	Number of mice whose tumour reached 263mm^3
Control	25	19
Celecoxib	12	9
Celecoxib from tablets	12	9
Ibuprofen	12	9
Indomethacin	12	10
Total	73	56

Dr Richards intended method of analysing these data was to implement survival analysis techniques using the time to androgen independence as the outcome. He was enrolled in the survival analysis course at the School of Public Health due to commence in two months time, although he had already begun his experiment and had been monitoring the mice for over a month. By the time the course began his experiment would be complete and his data would be ready to analyse.

Initial advice

My advice to Dr Richards was that survival analysis would be an appropriate method to analyse such data because the outcome he wished to compare between groups was the time until an event happened (from implant until androgen independence) and because some of these times may be censored (the time till androgen independence was not available due to mice being euthanased or not reaching androgen independence before the experiment was complete). However, there were a number of issues that he should consider:

1) the total sample size was 73 mice and these were allocated to five different treatments (Table 1). Before doing any power calculations I suggested that although 73 may seem a reasonable sample size when this is split amongst five groups the power of a test to detect a difference between the treatments may be low. I asked whether there could be a way of grouping the treatments, or was the difference between each treatment important?

   Dr Richards stated that although four different types of NSAIDs had been given, whether an NSAID of any kind would have an effect on tumour growth was also of interest.

2) The mice were randomly allocated to each of the treatments by the box to which they were allocated and each box contained six mice (with the exception of one box which contained seven mice). Were the mice in different boxes managed and monitored in the same way and were there any other potential factors that he had recorded (or not recorded) that may influence the growth of the tumours?

   The mice had been bred to be as similar as possible and since birth had received the same feeding and care. There were no other factors that Dr Richards could think of that would influence the tumour growth.

3) Were there any other ways of defining the outcome of treatment other than time to androgen independence that he was aware of and would have some clinical meaning?

For example, had he considered calculating the area under the curve of the tumour growth for each mouse, or was the rate of increase in tumour size over time a measure that would have some clinical meaning?

The rate of increase in tumour size was a clinically meaningful measure that Dr Richards had come across, however he was not aware of the usefulness of the area under the curve.

Having considered these issues, my advice to Dr Richards was that the data could be analysed using survival analysis with time to androgen independence as the outcome. This would involve comparing survival curves of mice given the placebo and mice given an NSAID using the log-rank test. The mice on the different NSAIDs would be grouped together should the power of the study to identify a meaningful difference be too low when all five groups were used. I asked Dr Richards to provide me with an estimate of the difference in survival rates he would consider to be clinically meaningful. The estimate he provided was an increase in time to androgen independence of 50% and he estimated that the average time to androgen independence would be around 30 to 60 days.

As there were no other potential factors to be accounted for, there was no need to conduct more complicated survival analyses using, for example, Cox regression. The analysis to be undertaken would be relatively straight forward and Dr Richards would be able to carry this out after his first few weeks of his survival analysis course, although if he wished the results sooner then I would be available to carry out the work.

In addition to comparing survival analysis I also suggested that other analytical methods could be employed that may be more powerful than the log-rank test. These methods rather than using one single data point for each mouse – time to androgen independence – would utilise all the tumour size measurements to describe the differences in tumour growth between the groups. The methods I suggested were: a summary measures approach where a single measure is calculated for each mouse (eg area under the curve) and then these measures are compared between groups using simple statistical tests such as the two sample t-test (Matthews, 1990). Other possible analytical methods would be the use of Generalised Estimating Equations (GEE) (Diggle, 1994) or mixed models which describe each mouse's tumour growth over time (Verbecke, 2000).

Original data

I received an Excel dataset from Dr Richards which contained for each mouse the tumour size at each occasion it was measured. The table below is an example of the data for the first seven measurements for mice that were placed in box 2, one of the boxes containing mice that were in the control group. Each box contained six mice and mice within each of the boxes had their tumour volume measured on average every 3.5 days. A row of observations is the measurements made for a single mouse and each mouse within a box was identified by ink marks on their body (shown in the first column).

Table 2: Example of Excel data received

VOLUMES (l^2 *w)							
Day	21	24	27	31	34	38	41
DATE	21.6.05	24.6.05	27.6.05	1.7.05	4.7.05	8.7.05	11.7.05
BOX2 (Control)							
left	119.4	148.1	140.6	174.0	241.3	**275.0**	260.7
right	103.6	131.4	107.1	125.2	146.8	117.4	153.7
left and right	196.1	199.2	**267.8**	199.6	208.9	348.2	320.6
left * 2	138.6	183.5	**291.8**	335.4	301.8	545.7	
right *2		45.9	80.3	75.1	150.1	191.4	169.5
no mark	132.9	**289.7**	219.7	212.3	343.5	311.6	272.8

For each mouse the first number to appear in bold indicates the occasion when the tumour reached an index volume of 263mm^3 and so when the mouse was surgically castrated and fitted with the surgical device that delivered the drug.

Dr Richards also provided an Excel datasheet which would be more amenable to survival analysis methods. An example of these data are showing Table 3. This table uses the original coding scheme that Dr Richards used when he entered the data.

Table 3: Excel spreadsheet used to derive data for survival analysis

	uniqueid	group	tumour formed*	event+	Time at event#
BOX2 (Control)					
left	7	1	0	0	17
right	8	1	0	2	28
left and right	9	1	0	0	16
left * 2	10	1	0	0	10
right *2	11	1	0	0	30
no mark	12	1	0	0	27

* 0 = Yes, 1 = No

+ 0 = Androgen independence, 1 = euthanased at 2 months, 2 = other death

Number of days from implanting of drug delivery device until the event

These data gave each mouse a unique identifier and identified to which treatment group they belonged. The variable tumour formed indicated whether the tumour reached the index size of 263mm^3 and so whether the mice were implanted with the drug delivery device. The variable event recorded whether after implantation the tumour increased in size by 105 mm^3 (indicating the tumour had reached androgen independence) or whether the mouse had to be euthanased before this time. The time at event was the time from the implanting of the drug delivery device until the tumour reached androgen independence, or the time of euthanisation.

Evaluation of the appropriate analysis methods to be used

Right and interval censored data

One of the simplest ways to compare two survival curves is using the log-rank test. In brief, this test compares the observed number of events in each treatment group with the number of events that would be expected if the survival curves were the same for each group. This comparison takes place at every time when an event occurs and the differences between the observed and expected numbers of events are then summed over the total follow up time. Large differences between the observed and expected numbers suggest that the survival curves are not the same. Whether the differences observed may have occurred by chance can be determined by comparing the result of the calculations with a Chi-squared distribution. However, use of the log-rank test assumes that the actual event times are known. On seeing the first Excel datasheet I realised that this may not be the case with these data.

The problem I saw was that although the endpoint was defined as being the time from surgery until the tumour grew by $105mm^3$ it was not possible to exactly determine when this happened as the tumour volume was only measured, on average, every three days. These types of data are known as interval censored data as the time when the event occurred is known only to have occurred within a specific interval (Hosmer, 1999).

The most common type of data encountered in survival analysis is right-censored data and the simple survival analysis methods, such as the log-rank test, assume that the data are right-censored. If data are right censored then we know the time at which the event happened, or the last time at which the subject was known not to have experienced the event. Interval censored data require different methods of analysis to right-censored data. However, having examined recent papers that analysed data from similar trials (Nicoholson, 2004. Eigl, 2005) it seemed that researchers ignored the interval censored nature of these data and proceeded as if the data were right censored. They provided no justification for this.

I decided to analyse the data ignoring the interval censored nature with the following justification. In the worst case scenario we would underestimate the median time until androgen independence of one group by the maximum number of days between measurements and overestimate the median for the other group by the same number of days. The maximum time between measurements was 4 days. This means that there could be a potential bias in the estimate of the difference between the median times to androgen independence of 8 days.

Dr Richards indicated that a meaningful difference between treatments would be an increase in time to androgen independence of 50% - this would equate to 15 days if the median in the control group was 30 days. Therefore when interpreting the final result comparing the median time of the two groups it should remembered that the difference between median times may be biased by up to 8 days and a difference in median times of between 7 and 15 days may be suggestive of a meaningful difference.

Power calculations for five and two group comparisons

Having advised Dr Richards that he would have a more powerful test if he compared two groups rather than five I then carried out power calculations using the power and sample size program PS (Dupont, 1997)) for the two group comparison. For the five group comparison the formula (Lee, 1992)–

$$n_d = \frac{2\tau(K-1,\alpha,\beta)}{(\log_e a_k)^2}$$

provides the sample size required in each group where K is the number of groups $2\tau(K-1,\alpha,\beta)$ is a non-centrality parameter. The non-centrality parameter depends upon the number of groups being compared (K), the power ($1-\beta$) and the required significance level (α). In this instance it equals 11.935 where 5 groups are to be compared with a power of 80% and a significance level of 0.05. Finally, a_k is the largest ratio of the mean survival times between groups. In this experiment it is 1.5. The required number of mice in each group is calculated to be 146.

In the experiment there were on average 14.6 mice per group, ten percent of what is required. Therefore, the power with 14.6 mice per group (on average) to detect the difference of 50% would be approximately 6% (Armitage, 2002). In fact there were 12 mice in the four drug groups and 25 in the control group. The unequal numbers in the groups would actually provide less power than 6%.

If all the treatment groups are combined then the power can be calculated for a test comparing two groups - the treatment group with the control group. Table 4 shows that although comparing two groups provides a more powerful test the chance of detecting a difference should it exist is still relatively poor. We would detect a difference in median time of 50% in only 38% of experiments. Commonly a power of 80% is used when designing experiments. The 38% power is still much better than the 6% power of the 5 group comparison. Therefore I decided that the best approach would be to combine the four drug groups in one and proceed comparing the mice who received an NSAID with those in the control group.

Table 4: Power of the log-rank test to detect a difference in median time till androgen independence of 50%

Median time till androgen independence for control group	Number of groups to be compared	Power to detect a difference of 50% in median time
30 days	2	38%
	5	6%

Summary measures analysis

The main outcome of the study is time to androgen independence and this is defined by an increase in tumour volume of 105mm^3. The tumour volume was measured every 3 days and the measurements for an individual mouse can be used to produce a summary measure (eg rate of tumour growth) for that mouse. The summary measures are then compared between the groups. Two approaches to a summary measures analysis that would be appropriate for this experiment would be to use the area under the curve or the rate of increase in tumour volume (Matthews, 1990).

In the area under the curve approach for each mouse the total area under their growth curve is calculated and then the mean areas for the two groups (those on an NSAID and those in the control group) is compared using a two sample t-test. For the rate of increase method ordinary least squares is used to fit a line to the data for each mouse. The mean regression coefficient for the two groups is then again compared using a two sample t-test. It would also be possible to extend this method and fit a quadratic curve to the growth curve for each mouse. This would allow tumour size to fall before beginning to increase which is biologically plausible given Dr Richards' description of tumour growth – "We know that, after initial inhibition or shrinkage of tumours with castration, they will all begin to grow again". Fitting a quadratic would mean having two parameters to compare for each group of mice – the rate of increase at a certain point and the curvature of the fitted line. As the rate of change will alter as time increases then the point at which the curves are compared is important. Comparing the curves at the time of castration may provide a different result to another time, so a clinically meaningful point should be chosen.

These summary measures analyses are problematic in this dataset. Some mice were euthanased before the end of the study, either because they had reached androgen independence or for other

medical reasons, and so the number of tumour volume measurements available for each mouse was different. This would mean that mice who were not euthanased would have more tumour measurements recorded and so their area under the curve may be greater. In addition, the regression coefficients estimated for those mice euthanased and those not would have been based on different numbers of measurements. Those euthanased would have fewer measurements and so the estimated rate of increase would have been less accurate in this group. The summary measures approach was decided not to be appropriate for these data for these reasons.

GEE or mixed models

The two other methods of analysing longitudinal data – generalised estimating equations (GEE) approach and linear mixed models – could possibly be used to analyse these data. If all of the data recorded for each mouse are to be utilised then these methods are necessary as the tumour volumes for each mouse are correlated. Simple linear regression assumes that each observation to be analysed is independent of the others. The two methods – GEE and mixed models – take into account the correlation amongst observations and so provide correct point estimates and standard errors for the estimated differences between groups. However, which of the two methods is appropriate for these data is dependent upon the model that we fit to the data and the assumptions we make about that model.

In the GEE approach a marginal model is specified. This models the expected value of the outcome (tumour growth) as a linear function of the covariates (drug group and time since surgery) with the correlation between observations specified separately to have a particular form, and those observations with the same values of covariates are assumed to have the same correlation structure. Generalised estimating equations are then used to provide estimates and standard errors of the parameters of interest. The results obtained provide the population average effect of the covariates on the outcome. However, if the correlation structure is wrongly specified then robust methods will still provide valid estimates of the parameters and standard errors, so long as certain conditions regarding missing data hold (see the discussion of missing data below).

The linear mixed model also assumes a linear relationship between the outcome and the covariates. However, in this case the parameters of the linear relationship are allowed to vary between individuals. These are the random effects of the model. The term mixed comes from the

fact that these random effects are mixed with fixed effects. In the mixed model the fixed effects provides an estimate of the effect a covariate would have on an individual's outcome, and in the linear mixed model this also gives the population average effect of the covariates. This is not the case in a non-linear mixed model.

In the current analysis, each individual mouse has their own individual parameters that describe the growth in tumour size over time (random effects) but there is also a population average growth curve (fixed effect of time). The correlation that is observed between observations over time for an individual is due to their underlying growth curve which cannot be observed (Liang, 1993), but rather than specifying a particular form for this correlation structure as in the GEE method, the parameters that are included in the model as random effects will determine how this observed correlation is estimated. One result of the more detailed specification of the linear mixed model is that the effect of the covariates for a particular individual (ie subject specific effects) can be estimated rather than simply the population average effect.

Another point of difference between the GEE and mixed models approach is how the two methods deal missing data. Missing data can be classified into missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). In terms of the two approaches the difference lies in whether the data are missing completely at random or missing at random.

Missing completely at random means that the probability of a particular observation being missing (such as the tumour volume measured on a particular day for an individual mouse) does not depend upon any other observation. Missing at random means that the probability of and observation being missing may depend on other observed values, but the fact that it is missing does not depend on its own value.

The data from the mice experiment are missing at random, not missing completely at random. A mouse that has been euthanased cannot have a tumour volume measured, the measurement is missing. The reason the measurement is missing is because of the previous values of the tumour volume, however it does not depend upon the current tumour volume (if it could have been measured).

In order to use the GEE method we must assume that any missing data are missing completely at random. It is possible to relax this assumption so that the data are only assumed to be missing at random. However, using a GEE method in this case we would have to correctly specify the correlation structure. An incorrectly specified model fitted using GEE where the data are missing at random produces biased results (Cnaan, 1997). The mixed model approach, in contrast to the GEE approach, provides unbiased estimates in situations where the data are missing at random.

The mixed models approach has a further advantage over the GEE approach. In the current experiment mice were kept in boxes with six mice in each box. It could be possible that mice in different boxes received slightly different treatment and so mice from the same box would be more likely to have similar experiences than mice in different boxes. There is therefore the potential for some for of clustering of data within boxes. The data are already clustered over time as repeated measures have been made for each mouse. The added clustering level of box results in two different levels of clustering. The mixed models approach can incorporate many levels of clustering, whereas in the simple GEE method we are restricted to one – either over time, or within box. Although this question will not be addressed in the current project it would be another reason for choosing the mixed models approach over the GEE method.

The mixed models approach was therefore chosen to analyse these data as the tumour volumes in the study were missing at random, rather than missing completely at random, and also because it would be of interest to develop a model that could estimate the tumour growth for an individual. Although less important in this animal experiment, it would become more important if the study was extended to human participants.

Within this mixed models approach choices have also to be made regarding the form of the model. One of the choices is how to model the variance in the outcome. The variance in the outcome is determined by how we model the random effects and what correlation we assume there is between observations for an individual. When the data are highly unbalanced, such as in this study where at later observation times there were relatively few mice still alive, the recommended approach is to assume that observations are independent within an individual. The structure of the variance in the outcome is then determined wholly by how the random effects are specified (Cnaan, 1997).

Analysis proposed

Compare the mice in the control group with those who were in the drug group using

- Kaplan Meier curves and descriptive statistics of time from implantation with drug delivery device until androgen independence
- Log-rank test to determine whether the survival curves are different
- Mixed model analysis of time from implantation with drug delivery device until androgen independence

Results

Mice who reached initial tumour volume

Only 56 of the mice implanted with the tumour reached the initial tumour size of 263mm^3. This reduced the power of being able to detect the proposed difference to 30%. However, one positive observation was that the median tumour volume was identical for the mice in the two groups which reached the initial tumour size (table 5). As the mice up until this point were virtually identical any difference would have been due to chance variation, or a bias in the measurement of tumour volume as described earlier. This provides some evidence that such a bias has not occurred.

Table 5: Description of mice whose tumours reached index volume (> 263mm^3)

	Number of mice	Mean tumour volume (mm^3) at time of surgical implant	Median tumour volume (mm^3) at time of surgical implant
Control	19	303	291
Drug	37	297	291

Log rank test

Table 6 provides the median time to AIC for both the drug and the control groups and the results of the log-rank test comparing the survival curve between groups. The drug group had a higher median time too androgen independence (35 days) compared to the control group (30 days) although this difference was not significant ($\chi^2 = 0.39$, df = 1, p = 0.53). For the great majority of the follow-up period the survival curve for the drug group remains above that of the control group (figure 1).

Figure 1

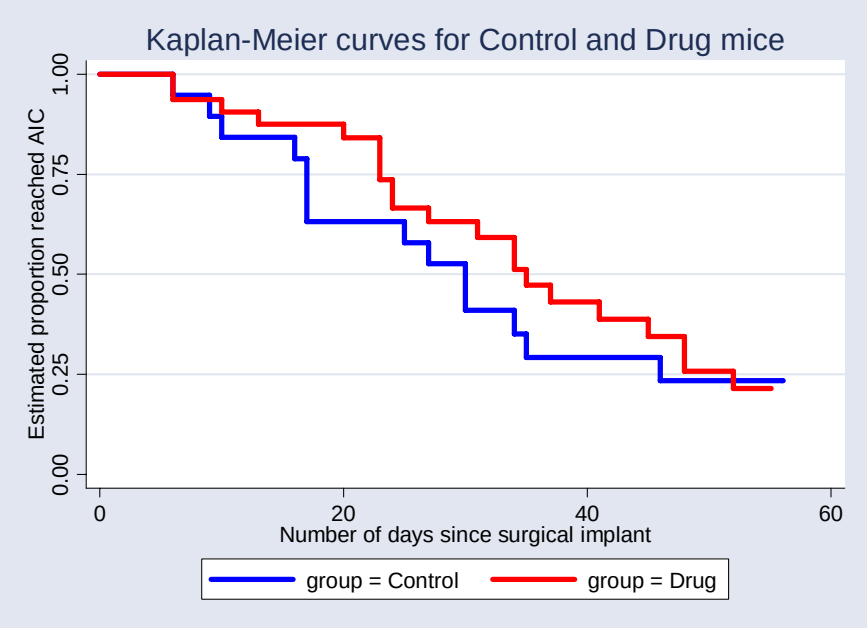


Table 6: Log rank test comparing time until androgen independence

	Number of mice	Number who reached AIC	Number censored	Median time till AIC (days)	95% Confidence Interval	Log-Rank test	Df	p-value
Control	19	14	5	30	(17, 46)	0.39	1	0.53
Drug	37	21	16	35	(24, 48)			

Mixed models

The mice were examined on average every 3.5 days and post-surgery the median number of tumour volumes recorded for the mice was 9. Figure 2 shows the tumour volume trajectories for each mouse in the drug and control groups. Individual trajectory plots were also examined for each mouse in order to determine whether a transformation of the data would be appropriate. This was not felt to be necessary.

Figure 2: Trajectory plots for tumour growth with mean at each time point overlaid

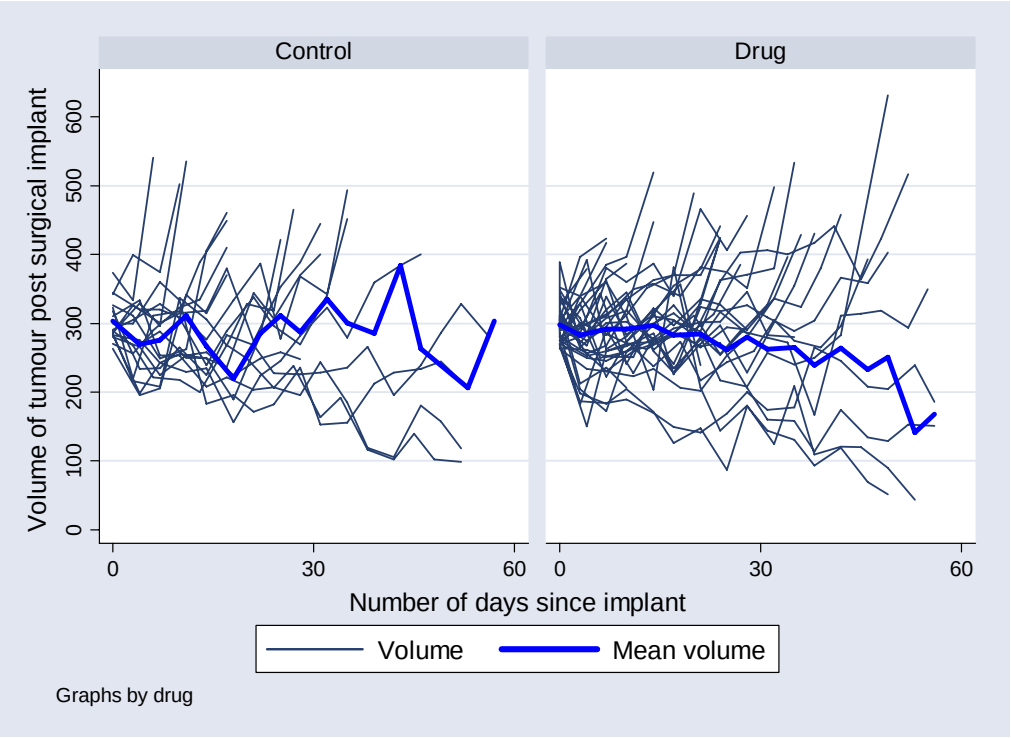
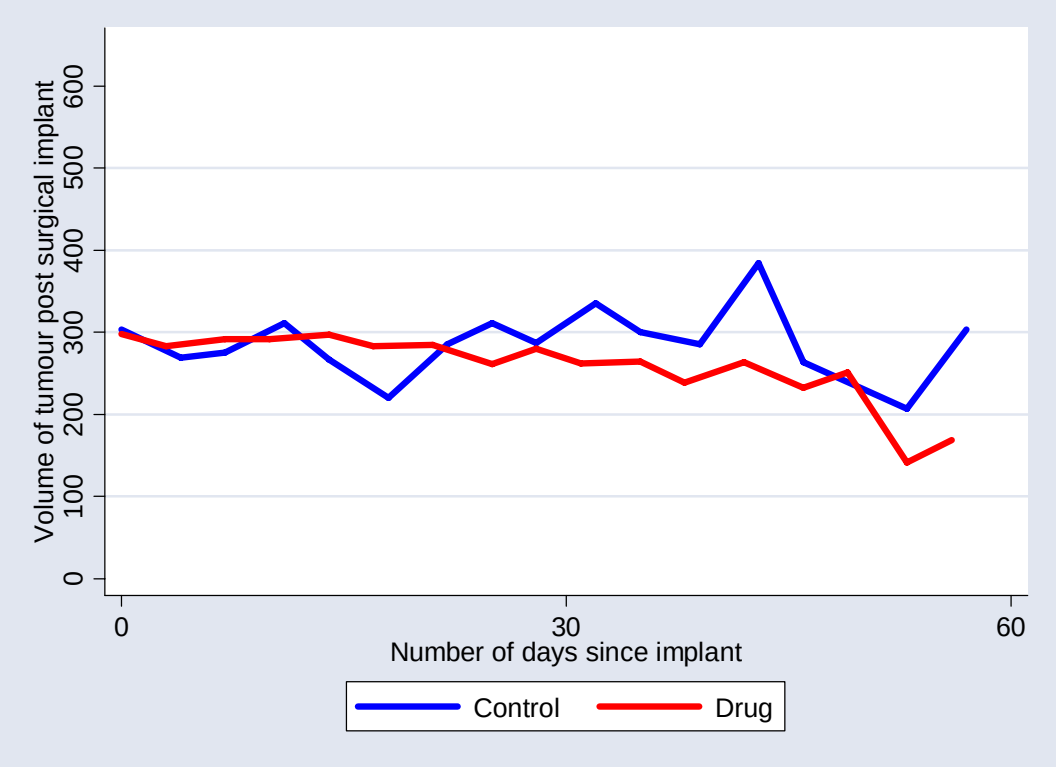


Figure 3: Mean tumour volume by drug group



The trajectory plots show little indication that the drug groups tumours are growing at a slower rate. Figure 3 plots on the same graph these mean tumour volumes. This plot again provides little evidence of any difference between the two groups in tumour volume.

Two different random effects models were fitted to the data. One assumed that the growth in tumour volume was linear, ie tumour volumes changed at a constant rate. The second assumed that growth was quadratic, ie the rate of change in tumour volumes changed over time. The quadratic model would thus be able to capture the change in tumour volumes that Dr Richards had described in his original description of the study - "after initial inhibition or shrinkage of tumours with castration, they will all begin to grow again.

The models were constructed as two-level models. The level 1 model describes how each mouse's tumour volume changes with time, and the level 2 model describes how these changes differ between mice. So for the linear model, level 1 states that an individual mouse's tumour grows in a linear fashion and level 2 states that the starting volume (intercept) and the rate of change of tumour volume (slope) for each mouse may be different. One advantage of specifying

the model in two levels rather than directly specifying the mixed model is that it ensures that in the final mixed model each random effect had a corresponding fixed effect.

Model assuming linear change

Level one model

$$Y_{ij} = \pi_{1i} + \pi_{2i} post\_surg_{ij} + \varepsilon_{ij}$$

Y_{ij} = volume of tumour for mouse *i* at time *j*

π_{1i} = intercept for mouse *i* (ie tumour volume at time 0)

π_{2i} = slope for mouse *i* (ie rate of change in tumour growth per day)

$\varepsilon_{ij} \sim N(0, \sigma^2)$ residuals around the fitted line for mouse *i*

Level two model

$$\pi_{1i} = \beta_{10} + \beta_{11} drug_i + \lambda_{1i}$$

β_{10} = average population intercept for mice in drug group 0 (control mice)

β_{11} = change in the average population intercept associated with being in drug group 1 (drug mice)

$drug_i$ = drug group for mouse *i*. 0 if control mouse, 1 if drug mouse.

$$\pi_{2i} = \beta_{20} + \beta_{21} drug_i + \lambda_{2i}$$

β_{20} = average population slope for mice in drug group 0 (control mice)

β_{21} = difference in the average population slope associated with being in drug group 1 (drug mice)

$\begin{pmatrix} \lambda_{1i} \\ \lambda_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{11} & \tau_{21} \\ \tau_{21} & \tau_{22} \end{pmatrix} \right]$ variation between mice in intercepts (τ_{11}) and slopes (τ_{22}) and covariance between the two (τ_{21}).

Combined model

$$Y_{ij} = (\beta_{10} + \beta_{11} drug_i + \lambda_{1i}) + (\beta_{20} + \beta_{21} drug_i + \lambda_{2i}) post\_surg_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = \beta_{10} + \beta_{11} drug_i + \beta_{20} post\_surg_{ij} + \beta_{21} drug_i * post\_surg_{ij} + (\lambda_{1i}) + (\lambda_{2i}) post\_surg_{ij} + \varepsilon_{ij}$$

Volume since surgery - quadratic

Level one model

$$Y_{ij} = \pi_{1i} + \pi_{2i} post\_surg_{ij} + \pi_{3i} (post\_surg_{ij})^2 + \varepsilon_{ij}$$

Y_{ij} = volume of tumour for mouse *i* at time *j*

π_{1i} = intercept for mouse *i* (ie tumour volume at time 0)

π_{2i} = rate of change in tumour growth at time 0 for mouse *i*

π_{3i} = curvature of the quadratic for mouse *i*

$\varepsilon_{ij} \sim N(0, \sigma^2)$ residuals around the fitted line for mouse *i*

Level two model

$$\pi_{1i} = \beta_{10} + \beta_{11} drug_i + \lambda_{1i}$$

β_{10} = average population intercept for mice in drug group 0 (control mice)

β_{11} = change in the average population intercept associated with being in drug group 1 (drug mice)

$drug_i$ = drug group for mouse *i*. 0 of control mouse, 1 if drug mouse.

$$\pi_{2i} = \beta_{20} + \beta_{21} drug_i + \lambda_{2i}$$

β_{20} = average population rate of change in tumour growth at time 0 for mice in drug group 0 (control mice)

β_{21} = difference in the average population rate of change in tumour growth at time 0 associated with being in drug group 1 (drug mice)

$$\pi_{3i} = \beta_{30} + \beta_{31} drug_i + \lambda_{3i}$$

β_{30} = average population curvature of the quadratic for mice in drug group 0 (control mice)

β_{31} = difference in the average population curvature of the quadratic for associated with being in drug group 1 (drug mice)

$\begin{pmatrix} \lambda_{1i} \\ \lambda_{2i} \\ \lambda_{3i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{11} & \tau_{21} & \tau_{31} \\ \tau_{21} & \tau_{22} & \tau_{32} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{pmatrix}\right]$ variation between mice in intercepts (τ_{11}), slopes (τ_{22}) and curvature (τ_{33}) and covariance between the three parameters.

Combined model

$$Y_{ij} = (\beta_{10} + \beta_{11} drug_i + \lambda_{1i}) + (\beta_{20} + \beta_{21} drug_i + \lambda_{2i}) post\_surg_{ij} + (\beta_{30} + \beta_{31} drug_i + \lambda_{3i})(post\_surg_{ij})^2 + \varepsilon_{ij}$$

$$Y_{ij} = \beta_{10} + \beta_{11} drug_i + \beta_{20} post\_surg_{ij} + \beta_{21} drug_i * post\_surg_{ij} +$$
$$\beta_{30}(post\_surg_{ij})^2 + \beta_{31} drug_i * (post\_surg_{ij})^2 + (\lambda_{1i}) + (\lambda_{2i}) post\_surg_{ij} + (\lambda_{3i})(post\_surg_{ij})^2 + \varepsilon_{ij}$$

The data were also centred at 14 days post surgery. The reason for this was that in the models specified the parameter estimates provide the differences between the two groups at time zero. If time zero was taken to be the day of surgery then we would expect there to be no difference between the groups in their rate of change in tumour volume as they had not begun their different treatments. Fourteen days post treatment was decided as a reasonable time for treatment difference in rate of change to become apparent.

These two models were fitted using PROC MIXED (see appendix for program). Goodness of fit was assessed using AIC and BIC (Table 7), where lower values imply a better fit. The REML criteria were not used to judge goodness of fit as the two models that have been fitted have different mean models specified. This means that the REML functions are based upon different observations and so can no longer be directly compared (Verbecke, 2000). The model that assumed quadratic change in tumour volume fitted the data better using the AIC and BIC measures as they were both lower for the quadratic model.

Table 7: Model fit statistics

	AIC	BIC
Quadratic increase in volume	5564	5578
Linear increase in volume	5634	5642

The fitted population average curve for the control mice is

$$Y = 319.1 + 4.3\,post\_surg + 0.183 * \left(post\_surg\right)^2$$

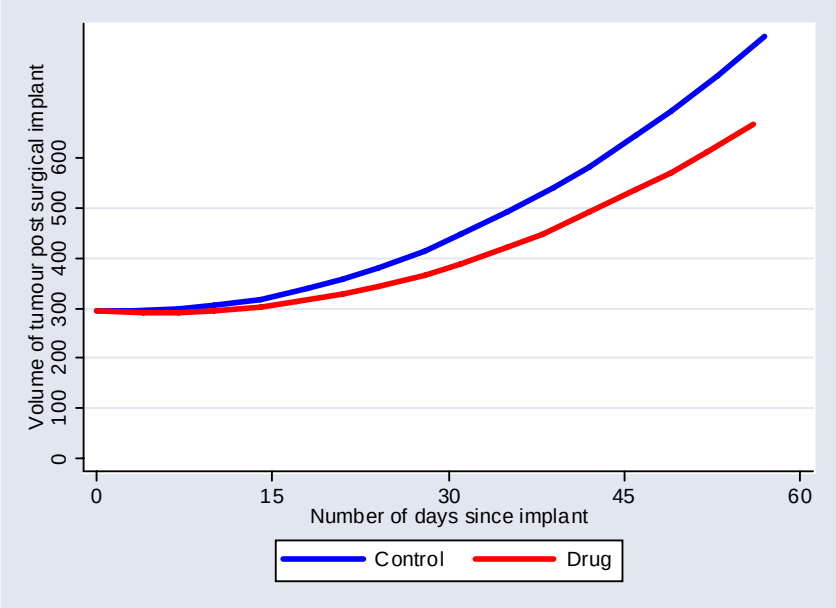
and for the mice in the drug group is

$$Y = 302.1 + 2.6\,post\_surg + 0.145 * \left(post\_surg\right)^2$$

Figure 4 shows the fitted population average growth curve using the quadratic model for drug and control groups. There was no evidence of any difference between these curves ($F_{2,19.1} = 0.54$, p = 0.59). There was no evidence of a difference in rate of growth at day 14 (t = -0.94, df = 18.5, p = 0.37) or in the curvature of the lines (t = -0.73, df = 20.7, p = 0.47). Centring the data at 28 days rather than 14 days had no effect on results. Centring beyond 28 days resulted in the model failing to converge.

Note that as multiple random effects have been modelled and the data are unbalanced the t and F tests are only approximate, with the degrees of freedom estimated by the Satterthwaite approach. This method often produces non-integer values for the degrees of freedom as seen here.

Figure 4: Fitted population average growth curves



These curves look quite different to the observed values. The reason for this is that there are very few mice with observations at the end. Fitting the individual predicted curves we see that the model fits the data quite well (figure 5). This is also demonstrated by plotting the average of the observed and predicted volumes of these mice still alive (figure 6).

Figure 5: Trajectory plot showing observed tumour growth and subject specific predicted growth curves

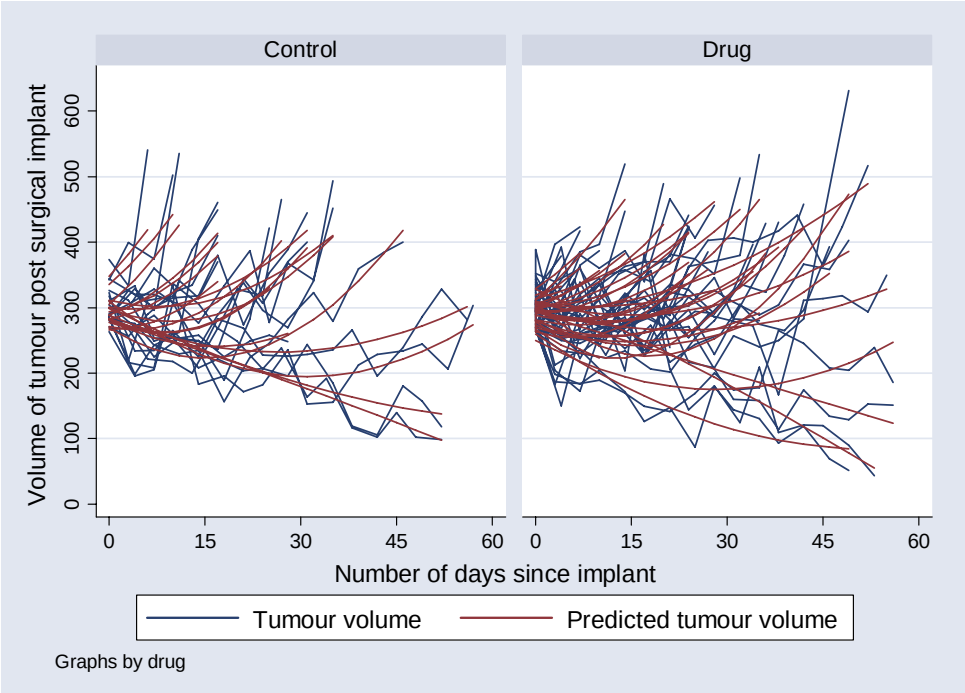
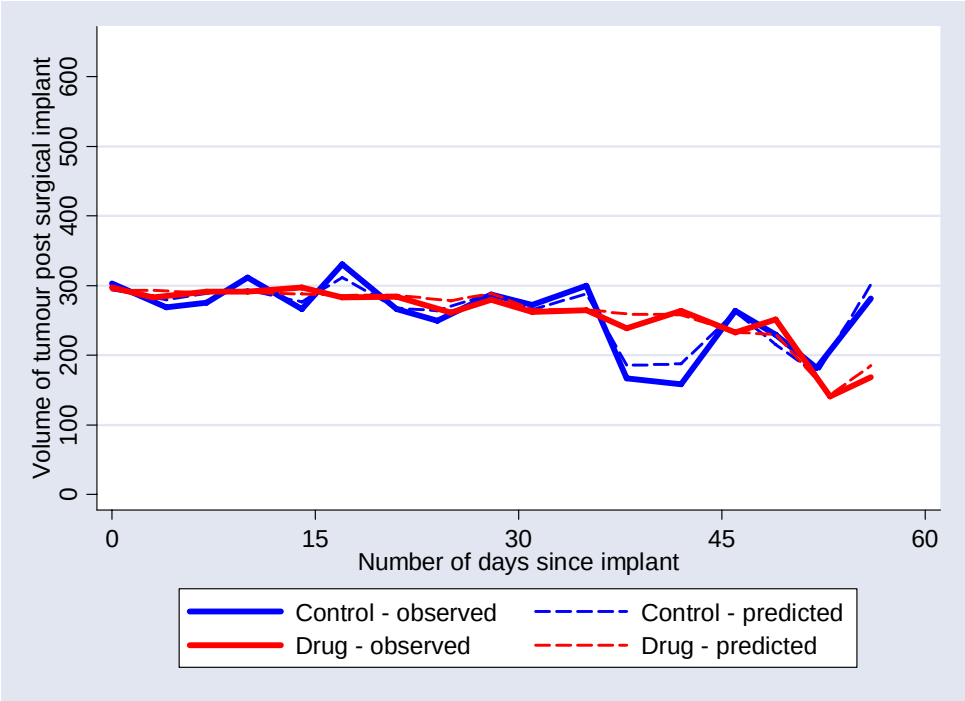


Figure 6: Mean of observed and predicted tumour volumes



Discussion

The analysis has provided little evidence that the use of NSAIDs delays the growth in prostate tumours and hence prolongs the time until androgen independence. Two methods of comparing the change in tumour volume have been used. The log-rank test compared the time until androgen independence for the two groups and the linear mixed models approach compared the rate of change in tumour volume. One of the reasons for a failure to detect any difference may have been the low power of the experiment.

The log-rank test had very low power (30%) to detect the stated difference of an increase in time to androgen independence of 50%. An increase of 50% is a rather large increase to detect. A smaller intended detectable difference would have resulted in an even lower power. Therefore the problem of low power in this experiment is the sample size used, and perhaps the unequal allocation between groups.

There were only 56 mice that were included in the final analysis. These are the mice whose tumours reached the required initial volume. Thirty-seven of these mice were in the drug group

and 19 in the placebo: ratio of roughly 2:1. In experiments with a given sample size a more powerful study usually results when equal numbers of subjects are in the comparison groups. In some studies this is not possible, for example case-control studies of rare diseases and hence multiple control subjects may be chosen for each case. However, the total number of subjects is less for an equivalently powered study where there are equal numbers. In this experiment if there had been equal numbers in the drug and control group (28 in each group for a total of 56) the power would have increased to 33% from 30% with the unequal groups used - a very small gain in power.

This means that the only way that this study could be repeated and have any chance of identifying a meaningful difference of 50% would be to increase the sample size. In order to achieve the standard 80% power, with a significance level of 0.05 and a difference to be detected of 50% there would have to be 192 mice in total, assuming equal allocation, and 214 mice, assuming the ratio of drug to placebo mice was 2:1.

In my advice to Dr Richards I had stated that the advantage of using the mixed models approach rather than the log-rank test would be that it would provide a more powerful test. The only justification I gave for this statement was that the mixed models approach uses more data than the log-rank test. The mixed models approach used all of the weights recorded and the times that they were recorded, whereas the log-rank test only uses the time that the tumour volume takes to increase by a certain size. Matthews in his description of the summary measures approach cautions against being seduced by the vast amounts of data that are available in longitudinal studies. However, it is true that the mixed models approach will provide a more powerful test as the outcome is continuous.

The book Analysis of Longitudinal (Diggle, 1994) data provides a formula for the sample size required in longitudinal studies where the measure of interest to compare between two groups is the rate of change. The formula for the sample size is

$$m = \frac{2\left(z_\alpha + z_\beta\right)^2 \sigma^2 \left(1-\rho\right)}{n s_x^2 d^2}$$

z_α and z_β are the usual points on the standard Normal distribution based on the 2-sided significance level (α) and the power (1-β). The value σ^2 and ρ are the variance of the variable under study (eg tumour volume) and the correlation of these repeated measurements within

subjects – the formula assumes a constant variance and correlation over time. The number of measurements per subject is n and s^2_x gives the variance of the measurement times. The smallest detectable difference in the rate of growth is given by d.

We can apply this formula to the mice experiment in the following way. On average there were 3.5 days between measurements, allowing for 9 measurements per mouse total follow-up would have been 31.5 days (the maximum a mouse was followed for was actually 56 days).

The median time to reach androgen independence for the control group was 30 days, and androgen independence was defined as an increase in tumour volume of 105 mm^3. This equates to a rate of growth of 3.5mm^3 per day. An increase in the median time to androgen independence of 50%, to 45 days, would equate to a rate of growth of 2.3mm^3 per day. So the difference in rate of growth we wish to detect is 1.2mm^3 per day.

From the data the estimate of σ (=71.2) was obtained by taking the average of the standard deviations of the volumes measured between the time of surgery and 30 days following it. Strictly, σ^2 is the variation in tumour volumes not explained by the model. The estimate obtained from the data would be an overestimate of this because, although the variation due to time has been accounted for by taking the average variation at different time points, the variation due to the random the random intercepts and slopes of the mice has not. The overestimation of σ will result in an overestimation of the sample size. The correlation (ρ) between measurements at surgery and 3 days post surgery was 0.28. The table below displays the sample size required in each group for various σ and ρ.

Table 8: Sample size required in each group to detect a 50% change in median time until androgen independence with 80% power and significance of 0.05

		ρ			
		0.1	0.3	0.5	0.8
σ	50	27	21	15	6
	70	52	41	29	12
	90	86	67	48	19

The number required when at the higher variance and when the correlation between repeated measures is lowest (m = 86) gives a total sample size required of 172 compared to 192 when a

log-rank test comparing the survival curves is used. In the current data the correlation was close to 0.3 and the standard deviation closer to 70 which would suggest the total sample size required would be 82 (41 in each group). This is larger than the sample size that the experiment was begin with (73) and the number that were finally available to be analysed (56).

There are, of course, a few caveats that much be attached to the above sample size calculations. The formula assumes that the variance is constant over time and that there are equal numbers of subjects in each group. This has partly been addresses by taking the average of the variances over a period of 30 days and the sample size can be adjusted for unequal numbers by using the formula –

$$m = \frac{(r+1)n}{2r}$$

m is the number of subjects in one group, *rm* is the number in the other group and *n* is the sample size required if the two groups have equal numbers of subjects (SMMR). Using this formula, a study with 41 in each group would be equivalent to having 31 in one group and 62 in the other – a total sample size of 93. This is still less than the 192 required for the survival analysis.

There are two issues that remain to be resolved with the use of the above sample size formula. The formula here assumes a linear increase in tumour volume. The data actually show a quadratic increase which means more parameters have to be estimated. Hence, a greater sample size would need to be used to provide a similar power.

A final issue is that the calculations assume that each mouse has a measurement at each of the time points and every mouse has the same number of measurements. This is not the case in this experiment due to the censoring of the data. Again this is likely to inflate the sample size required for a given power. The power of the test, and hence sample size required, in repeated measures depends not only the number of measurements per individual but also the pattern of missing observations (Tu, 2004).

References

Australian Institute of Health and Welfare (2006). Cancer FAQs. Avaliable at http://www.aihw.gov.au/cancer/faqs.cfm#ccancer Last accessed 24 June 2006.

Armitage P, Berry G, Matthews JNS. (2002) Statistical Methods in Medical Research. Oxford: Blackwell Science.

Cnaan A, Laird NM, Slasor P. (1997) Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Statistics in Medicine, Vol 16, 2349-2380.

Diggle PJ, Liang KY, Zeger SL. (1994) Analysis of longitudinal data. New York: Oxford University Press.

Dupont WD, Plummer WD. (1997): PS power and sample size program available for free on the Internet. Controlled Clin Trials,1997;18:274

Eigl BJC, Eggener SE, Baybick J, Ettinger S, Chi KN, Nelson C, Wang Z, Gleave ME. (2005) Timing is everything: Preclinical evidence supporting simultaneous rather than sequential chemohormonal therapy for prostate cancer. Clinical Cancer Research; 11(13) July 1, 2005.

Hosmer DW, Lemeshow S. (1999) Applied survival analysis: Regression modeling of time to event data. New York: Wiley.

Lee ET. (1992) Statistical Methods for Survival Data Analysis. New York: Wiley.

Liang KY, Zeger SL. (1993) Regression analysis for correlated data. Annual Review of Public Health 1993. 14: 43-68.

Matthews JNS, Altman DG, Campbell MJ, Royston P. (1990) Analysis of serial measurement in medical research. British Medical Journal 300, 230-235.

Nicholson B, Gulding K, Conaway M, Wedge SR, Thedorescu D. (2004) Combination antiangiogenic and androgen deprivation therapy for prostate cancer: A promising therapeutic approach. Clinical Cancer Research; Vol 10 8278-8734, December 15, 2004.

Power and Sample size program

Singer JD, Willett JB (2003) Applied longitudinal data analysis: Modeling change and event occurrence. New York: Wiley.

Tu XM, Kowalski J, Zhang J, Lynch KGM Crits-Christoph P. (2004) Power analysis for longitudinal trials and other clustered designs. Statistics in Medicine 2004: 23;2799-2815.

Verbecke G, Molenberghs G. (2000) Linear mixed models for longitudinal data. New York: Springer.

Appendix 1

```
* sort out data for proc mixed;
data wpp.mice_post_mixed;
set wpp.mice_post;

* centre post_surg at 14 days;
post_surg_cen = post_surg -14;

* squared term;
post_sq = post_surg_cen*post_surg_cen;

*class variable;
cpost = post_surg_cen;

run;


* post surgery;
*linear;
title "Linear";
proc mixed data=wpp.mice_post_mixed covtest;
where volume ne .;
  class id cpost;
  model volume= drug post_surg_cen drug*post_surg_cen / solution ddfm =
satterth;
  random intercept post_surg_cen / subject=id type=un;
  run;


* quadratic;
  title "Quadratic";
proc mixed data=wpp.mice_post_mixed covtest;
where volume ne .;
  class id cpost;
  model volume= drug post_surg_cen post_sq drug*post_surg_cen
drug*post_sq/
            solution ddfm = satterth  outp = test outpm = test2;
  random intercept post_surg_cen post_sq/ subject=id type=un;
  repeated cpost;
    contrast "Contrast" drug*post_surg_cen 1,
                        drug*post_sq 1 ;
  run;
```

Project B:
The relationship between chronic kidney disease and individual-level Socio-economic Status: the Three Continent Kidney Disease (3CKD) Study

Project B

Project title

The Relationship Between Chronic Kidney Disease and Individual-Level Socio-Economic Status: The Three Continent Kidney Disease (3CKD) Study

Location and dates

March – June 2006, School of Public Health, University of Sydney

Context

Researchers from the School of Public Health and the George Institute at the University of Sydney collaborated on a project to investigate the association between socio-economic status (SES) and chronic kidney disease (CKD). The researchers had access to three different population surveys from the USA, Thailand and Australia. Each of the surveys recorded measures of kidney function and also information that could be used as measurements of socio-economic status. The measurements chosen to use as measures of SES were educational status, total household income and employment status. Ethnicity was also recorded in one of the surveys and a secondary aim of the study was to investigate whether the association between SES and CKD was similar across different ethnic groups.

Student contribution

Several meetings were held with all the researchers to plan the analysis and discuss the results. I also met with Dr Mike Jones on many occasions to discuss the ongoing analysis and the responses to questions from the other researchers.

I carried out all data management tasks, investigated appropriate ways to code the selected variables, performed the agreed statistical analysis and provided interpretation of the results. I also carried out additional analyses to provide supporting evidence for the choices made regarding the coding of variables.

Statistical issues

The primary research question was addressed using the relatively simple method of logistic regression. However, as each of the surveys had complex sampling designs this had to be taken into account in order to obtain correct parameter estimates and standard errors.

Other statistical issues encountered were how to assess the agreement between two measurements and which type of model is most appropriate in combining results from studies.

Acknowledgements

I would like to thank Dr Mike Jones for the time he has given me and his invaluable advice. I'd also like to thank the other researchers on the project and Dr Timothy Dobbins for providing advice on the project as it developed.

Student declaration

I declare that this project is my own work, with guidance provided by my project supervisor, Dr Mike Jones, and that I have not previously submitted it for academic credit.

_____

Kevin McGeechan

_____

Date

Supervisor declaration

I supervised Kevin's involvement in this project which was conducted jointly with staff from the School of Public Health, George Institute and Royal Prince Alfred Hospital. Kevin took the lead role with only minimal input from me in data management and the design, implementation and interpretation of the statistical analysis and its input into a conference abstract and draft manuscript which is planned to be submitted to a peer-reviewed journal shortly. At team meetings Kevin was indistinguishable from the more experienced researchers around the table with respect to contribution of creative ideas to the group. I am confident that my views are shared by other members of the collaboration.

_____

Dr Mike Jones

_____

Date

Introduction

Researchers from the School of Public Health and the George Institute at the University of Sydney collaborated on a project to investigate the association between socio-economic status (SES) and chronic kidney disease (CKD). The researchers had access to three different population surveys from the USA, Thailand and Australia. Each of the surveys recorded measures of kidney function and also information that could be used as measurements of socio-economic status. The measurements chosen to use as measures of SES were educational status, total household income and employment status. Ethnicity was also recorded in one of the surveys and a secondary aim of the study was to investigate whether the association between SES and CKD was similar across different ethnic groups.

The three surveys used in this analysis were each national representative surveys, however none were designed with the specific aim of investigating the association between chronic kidney disease (CKD) and socioeconomic status (SES). The National Health and Nutrition Survey (NHANES III) which was carried out between 1988 and 1994 is one of a series of surveys undertaken in the USA to provide national estimates of the health and nutritional status of the US population (National Center for Health Statistics Centers for Disease Control and Prevention, 1996). The InterAsia Collaborative Study of Cardiovascular Disease in Asia (InterAsia), carried out between July 2000 and March 2001, was designed to provide estimates of cardiovascular risk factors in Thailand and China (He, 2004). The Australian Diabetes, Obesity and Lifestyle Study (AusDiab), carried out between May 1999 and December 2000, was designed with the primary aim of providing national and state estimates of the prevalence of diabetes (Dunstan, 2002). However, each of the studies recorded information on creatinine levels, which were used to provide an indicator for the presence of CKD, socioeconomic status and other factors that should be taken into account when assessing the relationship between CKD and SES.

Data management

The first task was to combine the data from the three surveys into a common dataset to be analysed. This involved translating the datasets into a common format, identifying the SES constructs measured in common across the surveys and the data items by which they were represented and coding these to a common system. The datasets from each of the surveys were stored in three different formats. The NHANES data was stored as SAS transport files, the InterAsia data was stored as a Stata file and the AusDiab data was stored as SPSS datasets with its own specially designed interface where specific SPSS datasets could be created. Each of the datasets was translated into SAS datasets for analysis.

The variables to be used in the analysis then had to be created from the data recorded in each of the surveys. Each survey asked questions slightly differently and recorded information in different coding schemes. Questions that could be used or combined to provide a common variable had to be found and the coding scheme standardised. There was extensive documentation available for the NHANES study which made examining the 3,668 variables recorded for NHANES relatively easy. There was also good documentation for AusDiab and the subset of data that we requested for this analysis had only 174 variables. There was minimal documentation for the InterAsia data which made the examination of its 412 variables more difficult.

Survey design

While the data management tasks were undertaken several meetings took place with the co-investigators to clarify the research questions and to agree on an analysis plan. After having identified the variables required for the analysis, standardised the coding for each survey and agreed on the analysis plan, the next step was to analyse the data. All of the studies used complex sampling designs which involved stratified sampling with unequal sampling fractions and clustering. This means that simple methods of analysis are no longer appropriate as these methods are based on the assumption that individuals within the dataset are independent of each other. The sampling sheme for each of the surveys means that this assumption no longer holds.

AusDiab was the simplest design in that it stratified the Australian population by state then within each state six census districts were randomly selected (a census district is the smallest geographical unit defined by the Australian Bureau of Statistics and on average consists of 225 dwellings). These census districts formed the clusters and within the cluster all residents were invited to take part in the survey. The InterAsia used four levels of stratification – 1. provinces, 2. political districts, 3. slum or non-slum enumeration district and 4. age group by gender. As individuals come from the same area the data are also clustered so this also has to be taken into account. NHANES had four levels of clustering. The first level was the county level and 81 counties were randomly selected. These were then randomly split between the two time periods over which the survey was carried out (phase 1 was carried out over the period 1988-1991 and phase 2 was carried out from 1991 to 1994). Within each county, city or suburban blocks were then randomly selected and within a block households of individuals were selected.

Each of the datasets also contained sampling weights to be used in any analysis. The sampling weights are the inverse of the probability of being sampled (in NHANES African-Americans and Mexican-Americans were over sampled and so had a higher probability of being sampled). These weights also provided adjustments for non-respondents.

The choice of sampling design was based on practical as well as statistical considerations. For example, the AusDiab study sampled equal numbers from each state. A more precise estimate of the national prevalence of diabetes would have been obtained if numbers were sampled proportional to the population of each state. However, a secondary aim of AusDiab was to provide estimates at the level of each state therefore precision at the national level was sacrificed in order to obtain this. Another example is the over sampling in NHANES of African-Americans and Mexican-Americans.

The sampling design needs to be taken into account when estimates are calculated. Estimates derived from these studies need to take into account the sampling weights to provide accurate estimates of prevalence. The analysis also needs to take into account the stratification and clustering within the design. These elements of the sampling design will affect the estimates of the variance of parameters. For example, ignoring the clustering within a dataset would generally provide lower estimates of the variance, and hence standard errors, as the individuals would be more alike than if the sample had been obtained from a simple random sample. Lower standard errors will result in lower p-values thereby increasing the type I error rate and hence the chance of declaring an association between SES and CKD as significant when in fact no relationship exists.

Methods of analysis

The association between SES and CKD was estimated using logistic regression. This method was chosen because the definition of CKD used in this study was the dichotomous outcome of whether the person had a glomerular filtration rate (GFR) below 60ml/min/1.73m^2 or not. The design of each of the surveys was taken into account when analysing the data by using the SAS procedure PROC SURVEYLOGISTIC. This procedure carries out logistic regression where the data have been collected from complex surveys and it requires that the variables that identify the strata, primary sampling units and sampling weights be specified. Each survey dataset provided fields for primary sampling units and stratification variables.

A random effects model was also used to combine the estimates across the three surveys using techniques from meta-analysis. The random effects model was chosen as it was assumed that the effect of SES varied between the countries, and that this variation was not simply due to sampling variability. The fixed effects model would assume that there is one true measure of the effect of SES on CKD and that the measures between surveys only differ because of sampling variability. In the random effects model it is assumed that the effect of SES on CKD is itself a random variable.

The assumptions for the random effects model appear to be more plausible for these data. It is unlikely that the effect of SES on CKD would be the same in each population given the wide cultural and also health access differences between the countries. Also, for some measures of SES a relative measurement was used (eg below versus above median income). For such variables, how big the effect of SES on CKD is will depend upon how much variation there is in the SES measure.

Comparison of two formulae to estimate GFR

Measuring agreement using the Bland-Altman plot

The outcome investigated in this study was a glomerular filtration rate (GFR) below 60ml/min/1.73m^2. GFR is a measure of kidney function and a GFR below 60ml/min/1.73m^2 is one commonly used indication of chronic kidney disease. The GFR is calculated using the serum creatinine and other factors such as age, race or body surface area depending on which formula is used. Two of the most common formulae used to calculate GFR are the MDRD formula (named after the US Modification of Diet in Renal Disease Study) and the Cockcroft-Gault formula. (The Australasian Creatinine Consensus Working Group, 2005).

The GFR calculated using the MDRD equation was the method employed in this study as it is considered to be the more widely validated and is currently recommended by The Australasian Creatinine Consensus Working Group (2005). However, as a previous publication investigating the prevalence of kidney disease amongst the Australian population (Chadban, 2003) used the Cockcroft Gault formula as at the time it was the more commonly used formula in Australia, it was of interest to compare whether using either equation influenced the estimated association between SES and CKD. If the two equations provide similar estimates of GFR (ie agree with each other) then it would not matter which equation was used as to estimate GFR as the effect on the association between SES and CKD would be small.

One way to investigate the agreement between the two measurements would be to plot each variable against the other. The resulting plot should be on the diagonal line of equivalence if the measures agreed. The data from each of the three surveys is shown in Figures 1, 2 and 3. Also included on the plots are reference lines at GFR equal to 60ml/min/1.73m^2. Simply calculating the correlation between the two measures does not give an indication on whether the measures agree. For instance, their may be perfect correlation but one measurement is consistently half of the other measurement.

It is difficult to determine whether the data lie on the diagonal due to the number of observations although it is clear that the variation increases as the estimated GFR increases. Bland and Altman have suggested a more appropriate graph to compare two measurements which plots the difference between the two values against their average (these are shown in figure 4, 5 and 6) (Bland, 1986). If the measures agreed then the data points should be a straight band centred around the horizontal zero line, with no fanning out or tendency to increase or decrease. (The SAS program for the Bland-Altman plots was based on the program presented by Schneider at the SAS Users Group International conference (2001).

Figure 1: Comparison of GFR estimated by the MDRD and Cockcroft Gault methods using NHANES data

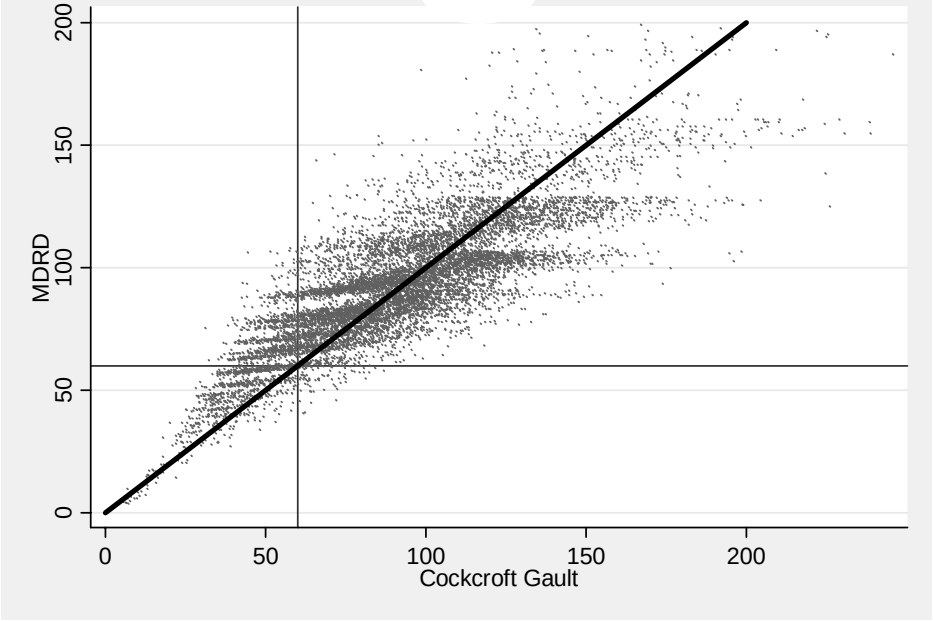


Figure 2: Comparison of GFR estimated by the MDRD and Cockcroft Gault methods using InterAsia data

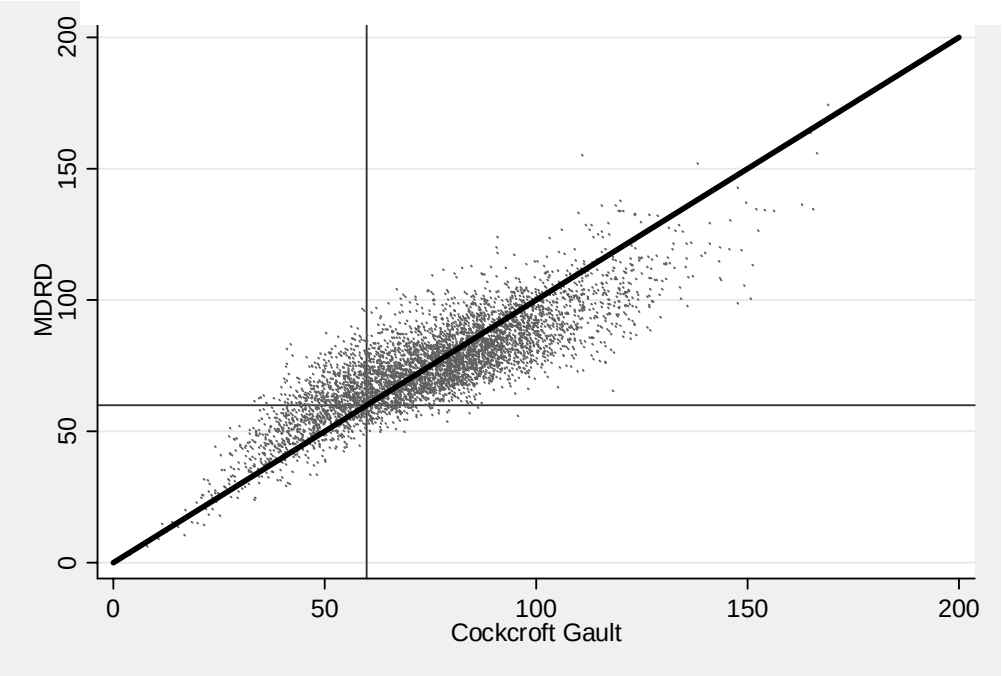


Figure 3: Comparison of GFR estimated by the MDRD and Cockcroft Gault methods using AusDiab data

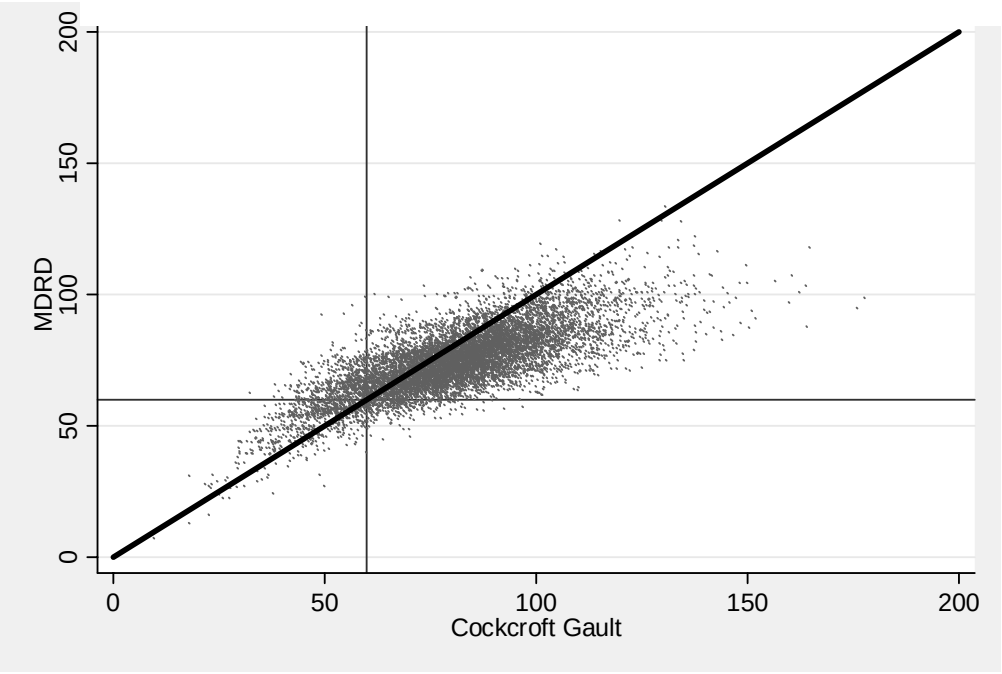


Figure 4: Bland-Altman plot for GFR using NHANES data

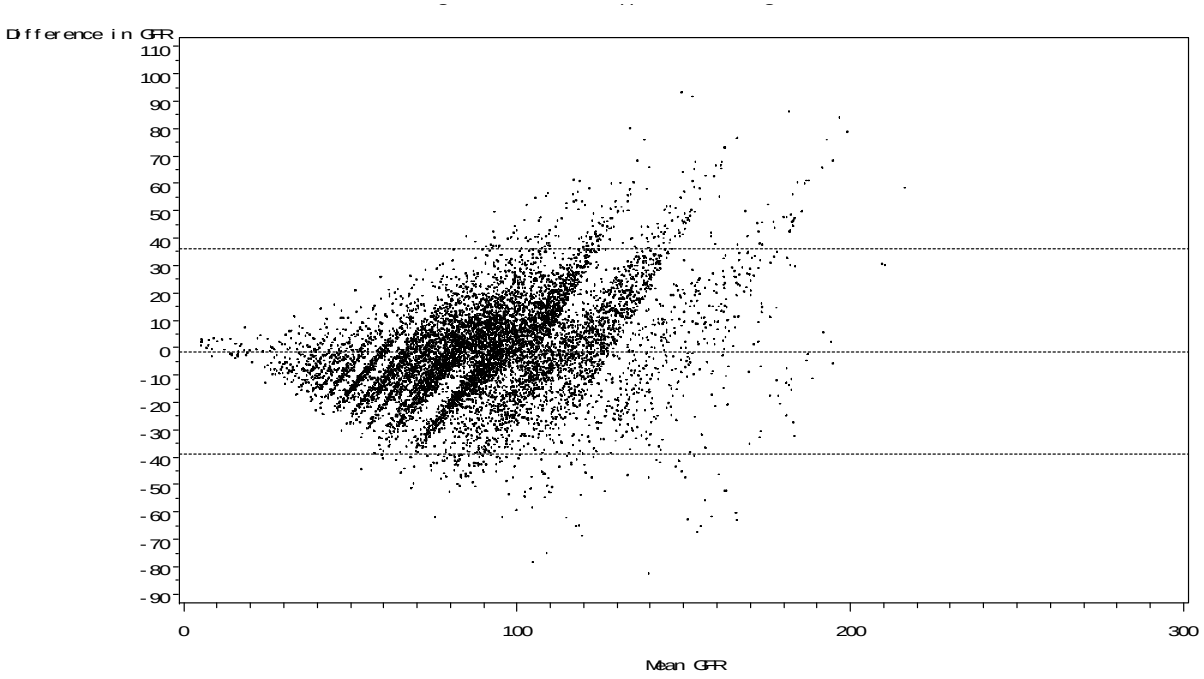


Figure 5: Bland-Altman plot for GFR using InterAsia data

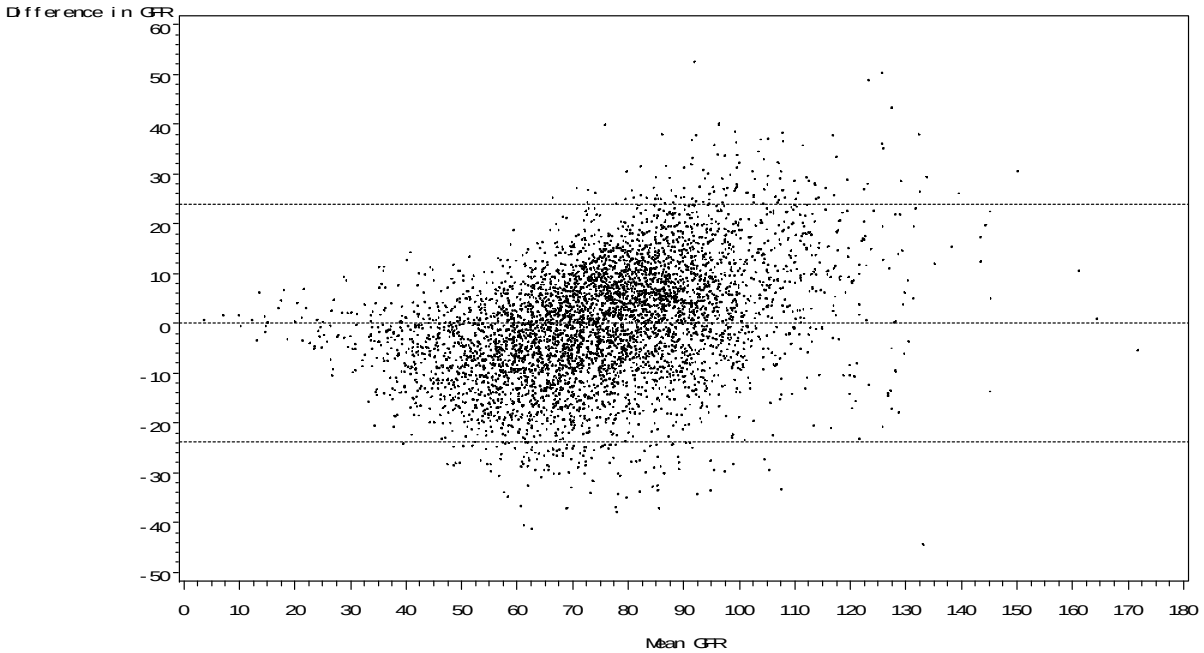
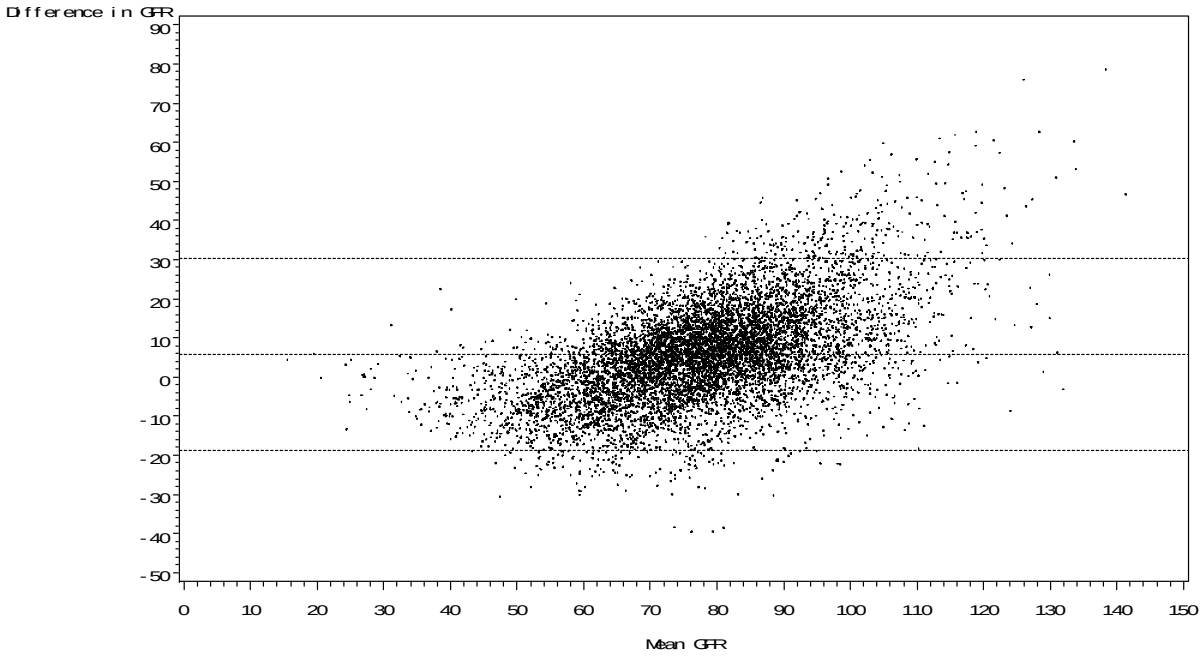


Figure 6: Bland-Altman plot for GFR using AusDiab data



In all three studies there is a clear increase in variation between measurements as the GFR increases. This indicates increasing disagreement between the estimates as the true value of GFR increases. Also, in all three studies there appears to be a trend to an increasing difference between estimates as the GFR increases. This is particularly evident for the AusDiab study. This suggests that the Cockcroft Gault equation estimates a higher GFR value than the MDRD equations at higher values of GFR, and lower GFR values at lower values.

Measuring agreement using the Kappa coefficient

The above analysis has investigated whether the estimated GFR values agree when computed using the two equations. In this study however the GFR estimate was not used directly but rather the GFR values were categorized into low (<60ml/min/1.73m^2), which would indicate the presence of CKD, or not (≥60ml/min/1.73m^2). It would be appropriate then to also determine whether the classification of people as having low GFR was the same when the two equations are used. The Kappa coefficient (Armitage, 2002) was used to determine this.

The Kappa coefficient measures the agreement between categorical variables over and above that which may have arisen by chance. The formula for Kappa is

$$\kappa = \frac{I_o - I_e}{1 - I_e}$$

where I_o is the observed agreement and I_e is the expected agreement. The Kappa coefficient can be supplemented by two measurements which provide information on whether the two classifications agree when the condition is present and when the condition is absent. The formulae for the two measurements are

$$p_{pos} = \frac{2a}{2a + b + c}$$
$$p_{neg} = \frac{2d}{2d + b + c}$$

where p_{pos} is the number of agreed positives divided by the average of those classed as positives by the two classifications. p_{neg} is similarly defined for the negatives, and *a, b, c* and *d* are the usual four cells from the 2x2 table of agreement.

Table 1: Comparison of people classified as low GFR using two methods

		NHANES		**InterAsia**		**AusDiab**	
		GFR estimated by Cockcroft Gault					
		<60	**≥60**	**<60**	**≥60**	**<60**	**≥60**
GFR estimated by MDRD	<60	858	76	741	120	710	267
	≥60	891	8770	565	3672	622	8134
	Kappa	0.59		0.60		0.57	
	p_{pos}	0.64		0.68		0.62	
	p_{neg}	0.94		0.91		0.95	

The agreement in each of the studies is reasonable (but also far from perfect) between the two classifications, and for each study there is more agreement in classifying subjects who have a GFR≥60 ml/min/1.73m^2, than below 60ml/min/1.73m^2, as indicated by the high values of p_{neg} and the lower values of p_{pos} (Table 1) . Although the Bland-Altman plot indicates that the Cockcroft Gault and the MDRD equations are not producing equivalent estimates of GFR, the Kappa coefficients suggest that once people are classified as having low GFR then the difference between the two equations matters less. However, the Cockcroft Gault formula does classify a higher proportion of people as having low GFR.

In this study a low GFR (below 60ml/min/1.73m^2) is regarded as indicating chronic kidney disease and the purpose of the study was to examine whether various SES factors were associated with CKD. Table 2 below presents the odds ratios calculated for the three different SES factors under study (education, income and employment) when CKD is determined using the MDRD formula or the Cockcroft-Gault formula. These odds ratios are presented separately for the three surveys – NHANES, InterAsia and AusDiab. Using either method to define CKD there appears to be an increase in risk of CKD with lower SES across all three surveys, however many of the confidence intervals for the odds ratios contain one. There are no dramatic differences between the results when CKD is defined by the MDRD or Cockcroft-Gault formula.

Table 2: Effect on odds ratios when GFR is estimated using MDRD or Cockcroft-Gault

		MDRD			*Cockcroft Gault*		
Survey	***Effect***	***Odds Ratio***	***95% Lower Confidence Limit***	***95% Upper Confidence Limit***	***Odds Ratio***	***95% Lower Confidence Limit***	***95% Upper Confidence Limit***
NHANES	Education: Average vs Greater than average	1.01	0.79	1.30	1.01	0.77	1.31
	Education: less than average vs Greater than average	1.31	1.00	1.71	1.37	1.06	1.77
	Income: Below median income vs Above median income	1.31	0.97	1.76	1.12	0.88	1.42
	Employment status: Not employed vs Employed	1.12	0.86	1.45	0.99	0.73	1.34
	Employment status: Retired vs Employed	1.52	1.13	2.03	1.27	0.97	1.66
InterASIA	Education: Average vs Greater than average	1.35	0.96	1.90	1.38	0.99	1.92
	Education: less than average vs Greater than average	1.17	0.73	1.88	1.58	1.07	2.31

		MDRD			Cockcroft Gault		
Survey	Effect	Odds Ratio	95% Lower Confidence Limit	95% Upper Confidence Limit	Odds Ratio	95% Lower Confidence Limit	95% Upper Confidence Limit
	Income: Below median income vs Above median income	1.07	0.80	1.42	1.72	1.27	2.32
	Employment status: Not employed vs Employed	1.25	0.92	1.70	1.24	0.80	1.92
	Employment status: Retired vs Employed	1.11	0.80	1.53	1.00	0.68	1.49
AusDiab	Education: Average vs Greater than average	0.70	0.52	0.95	0.88	0.68	1.13
	Education: less than average vs Greater than average	1.07	0.88	1.30	1.07	0.73	1.58
	Income: Below median income vs Above median income	1.29	0.91	1.83	1.79	1.18	2.71
	Employment status: Not employed vs Employed	1.40	0.96	2.05	1.32	0.80	2.16
	Employment status: Retired vs Employed	2.01	1.28	3.17	2.03	1.31	3.14

Neither of the two methods for estimating GFR provides the true value of GFR therefore when these methods are then used to classify someone as below or above a certain GFR some people will be misclassified. Unfortunately, within the surveys there is no gold standard of GFR to determine which of the two estimates has less misclassification and would thereby provide a more accurate result. In logistic regression where there is misclassification in the outcome, the resulting estimate of the association between the outcome and risk factor will be biased (Luan 2005). The direction of this bias will depend on whether there is differential or non-differential misclassification in the outcome. If there is non-differential misclassification the bias would be towards the null value, but for non-differential misclassification the bias could be away from the null rather than towards. In the present

study if SES status is associated with GFR level then there may be non-differential misclassification with lower SES more likely to be misclassified. However, in this instance the estimated association would still be biased towards the null value. Although the estimates of the association between SES and CKD displayed in table 2 may be biased when either equation is used, the associations described appear to be broadly consistent no matter which equation is used.

Residual confounding

Residual confounding occurs when a confounder is not completely controlled for in an analysis. This could happen if the variable that measures the confounder is categorised into broad groups (Webb, 2005). Within each broad group the association between the outcome and the covariate may then still be affected by the true value of the confounder. In our analysis the outcome is CKD, the covariate is SES and the confounder is age which can be categorised into age groups.

It may seem strange to move from describing the outcome to describing whether residual confounding is an issue. Normally, one would raise the possibility of residual confounding in the discussion, after the analysis has been completed. This is what was done in the two abstracts submitted from this analysis and in the first draft of the paper. In these it was mentioned "residual confounding by age is possible in our analysis". The reason residual confounding may have been a possibility was that having low GFR was associated with increasing age, having low SES is also associated with increasing age and in the analysis age has been grouped into ten year age bands. So within these ten year bands the effect of age may still be confounding the association between SES and CKD. It was only after the abstracts were submitted and the draft paper written that I began to think about how to investigate whether residual confounding was a problem and how it could be dealt with.

One of the ways of tackling the problem of residual confounding would be to classify age into narrower age bands, for example using five year age bands rather than ten year age bands. Another way would be to include age as a continuous variable rather than grouping it at all. I had chosen ten year age bands, rather than five years, to make the reporting of the results simpler (fewer categories to report) and also to avoid having some age categories with small numbers. I had no clinical reason for choosing ten year groups. I had chosen to group the ages rather than fit age as a continuous variable as the first option would allow greater flexibility in describing the relationship between CKD and age. Fitting age as a continuous variable specifies a particular relationship between age and presence of CKD – there is a linear relationship between the logit and age. Fitting a grouped variable age specifies

no particular relationship. If there was a linear relationship between the logit and age then age could be fitted as a continuous variable and residual confounding would not be an issue, with the added benefit of a more powerful test.

The method of deciding whether there is a linear relationship between the logit and a covariate is described by Hosmer and Lemeshow (2000). It involves fitting the logistic model with the covariate grouped into categories (with the smallest category as the references) and then plotting the estimated coefficients for these categories against the mid-point of the category (taking the coefficient for the smallest category to be zero). If the plotted points are approximately linear then there is a linear relationship between the logit and the covariate and so the covariate may be fitted as a continuous variable. This procedure was carried out for each of the surveys and the plots are shown for age grouped into 10 year age bands (figures 7, 8 and 9). The assumption of linearity appears reasonable for all surveys. Plots were also obtained using five year age bands rather than ten years bands. Also, separate plots were created for each ethnic group in the NHANES survey as the logistic regression modelling was carried out stratified by ethnic group (not shown). In all cases the linearity assumption appeared reasonable. Age was then fitted as a continuous variable.

Figure 7: Estimated coefficients for each age group plotted against mid-point of the age group to determine the functional form of age in the logistic model using NHANES data

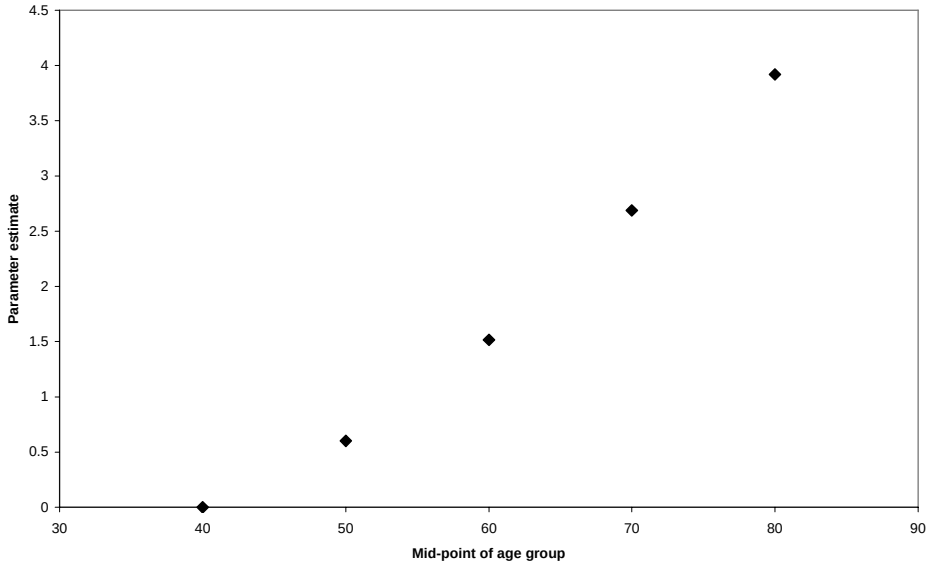


Figure 8: Estimated coefficients for each age group plotted against mid-point of the age group to determine the functional form of age in the logistic model using InterAsia data

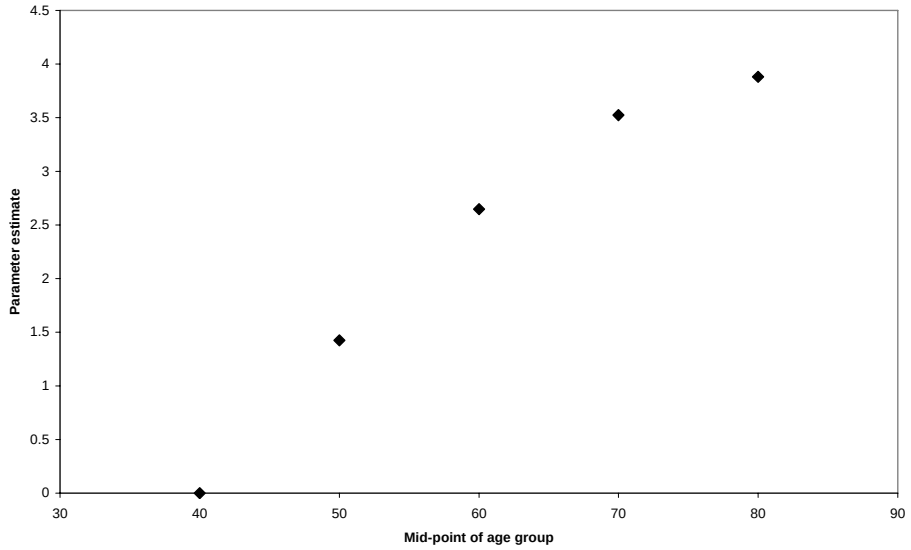


Figure 9: Estimated coefficients for each age group plotted against mid-point of the age group to determine the functional form of age in the logistic model using AusDiab data

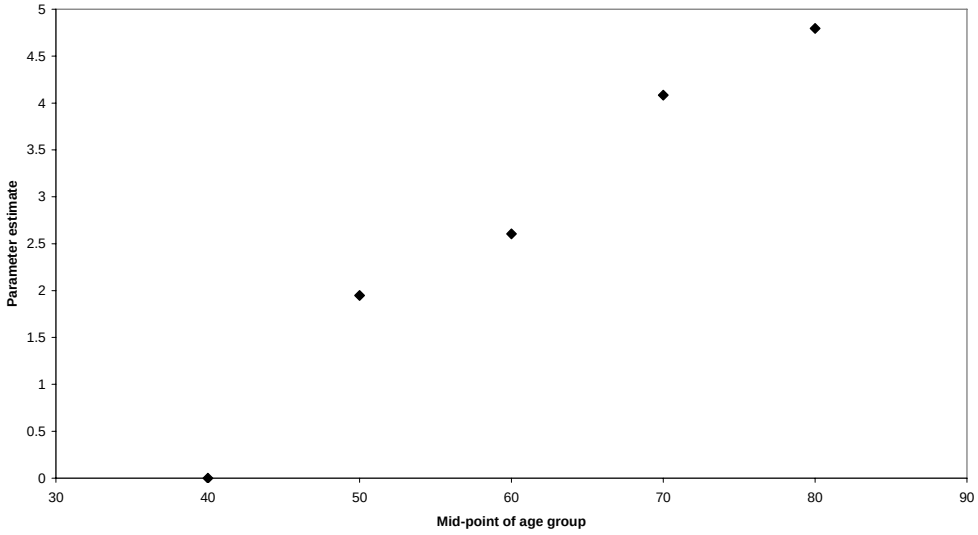


Table 3 shows the odds ratios for each of the SES factors when age is treated as continuous and when age is grouped into ten year age groups. From the tables there is evidence of residual confounding when age is analysed in groups as most of the estimated odds ratios are slightly closer to 1 when age is treated as a continuous variable. Although adjusting for age as a continuous variable had little effect on the overall results, it would be better to adjust for age as a continuous variable in the draft paper in order to pre-empt the issue of whether there may be residual confounding.

Table 3: Comparison of odds ratio calculated in models where age is fitted as a continuous or categorical variable

		Age fitted as continuous			*Age fitted as categorical*		
Survey	***Effect***	***Odds Ratio***	***95% Lower Confidence Limit***	***95% Upper Confidence Limit***	***Odds Ratio***	***95% Lower Confidence Limit***	***95% Upper Confidence Limit***
NHANES	Education: Average vs Greater than average	1.01	0.79	1.30	1.02	0.79	1.31
	Education: less than average vs Greater than average	1.31	1.00	1.71	1.33	1.02	1.73
	Income: Below median income vs Above median income	1.31	0.97	1.76	1.33	0.98	1.80
	Employment status: Not employed vs Employed	1.12	0.86	1.45	1.19	0.92	1.55
	Employment status: Retired vs Employed	1.52	1.13	2.03	1.61	1.19	2.17
InterASIA	Education: Average vs Greater than average	1.35	0.96	1.90	1.32	0.97	1.79
	Education: less than average vs Greater than average	1.17	0.73	1.88	1.25	0.82	1.90
	Income: Below median income vs Above median income	1.07	0.80	1.42	1.06	0.80	1.40
	Employment status: Not employed vs Employed	1.25	0.92	1.70	1.31	0.99	1.75

		Age fitted as continuous			*Age fitted as categorical*		
Survey	***Effect***	***Odds Ratio***	***95% Lower Confidence Limit***	***95% Upper Confidence Limit***	***Odds Ratio***	***95% Lower Confidence Limit***	***95% Upper Confidence Limit***
	Employment status: Retired vs Employed	1.11	0.80	1.53	1.27	0.94	1.70
AusDiab	Education: Average vs Greater than average	0.70	0.52	0.95	0.69	0.52	0.92
	Education: less than average vs Greater than average	1.07	0.88	1.30	1.13	0.93	1.38
	Income: Below median income vs Above median income	1.29	0.91	1.83	1.35	0.98	1.85
	Employment status: Not employed vs Employed	1.40	0.96	2.05	1.55	1.12	2.16
	Employment status: Retired vs Employed	2.01	1.28	3.17	2.30	1.54	3.45

Coding of the SES factors

Education

One of the socio-economic factors that was of interest was educational status. In AusDiab there were five different questions asked about different aspects of educational status, InterAsia and NHANES included only one – "What is the highest grade or year of regular school completed?". This question was not asked in AusDiab and the closest question matching this was "What is your highest level of education completed?" where the Reponses were categorical rather than numerical as in NHANES.

In Australia and the USA completion of high school requires 12 years of education and this is the norm for each country. It was decided to code educational status into three categories

- less than 12 years,

- 12 years,
- more than 12 years of schooling

Table 4 presents the percentage of people in each of the three categories for each of the three surveys. The split between each category for NHANES and AusDiab is very roughly one third, but for the InterAsia survey the percentage in the <12 years category is very high. Examining the InterAsia data it was found the distribution of number of years of schooling was quite different to that of NHANES and AusDiab. This reflects the quite different development, and current status, of the education systems in Thailand compared to Australia and the US.

Table 4: Number of years of schooling for participants in the three surveys

	Percentage		
	Survey		
	NHANES	***InterASIA***	***AusDiab***
Number of years of schooling			
Missing	0.5	0.0	0.1
< 12 years	26.9	89.3	42.7
12 years	32.3	4.3	18.4
> 12 years	40.3	6.4	38.8

The most frequently reported number of years of schooling was 4 years in the InterAsia survey compared to 12 in AusDiab and NHANES. Sixty-four percent of the InterAsia study reported number of years of schooling as 4.

The present coding of educational status may present a problem when interpreting the results for InterAsia. Being able to say that people would be at a decreased risk of kidney disease if they had 12 or more years of schooling would not be useful for Thailand where only around 10% of people in this survey had 12 or mores years of schooling. The census conducted in Thailand in 2000 reported average years of schooling to be 7.2 years (National Statistical Office Thailand, 2006). An alternative coding scheme would be to categorise numbers of years of schooling for each of the surveys into

- less than average,
- average,
- more than average

where "average" would be that which is considered the "norm" for the country. In Australia and the USA the average was taken to be 12 years of schooling, and in Thailand using the InterAsia data it was taken to be 4 years of schooling. This resulted in the following groupings –

Table 5: Number of years of schooling (recoded) for participants in the three surveys

	Percentage		
	Survey		
	NHANES	***InterASIA***	***AusDiab***
Number of years of schooling			
Missing	0.5	0.0	0.1
Less than average	26.9	11.5	42.7
Average	32.3	63.7	18.4
Greater than average	40.3	24.8	38.8

Two of my co-authors suggested that it would be useful to model the absolute effect of education status across the surveys rather than the relative effect within each survey as was chosen. One problem with this proposal is that it would require some way of equating one year of schooling across the three surveys. To estimate the absolute effect of 1 (or ten years) or schooling across the surveys would require a method for converting one year of Thai or Australian education into one year of US education. This would not be feasible and would not be investigated further.

Total household income

The SES factor total household income was recorded in a variety of ways in the three surveys. In the AusDiab study income was recorded as one of seven categories, NHANES had 28 categories and

InterAsia recorded the actual income. This means that how income was analysed would be constrained by the seven categories available in the AusDiab study if we are to make the results comparable.

It was decided that the simplest approach would be to assign the respondents in each survey to above and below the median household income for their survey. Consideration was given as to whether the median should be derived from official sources (eg census). However as in this analysis the population in each survey has been limited to those aged 35 and over the median from the census which would include all ages may not be applicable. As each of the three surveys was nationally representative the median calculated from within the survey should provide a reasonable estimate of the median household income for the population 35 and over.

After this decision was made to code income into above or below the median income, as derived from each survey, some of the authors of the paper suggested that we might be losing important detail not splitting income into smaller groups and also by using a relative measure of income rather than an absolute one. In response to this suggestion I sent the following reply –

We are constrained by how we can analyse income by the availability of data. This means if we are to use the same groupings, and hence, absolute measures we are constrained by the coarseness of recording in the AusDiab study.

In order to make the incomes comparable across the three surveys we also have to convert to the common currency of international dollars. This involved using the World Development Indicators conversion factors for 2002 (World Bank, 2004) to convert each country's currency to international dollars and increasing the US figure by a factor of 1.264 to account for inflation (the inflation from 1991 to 2000 in the US - (U.S. Department of Labor, Bureau of Labor Statistics, 2006) as the NHANES survey was carried out in 1988-94 whereas AusDiab was conducted in 1999-2000 and InterAsia 2000-2001.

Table 6 converts the seven AusDiab categories to international dollars and show the percentage of people from each survey in each category. Using these would result in a loss of detail in the InterAsia and NHANES studies as AusDiab studies have more responses in the middle, whereas InterAsia and NHANES are skewed to the low and high incomes, respectively.

Table 6: Percentage of population by household income (International$)

Total household income (International $)	NHANES	InterASIA	AusDiab
<2,972	0.6%	31.4%	0.5%
2,972 – 7,423	2.0%	36.8%	8.6%
7,423 – 14,858	10.0%	18.7%	17.6%
14,859 – 22,286	11.1%	5.6%	16.9%
22,287 – 29,715	6.7%	3.3%	12.0%
29,716 – 55,715	31.3%	2.9%	26.7%
≥55,716	38.4%	1.3%	17.7%

An alternative approach would be to split up the data into quartiles for each survey, as in table 7 (this is also equivalent to splitting up the surveys into quartiles using the original incomes rather than income in international dollars).

Table 7: Percentage of population by household income (International$) - quartiles

	Total household income (International $)	Percent of population
InterAsia	<2,500	24
	2,500 - 3,999	20
	4,000 - 7,999	29
	=>8,000	27
AusDiab	< 15,000	27
	15,000 - 29,999	29
	30,000 - 59,999	27
	=> 60,000	18
NHANES	< 22,000	24
	22,000 - 43,999	24
	44,000 – 59,999	21
	=> 60,000	32

This new categorization of income was then analysed and the age/gender adjusted odds ratios are plotted in figure 11. After adjusting for age and gender there is no apparent effect of income. This is consistent with the result when income is categorized into below or above median income.

Figure 10: Estimated adjusted odds ratios for household income (Internaltional $)

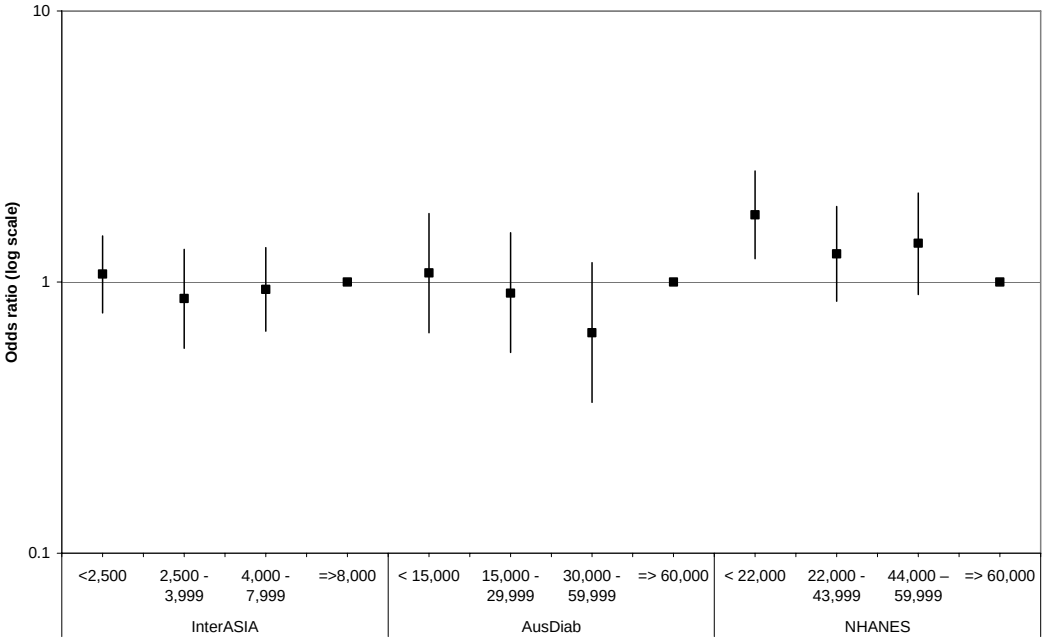
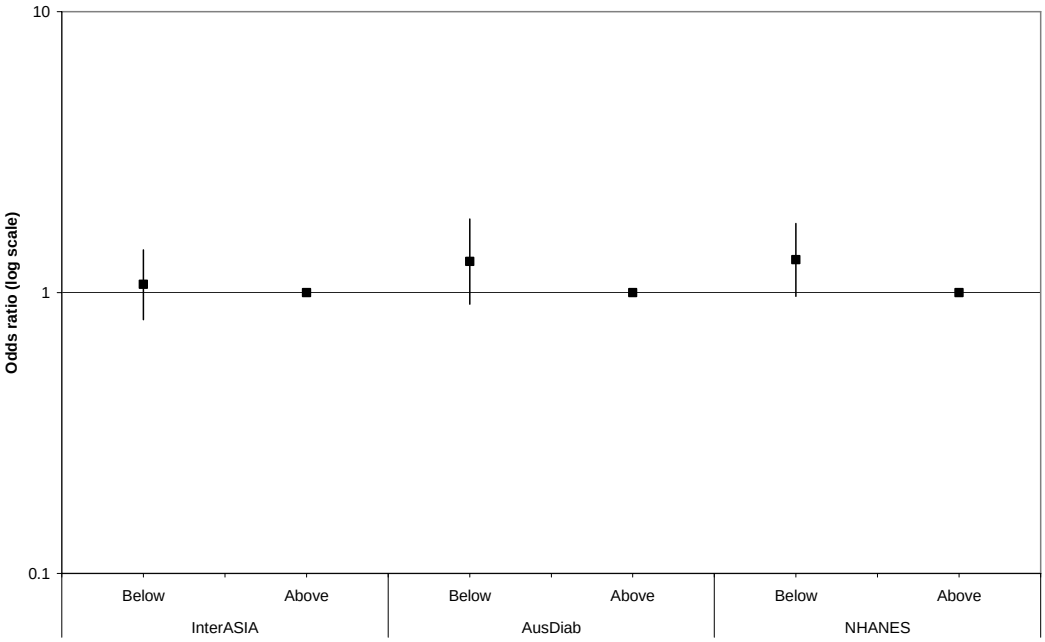


Figure 11: Estimated adjusted odds ratios for household income dichotomised into above or below median income



Employment

Each of the surveys recorded information on the subject's main occupation and they each recorded this to differing levels of detail. In InterAsia occupation was recorded as one of five categories, NHANES had 40 categories and the occupations recorded in AusDiab were coded to the ASCO (Australian Standard Classification of Occupations) classification which has 986 individual occupation categories (Australian Bureau of Statistics, 2997) which can be combined into nine major groups.

As the lowest number of categories was in the InterAsia study the other studies' occupation groups would in some way have to be mapped to these five categories. However, it was suggested that should we wish to examine the effect of occupation in more detail it would be useful to analyse the more detailed occupation categories in NHANES and AusDiab studies. In order to do this a recoding occupation for both studies to a common classification system would be required, and preferably to a system with international recognition as the surveys were carried out in separate countries. None of the authors of the paper had any specific knowledge regarding the classification of occupation so I sought the advice of Dr Tim Driscoll an expert on the analysis of occupational data.

Dr Driscoll suggested I look into the use of ISCO 88 (the International Standard Classification of Occupations created by the International Labor Office (1990)) and also provided advice on how the current data could be mapped to this classification. The ISCO classification scheme turned out to be very similar to that of ASCO. In ISCO 88 there are 10 major groups which correspond broadly to the nine major groups in ASCO. The extra group in ISCO 88 is for skilled agricultural and fishery workers, whereas in ASCO these occupations are split between two major groups – farm managers are assigned to Group 1 Managers and administrators, and skilled agricultural workers are assigned to Group 4 Tradespersons and related workers.

Although the ISCO classification is more internationally recognised it was decided to use the ASCO classification scheme to code both the AusDiab and NHANES studies. The reason for this was the purpose of the occupation classification is to provide a measure of socioeconomic status and ASCO provided a better reflection of SES as evidenced by the classification of farm workers. Also, the ASCO categories were broadly similar to that of ISCO 88 and the AusDiab data were already coded to ASCO so less work would be involved in coding these data to another scheme.

Table 8 provides the percentage of people in each major ASCO group for the AusDiab and NHANES surveys.

Table 8: Percentage in ASCO group for NHANES and AusDiab

	NHANES	AusDiab
ASCO group		
Missing	0.1	1.1
Managers and administrators	10.0	5.9
Professionals	9.1	14.7
Associate professionals	5.8	8.3
Tradespersons and related workers	7.5	6.3
Advanced clerical and service workers	2.0	3.1
Intermediate clerical, sales and service workers	13.8	7.9
Intermediate production and transport workers	5.8	3.2
Elementary clerical, sales and service workers	0.0	3.3
Labourers and related workers	5.7	3.2
Not employed	19.6	16.7
Retired	20.5	26.3

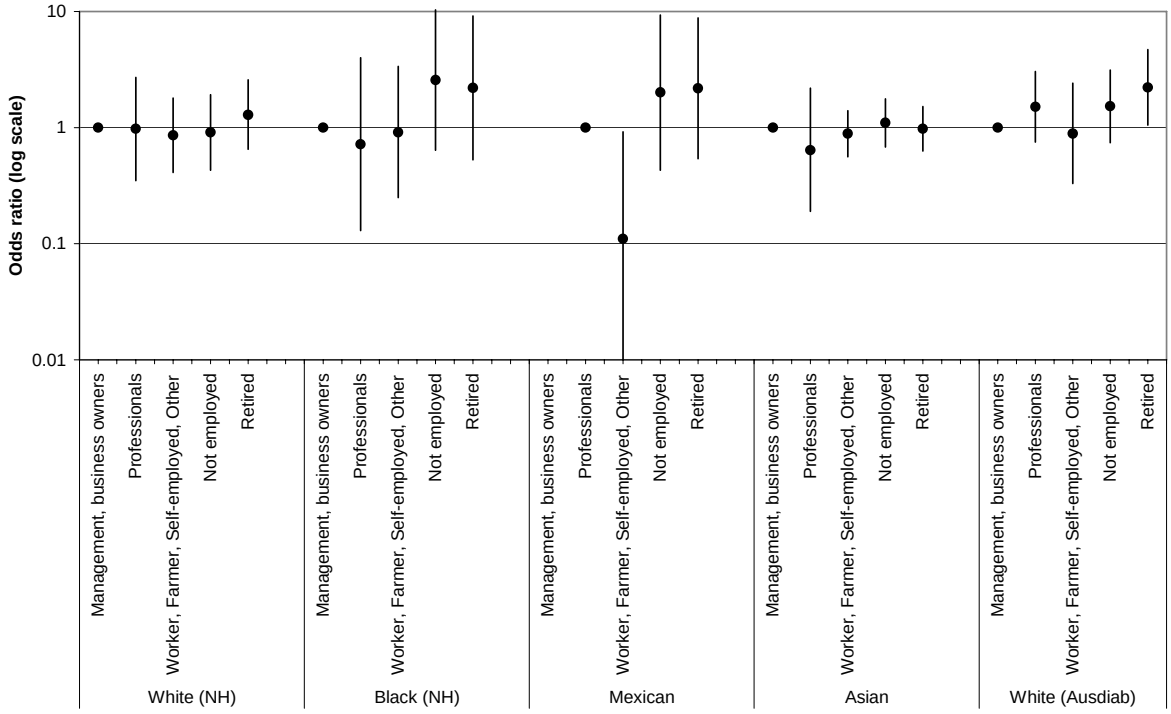
The nine ASCO groups were then mapped to the five occupation categories that were used in the InterAsia survey. However, two of the InterAsia categories – self-employed and business owner - had no direct equivalent in the ASCO classification. Therefore, some of the InterAsia categories were merged and the following classification scheme was created to be used in the analysis.

Table 9: Percentage in occupation groups

	NHANES	InterASia	AusDiab	Total
Employment categories (Thai grouped)				
Missing	0.1	0.3	1.6	0.3
Management, business owners	10.0	3.8	5.9	8.7
Professionals	14.9	4.3	23.0	13.6
Worker, Farmer, Self-employed, Other	34.8	68.9	27.0	40.2
Not employed	19.6	16.5	16.2	18.8
Retired	20.5	6.2	26.3	18.4

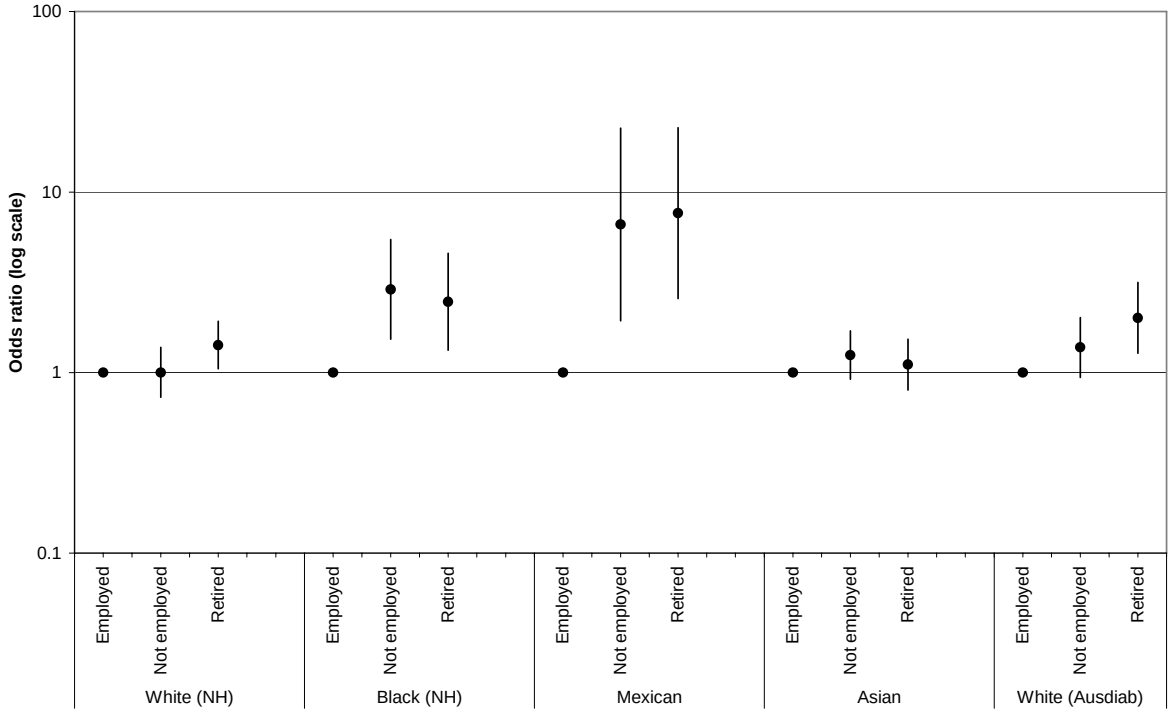
The odds ratios showing the association between employment and CKD, adjusted for age and gender, are shown in figure 12. Note that the reference employment group for all ethnic groups is Management and business owners, except for Mexican-Americans where the reference category is Professionals. There were no Mexican-Americans in the Management and business owners group who had CKD so it was not possible to use this as the reference category. Management and business owners has been kept as the reference category for the other ethnic groups for consistency with the analyses of the other SES factors where the highest SES category has been chosen as the reference category.

Figure 12: Estimated adjusted odds ratios for employment category by ethnic group



After analysing the data using these categories it was then decided that in order to cut down on the number of data points reported in the submitted paper it would be better to further group employment categories. Three categories were decided upon – Employed, Unemployed and Retired and the effect on the age and gender adjusted results is shown in figure 13. The pattern of results is broadly similar with the major change being the narrowing of confidence intervals due to the combining of smaller categories.

Figure 13: Estimated adjusted odds ratios for combined employment categories by ethnic group



Conclusion

The previous analyses have set out the many issues that were considered in examining the relationship between socio-economic status and chronic kidney disease. Much of this work will not be included in any formal summary of the work for publication. The draft paper which follows reports in detail only the associations between the SES factors and CKD and gives brief mention to the choice of method for estimating GFR and the implication of residual confounding. Although it can be seen from the previous analysis that the work involved in substantiating these brief mentions is substantial. Also, there was also a lot of discussion and work involved in determining an appropriate form for the SES factors. This work, although for the moment not required in the draft for publication, will be invaluable later, if and when, we are called upon by reviewers to justify the choices we have made in this study.

References

Australian Bureau of Statistics (1997). Australian Standard Classification of Occupations (ASCO) Second Edition, 1997. Available at http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/1220.0Contents11997?opendocument&tabname=Summary&prodno=1220.0&issue=1997&num=&view= Last accessed 20 June 2006.

Armitage P, Berry G, Matthews JNS. (2002) Statistical Methods in Medical Research. Oxford: Blackwell Science.

The Australasian Creatinine Consensus Working Group (2005). Chronic kidney disease and automatic reporting of estimated glomerular filtration rate: a position statement. MJA 2005; 183 (3): 138-141

Bland, J.M. and Altman D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet; February 8:307:310.

Chadban SJ, Briganti EM, Kerr PG, Dunstan DW, Welborn TA, Zimmet PZ, Atkins RC (2003). Prevalence of kidney damage in Australian adults: The AusDiab kidney study. Journal of the American Society of Nephrology 14: S131-S138.

Dunstan DW, Zimmet PZ, Welborn TA, Cameron AJ, Shaw J, de Courten M, Jolley D, McCarty DJ, AusDiab Steering Committee. (2002) The Australian Diabetes, Obesity and Lifestyle Study (AusDiab) – methods and response rates. Diabetes Research and Clinical Practice 57 (2002) 119-129.

He J, Neal B, Gu D, Suriyawongpaisal P, Xin X, Reynolds R, MacMahon S, Whelton PK. (2004) International collaborative study of cardiovascular disease in Asia: Design, rationale and preliminary results. Ethnicity and Disease , Vol 14, Spring 2004, 260-268.

Hosmer DW, Lemeshow S (2000). Applied logistic regression. New York: Wiley

International Labor Office (1990). International standard classification of occupations : ISCO-88. Geneva : International Labour Office.

Luan X, Pan W, Gerberich SG, Carlin BP. (2005) Does it always help to adjust for misclassification of a binary outcome in logistic regression? Statistics in Medicine 2005 Jul 30;24(14):2221-34.

National Center for Health Statistics Centers for Disease Control and Prevention (1996) Analytic and Reporting Guidelines: The Third National Health and Nutrition Examination Survey, NHANES III (1988-94). Available at http://www.cdc.gov/nchs/data/nhanes/nhanes3/nh3gui.pdf, last accessed 19 June 2006.

National Statistical Office Thailand (2006). Education indicators http://web.nso.go.th/eng/indicators/educat_e.htm - Last accessed 19 June 2006.

Schneider B . (2001). Using the SAS System to visualise inter-rater agreement for continuous measurements in medical statistics. SAS User Group International conference. Downloaded from http://www2.sas.com/proceedings/sugi25/25/po/25p207.pdf Last accessed 19 June 2006.

U.S. Department of Labor, Bureau of Labor Statistics, 2006. CPI Inflation Calculator. Available at http://data.bls.gov/cgi-bin/cpicalc.pl. Last accessed 19 June 2006.

Webb P, Bain C, Pirozzo S. (2005). Essential epidemiology : an introduction for students and health professionals. New York: Cambridge University Press

World Bank. 2004. World Development Indicators 2004. Washington, D.C. : International Bank for Reconstruction and Development.

First draft of paper for submission to journal

The Relationship Between Chronic Kidney Disease and Individual-Level Socio-Economic Status: The Three Continent Kidney Disease (3CKD) Study

White S [1,2], McGeechan K [3], Jones M [3], Cass A [2], Chadban S [1,4], K Polkinghorne [5], Perkovic V [2], Roderick PJ [6]

[1] Central Clinical School, The University of Sydney, Sydney, Australia

[2] The George Institute for International Health, Sydney, Australia

[3] School of Public Health, The University of Sydney, Sydney, Australia

[4] Royal Prince Alfred Hospital, Sydney, Australia

[5] Monash Medical Centre, Melbourne, Victoria

[6] University of Southampton, Southampton, United Kingdom

Abstract

The relationship between socio-economic status (SES) and chronic kidney disease (CKD) was examined using nationally-representative data from 3 countries (USA, Thailand and Australia). Individual data were obtained for participants aged 35 years or above from the NHANES III, InterASIA and AusDiab I studies. We examined the association between CKD prevalence (GFR <60 ml/min/1.73 m^2) and education, income, employment and area of residence (urban/rural). Household income below national median, below-average education, being unemployed/retired and living in rural settings were associated with increased odds of CKD (INSERT RESULTS). Lower SES was consistently associated with higher rates of CKD across different ethnic groups, however some variation in effect size was observed (INSERT RESULTS). Income, employment and education are associated with CKD risk, however the effects vary between different population groups and this requires further exploration. Our results suggest CKD prevention and management strategies should take account of the higher prevalence in low-income and poorly educated groups.

Introduction

The results of cross-sectional and prospective studies have established a relationship between socioeconomic status and health indicators, including physical and mental functioning, specific diseases and mortality (Pappas, 1993; Lynch, 1996; Lynch, 1997; Kunst, 1998). This relationship takes the form of a gradient, such that increasing levels of wealth are associated with increasingly better health (Adler, 1994), and cannot be fully explained by differences in burden of risk factors such as smoking, obesity, elevated cholesterol or blood pressure (Davey Smith, 1990). Further investigation of this association found specific links between socioeconomic status and cardiovascular morbidity and mortality (Lynch , 1996; Marmot, 1978; Von Rossum, 2000). Since chronic kidney disease is part of the spectrum of chronic vascular diseases, it is possible that similar mechanisms behind the observed association between SES and cardiovascular disease underlie a relationship between SES and chronic kidney disease (CKD). There is evidence from several studies to support an association between SES and CKD. Most have examined the link between individual- or area-level SES and treated end-stage kidney disease (ESKD) (Young, 1994; Byrne, 1994; Cass, 2001). Studies conducted in pre-end-stage populations also support an association between SES and CKD. (Fored, 2005).

This analysis aims to address questions concerning the association between individual-level socio-economic factors and CKD. The two principle questions which are the subject of this analysis are: i) is socio-economic status (SES) associated with chronic kidney disease (CKD), and; ii) is the association between SES and CKD different in different ethnic groups? We address these research questions in parallel across three countries, the United States, Thailand and Australia, using data from three independent cross-sectional population-based surveys, NHANES III (USA), InterASIA (Thailand) and AUSDIAB I (Australia).

Understanding the relationship between socioeconomic factors and early stages of chronic kidney disease is of particular importance given recent recognition of the effective impact primary prevention and early detection can have on the development and progression of chronic kidney disease. Local and international health promotion, education, screening and early intervention efforts will be improved through detailed understanding of which groups are at particular risk and what barriers exist to their detection and effective treatment.

Methods:

Study Design and Population

Individual participant data were obtained for participants aged 35 years or above from the NHANES III (n=10,625), InterASIA (Thailand; n=5,099) and AusDiab (n=9,852) studies who had a valid serum creatinine measurement. The minimum age cut-off for inclusion in the analysis was set at 35, as the inclusion criteria for the InterASIA study stipulated age of 35 years or above. The rate of CKD in the population under 35 is very low. There were 9,098 participant in NHANES, 5,063 in InterASIA and 9,329 in AusDiab who had valid measurements on all the variables reported. As one of the principle questions of this study was whether the association between SES and CKD differs between ethnic groups, the ethnic composition of the three study populations was considered. The NHANES III study over samples African Americans and Mexican Hispanics in order to look specifically at differential health outcomes in these two groups (**insert Coresh Ref**). In our analysis we therefore looked at white, African, and Mexican Americans separately. Although Indigenous Australians suffer a disproportionately high burden of chronic and end-stage kidney disease, only 88 people in the AusDiab study were classified as Aboriginal and Torres Strait Islanders and therefore this study was not powered to look at differential outcomes in this group. Due to small numbers, the Indigenous Australian participants of the AusDiab I study and the 423 participants in the 'Other' ethnic group from NHANES III were not included in the logistic regression analysis to evaluate the association between CKD and SES amongst the different ethnic groups. No data on ethnicity was collected as part of the InterAsia study.

(WPP: see *Survey Design* and *Methods of analysis sections* for a description of the surveys and a discussion on why adjustment for the survey design is necessary).

Outcome and Exposure Variables

The outcome of interest for this study was cross-sectional prevalence of CKD. Presence of CKD was defined as glomerular filtration rate (GFR) of less than 60 ml/min/1.73m^2. or Stage 3 CKD according to the K/DOQI classification of chronic kidney disease (***insert ref***). GFR was estimated using the Modification of Diet in Renal Disease (MDRD) study prediction formula:

$$\textit{Estimated GFR (ml/min/1.73 m}^2\textit{)} = 186.3 \times (\textit{serum creatinine}^{-1.154}) \times (\textit{age}^{-0.203})$$

*(0.742 if female), and *(1.21 if African-American)

Serum creatinine results for NHANES III and InterAsia had both been standardized to the same central laboratory, whereas serum creatinine results from AusDiab I had not been standardized. Data on urinary protein excretion was available only for NHANES III and AusDiab I, and was not collected as part of the InterAsia study.

(WPP: see the section *Comparison of two formulae to estimate GFR* for a discussion on using the alternative Cockcroft Gault formula to estimate GFR).

The socio-economic variables for which data was available from all 3 studies were education, household income, area of residence (urban/rural) and employment level. Data on education level was grouped into 'less than average', 'average' and 'greater than average' to reflect relative education standards in the United States (average=12 years), Australia (average=12 years) and Thailand (average=4 years). Data for total household income was similarly grouped into below median and above median (US: median= USD 30,000; Australia: median= AUD 41,600; Thailand: median= TBHT 60,000). Employment status was coded as 'employed, 'unemployed' or 'retired'.
(WPP: see the sections in *Coding of the SES factors* for discussions of the coding of education, income and employment).

Analysis

The demographic, health-related and socioeconomic descriptive statistics for each of the three studies were calculated, and weighted to represent the source populations. Univariate associations were calculated between each of the SES variables (education, household income, employment and rural/urban residence) and CKD (GFR <60 ml/min/1.73m^2). The univariate associations were estimated for each study (combined within country) and then stratified by ethnic group.

The association between SES and CKD may be confounded or mediated by a number of other variables. Logistic regression was used to determine whether any association between SES and CKD remains after adjusting for such variables. Variables considered included age, sex, hypertension, history of cardiovascular events (heart attack or stroke), diabetes, smoking status, obesity and cholesterol (total, LDL, HDL). Obesity was defined according to waist circumference, with the cut-offs of $\geq$ 80 cm if female, $\geq$ 94cm if male white or African American, $\geq$ 90cm if male Asian or Mexican-American (ref for waist circumference). Hypertension was defined as measured Systolic

blood pressure ≥140 mmHg and diastolic blood pressure ≥ 90 mmHg, if the participant is currently taking anti-hypertensive medication, or if a doctor has ever told the participant they had high blood pressure or hypertension. Diabetes was defined as plasma glucose level ≥ 7mmol/L, if the participant is taking insulin or oral glucose agents, or if a doctor has ever told participant they had diabetes, including gestational diabetes. The multivariate adjusted association between SES and CKD was estimated combined within country, and then separately by ethnic group.

The analysis used logistic regression to calculate the odds ratio for each of the risk factors. The standard errors were adjusted for the sampling design for each survey using the appropriate clustering, stratification and sampling weights. The analysis was carried out in SAS v9.1 using PROC SURVEYLOGISTIC.

Finally individual study estimates were pooled using a random effects meta-analysis model to give an overall estimate of the effect of SES on rates of GFR <60. Intercooled Stata 8.2 was used to implement the META macro to calculate the pooled odds ratios.

(WPP: see the section see *Survey Design* and *Methods of analysis* sections for description of the design of each survey, why adjustment for the survey design is necessary and the reasons for pooling the results using a random effects model).

Results:

Characteristics of the Study Population

TABLE 1: Demographic and SES characteristics of study participants

	%			
	Survey			
	NHANES	*InterASIA*	*AusDiab*	***Total***
Age group				
35-44	34.4	40.8	31.4	35.3
45-54	21.8	26.6	26.6	22.9
55-64	18.3	17.2	17.4	18.1
65-74	15.7	12.7	16.2	15.2
75+	9.7	2.6	8.5	8.5
Gender				
Male	46.5	48.7	48.6	47.0
Female	53.5	51.3	51.4	53.0
Ethnic group				
White (NHANES)	79.8	0.0	0.0	61.6
Black (NHANES)	9.3	0.0	0.0	7.2
Mexican-Amer	3.8	0.0	0.0	2.9
Other (NHANES)	7.1	0.0	0.0	5.5
Asian	0.0	100.0	0.0	16.4
White (AUSDIAB)	0.0	0.0	99.3	6.4
ATSI (AUSDIAB)	0.0	0.0	0.7	0.0
Area of Residence				
Urban	48.2	32.3	58.8	46.3
Rural	51.8	67.7	41.2	53.7

	%			
	Survey			
	NHANES	*InterASIA*	*AusDiab*	***Total***
Number of years of schooling				
Less than average	27.1	11.9	42.8	25.6
Average	32.4	63.1	18.4	36.6
Greater than average	40.5	25.1	38.8	37.9
Total household income				
Below median income	47.5	46.5	55.7	47.9
Above median income	52.5	53.5	44.3	52.1
Employment status				
Employed	59.9	77.1	57.0	62.5
Not employed	19.6	16.4	16.4	18.8
Retired	20.5	6.5	26.6	18.6

Demographic and socioeconomic characteristics of the NHANES III, InterASIA and AusDiab populations are given in Table 1. After restricting the 3 study populations to participants aged 35 years and above, the population of the InterASIA survey was younger than that of NHANES III and AusDiab. Table 2 summarises the relevant health characteristics of the study populations. The prevalence of CKD (GFR <60 ml/min/1.73m^2) was 6.6% in NHANES III, 13.9% in InterAsia and 10.0% in AusDiab. The InterASIA population had notably lower rates of hypertension and cardiovascular problems and NHANES had higher rates of obesity. After stratifying the NHANES population by ethnicity (*data not shown*), African- and Mexican-Americans were younger and had lower socio-economic status than Whites. In terms of health status, Whites had consistently better health than African-Americans. However, Mexican-Americans fared worst in terms of diabetes and obesity measures but best in terms of smoking, hypertension and cardiovascular measures. The prevalence of CKD in the NHANES III population stratified for ethnicity was 7.2% among Whites, 5.6% among African-Americans and 2.1% among Mexican-Americans.

Table 2: Health status characteristics of study participants

	% Survey			
	NHANES	*InterASIA*	*AusDiab*	***Total***
CKD				
=>60	93.4	86.1	90.0	92.0
<60	6.6	13.9	10.0	8.0
Smoking status				
Current smoker	24.3	24.8	14.6	23.8
Ex-smoker	32.7	13.9	29.3	29.4
Never smoked	42.9	61.3	56.1	46.8
Hypertension				
Yes	41.2	24.3	43.2	38.6
No	58.8	75.7	56.8	61.4
Stroke or heart attack				
Yes	7.3	0.8	6.4	6.2
No	92.7	99.2	93.6	93.8
Diabetes				
Yes	10.6	9.8	7.4	10.3
No	89.4	90.4	92.6	89.7
Obese (using waist circumference)				
Yes	69.4	37.8	32.0	61.6
No	30.6	62.2	68.0	38.4

Results of Univariate Analysis

Results of univariate analysis are shown in Table 3. Within each ethnic group all factors were associated with presence of CKD ($p<0.05$) with the exception of gender, area of residence, obesity amongst African and Mexican Americans and education amongst Mexican Americans. All univariate associations were in the expected direction except for smoking status – people who had never smoked or were ex-smokers were more likely to have CKD than current smokers. However, this association is due to current smokers being younger than non- or ex-smokers across all surveys and the association becomes non-significant after adjusting for age.

Table 3: Crude odds ratios (95% confidence interval) for the association between CKD prevalence and possible risk factors

		Ethnic Group				
		White (NHANES III)	African American	Mexican American	Asian	White (AusDiab II)
Income						
	Above median	1.00	1.00	1.00	1.00	1.00
	Below median	3.45 (2.54, 4.69)	3.46 (1.92, 6.24)	2.12 (1.13, 3.95)	1.50 (1.08, 2.08)	6.03 (4.43, 8.20)
Education						
	Greater than average	1.00	1.00	1.00	1.00	1.00
	Average	1.29 (0.99, 1.68)	1.79 (1.04, 3.08)	2.43 (0.63, 9.29)	2.55 (1.75, 3.72)	0.97 (0.76, 1.23)
	Less than average	3.18 (2.41, 4.20)	4.97 (2.94, 8.38)	2.37 (0.73, 7.68)	5.90 (3.79, 9.19)	2.69 (2.20, 3.28)
Area of Residence	*Urban*	1.00	1.00	1.00	1.00	1.00
	Rural	1.12 (0.86, 1.46)	1.36 (0.96, 1.93)	1.37 (0.64, 2.91)	1.31 (0.76, 2.25)	1.43 (0.86, 2.38)
Employment status	*Employed*	1.00	1.00	1.00	1.00	1.00
	Not employed	3.54 (2.52, 4.98)	4.72 (2.73, 8.18)	10.12 (3.66, 27.99)	3.38 (2.46, 4.64)	4.51 (3.02, 6.72)

	Retired	10.01 (7.55, 13.27)	11.10 (6.88, 17.91)	40.72 (16.23, 102.13)	5.74 (3.94, 8.36)	14.34 (10.52, 19.54)
Age	*35-44*	1.00	1.00	1.00	1.00	1.00
	45-54	1.50 (0.56, 4.02)	1.96 (0.79, 4.82)	-	4.14 (2.47, 6.94)	6.85 (4.86, 12.15)
	55-64	4.02 (1.77, 9.14)	8.07 (3.81, 17.09)	8.10 (2.81, 23.38)	14.11 (8.35, 23.83)	13.26 (7.49, 23.46)
	65-74	12.05 (5.75, 25.22)	18.70 (9.64, 36.27)	17.41 (6.71, 45.20)	34.08 (19.48, 59.62)	59.66 (34.27, 103.86)
	≥ 75	46.01 (22.16, 95.51)	38.51 19.38, 76.52)	41.93 (16.79, 104.73)	48.95 (24.84, 96.46)	121.15 (69.83, 210.20)
Sex	*Male*	1.00	1.00	1.00	1.00	1.00
	Female	1.58 (1.20, 2.05)	1.31 (0.90, 1.92)	0.99 (0.57, 1.71)	1.34 (0.87, 2.06)	2.32 (1.75, 3.07)
Smoking status	*Current smoker*	1.00	1.00	1.00	1.00	1.00
	Ex-smoker	2.97 (2.11, 4.18)	2.34 (1.48, 3.72)	3.39 1.31, 8.75)	1.85 (1.29, 2.66)	2.65 (1.71, 4.11)
	Never smoked	2.57 (1.83, 3.61)	1.83 (1.16, 2.88)	1.98 (0.96, 4.08)	1.68 1.05, 2.70)	2.51 (1.62, 3.90)
Hypertension	*No*	1.00	1.00	1.00	1.00	1.00
	Yes	4.85 (3.79, 6.20)	7.94 (5.35, 11.79)	6.09 (3.21, 11.58)	2.74 (2.05, 3.66)	3.62 (3.01,4.35)
Stroke or heart attack	*No*	1.00	1.00	1.00	1.00	1.00
	Yes	5.65 (4.42, 7.22)	5.38 (3.62, 7.99)	5.38 (2.90, 9.99)	3.16 (1.47, 6.80)	4.71 (3.43, 6.47)
Diabetes	*No*	1.00	1.00	1.00	1.00	1.00
	Yes	3.26 (2.30, 4.62)	4.71 (2.87, 7.73)	2.99 (2.33, 3.85)	1.97 (1.13, 3.41)	1.49 (1.16, 1.90)

Obese (waist circumference)	*No*	1.00	1.00	1.00	1.00	1.00
	Yes	2.06 (1.53, 2.79)	1.43 (0.91, 2.25)	1.48 (0.72, 3.06)	1.49 (0.72, 3.06)	2.27 (1.68, 3.07)

Income group, employment group, and level of education were all significantly associated with CKD ($p<0.01$). Having a household income below the national median was associated with a significantly increased rate of GFR <60 across all ethnic groups (Aus White OR=6.03; US White OR =3.45, African American OR=3.46, Mexican American OR=2.12, Asian OR=1.50). Similarly, decreasing levels of education were associated with higher rates of CKD for all ethnic groups except for Mexican Americans. Being unemployed or retired was associated with higher rates of CKD across all ethnic groups.

Adjusting for age, gender and other known risk factors for CKD

After adjusting for the age and gender of participants the odds ratios for each of the socioeconomic factors reduced. Amongst Asians none of the socio-economic factors remained associated with the prevalence of CKD ($p>0.1$). Employment status remained associated with CKD for each of the other ethnic groups ($p<0.01$), as did education ($p<0.05$), with the exception of amongst Mexican Americans. The odds ratios for income were greater than one for all the ethnic groups however their confidence intervals also contained one.

The effects of adjusting for other known risk factors (presence of diabetes, hypertension, CHD, obesity and smoking status) are shown in the last three columns of table 4. The impact on the odds ratios of each of these adjustments is much less than the adjustment for age and gender. As more factors are adjusted for the relationship between SES factors and CKD lessened. After adjusting for all confounders the only clear associations that remained were those for the effect of employment status for African and Mexican Americans and for the Whites in AusDiab.

Table 4: Odds ratios adjusted for other factors

		Crude OR	adjusted for age and sex	+ diabetes	+ hypertension	+ other confounders
Ethnic group	**Effect**	Model 1	Model 2	Model 3	Model 4	Model 5
White (NHANES)	Education: Average vs Greater than average	1.29 (0.99,1.68)	0.99 (0.75,1.29)	0.97 (0.74,1.27)	0.94 (0.72,1.24)	1.00 (0.76,1.30)
	Education: Less than average vs Greater than average	3.18 (2.41,4.20)	1.33 (1.01,1.76)	1.30 (0.98,1.71)	1.28 (0.97,1.70)	1.34 (1.01,1.77)
	Income: Below median income vs Above median income	1.79 (1.04,3.08)	1.63 (0.95,2.81)	1.58 (0.91,2.73)	1.48 (0.87,2.54)	1.31 (0.75,2.30)
	Employment: Not employed vs Employed	4.97 (2.94,8.38)	2.07 (1.17,3.68)	1.98 (1.08,3.63)	1.82 (1.00,3.34)	1.79 (0.92,3.47)
	Employment: Retired vs Employed	2.43 (0.63,9.29)	1.86 (0.47,7.31)	2.01 (0.49,8.31)	1.97 (0.48,8.05)	3.52 (0.86,14.44)
African-American	Education: Average vs Greater than average	2.37 (0.73,7.68)	0.86 (0.25,2.96)	0.79 (0.23,2.73)	0.79 (0.24,2.68)	1.40 (0.40,4.81)
	Education: Less than average vs Greater than average	2.55 (1.75,3.72)	1.35 (0.96,1.90)	1.38 (0.97,1.96)	1.43 (1.01,2.04)	1.31 (0.92,1.85)
	Income: Below median income vs Above median income	5.90 (3.79,9.19)	1.17 (0.73,1.88)	1.20 (0.75,1.92)	1.22 (0.77,1.93)	1.11 (0.70,1.76)
	Employment: Not employed vs Employed	0.97 (0.76,1.23)	0.70 (0.52,0.95)	0.70 (0.52,0.95)	0.70 (0.52,0.94)	0.72 (0.53,0.99)
	Employment: Retired vs Employed	2.69 (2.20,3.28)	1.07 (0.88,1.30)	1.07 (0.88,1.30)	1.07 (0.88,1.29)	1.01 (0.80,1.28)
Mexican-American	Education: Average vs Greater than average	3.45 (2.54,4.69)	1.28 (0.92,1.78)	1.25 (0.90,1.73)	1.24 (0.89,1.72)	1.29 (0.89,1.86)
	Education: Less than average vs Greater than average	3.46 (1.92,6.24)	1.79 (1.00,3.23)	1.67 (0.91,3.04)	1.53 (0.85,2.76)	1.65 (0.86,3.16)
	Income: Below median income vs Above median income	2.12 (1.13,3.95)	1.20 (0.66,2.18)	1.12 (0.61,2.07)	1.13 (0.63,2.04)	1.11 (0.55,2.24)
	Employment: Not employed vs Employed	1.50 (1.08,2.08)	1.07 (0.80,1.42)	1.07 (0.80,1.44)	1.11 (0.84,1.49)	1.06 (0.79,1.42)
	Employment: Retired vs Employed	6.03 (4.43,8.20)	1.30 (0.91,1.85)	1.30 (0.92,1.85)	1.27 (0.90,1.79)	1.21 (0.80,1.82)
Asian	Education: Average vs Greater than average	3.54 (2.52,4.98)	1.00 (0.73,1.38)	0.98 (0.71,1.34)	0.95 (0.69,1.30)	0.81 (0.57,1.16)
	Education: Less than average vs Greater than average	10.01 (7.55,13.27)	1.42 (1.05,1.92)	1.36 (1.02,1.82)	1.33 (0.99,1.77)	1.18 (0.86,1.61)
	Income: Below median income vs Above median income	4.72 (2.73,8.18)	2.89 (1.53,5.46)	2.73 (1.42,5.27)	2.47 (1.30,4.71)	2.38 (1.24,4.58)
	Employment: Not employed vs Employed	11.10 (6.88,17.91)	2.47 (1.33,4.58)	2.23 (1.18,4.22)	1.98 (1.06,3.68)	1.87 (0.97,3.58)

	Employment: Retired vs Employed	10.12 (3.66,27.99)	6.62 (1.94,22.64)	5.85 (1.64,20.81)	5.52 (1.51,20.15)	4.49 (1.15,17.55)
White (AUSDIAB)	Education: Average vs Greater than average	40.72 (16.23,102.1)	7.65 (2.57,22.76)	7.24 (2.37,22.16)	6.98 (2.21,22.02)	6.54 (1.93,22.13)
	Education: Less than average vs Greater than average	3.38 (2.46,4.64)	1.25 (0.92,1.70)	1.20 (0.88,1.64)	1.20 (0.88,1.63)	1.27 (0.92,1.77)
	Income: Below median income vs Above median income	5.74 (3.94,8.36)	1.11 (0.80,1.53)	1.03 (0.74,1.42)	0.96 (0.69,1.35)	1.00 (0.74,1.36)
	Employment: Not employed vs Employed	4.51 (3.02,6.72)	1.38 (0.94,2.02)	1.38 (0.93,2.05)	1.35 (0.91,1.99)	1.26 (0.88,1.83)
	Employment: Retired vs Employed	14.34 (10.52,19.54)	2.01 (1.28,3.16)	2.01 (1.28,3.15)	1.97 (1.27,3.05)	1.95 (1.26,3.01)

Pooled Results

The age and gender adjusted odds ratios for the three surveys were pooled using a random effects model and are displayed in the forest plot below**.** The pooled results show that lower SES as measured by each of the three factors is associated with an increased prevalence of CKD.

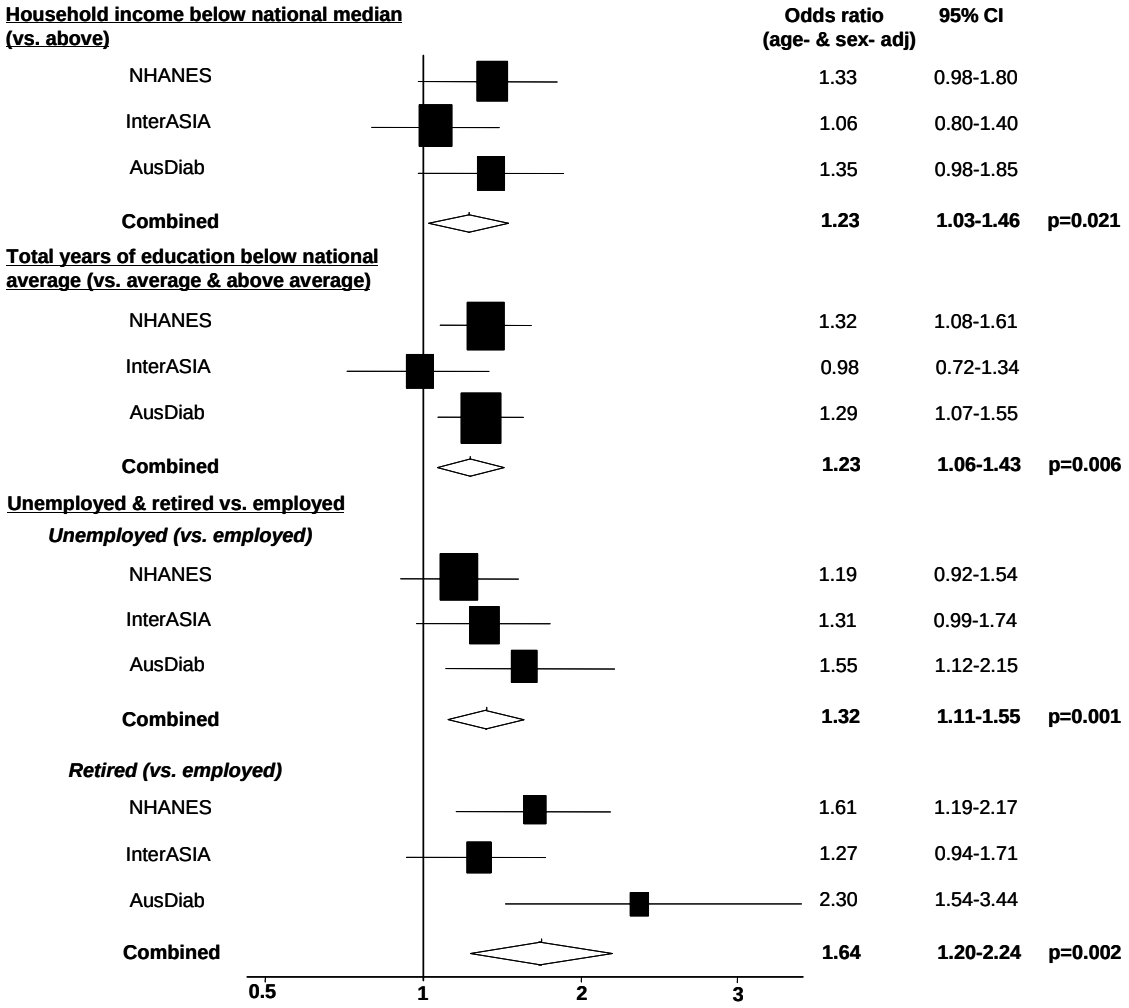


Figure 1: Forest plots of meta-analyses combining results from the three individual studies. Effect estimates are adjusted for age and sex. Pooled (combined) estimates are based on random-effects model.

Discussion

Our results are consistent with the findings from other studies of the relationship between SES and pre-end-stage CKD A Swedish case-control study of CKD (serum creatinine permanently in excess of 300umol/l in men and 250umol/l in women) in 18-74 year-olds found the risk of CKD was increased two-fold in households of unskilled workers only compared with households containing at least one professional, result adjusted for age, sex, BMI, smoking, alcohol or analgesic intake (Fored, 2003). The same study found that 9 years or less of schooling was associated with a 30% higher risk of CKD than 13 years or more. A US-based prospective study of the effect of area-level SES on progressive CKD found that amongst white men aged 45-64, living in the lowest quartile area for SES was

associated with more than twice the risk for progressive CKD compared to living in the highest quartile (Merkin, 2005). This trend also applied to African American women, but not to white women or African American men, a result implying a differential effect of area-level SES on progressive CKD according to race and sex, although possibly biased by the sampling frame.

Incidence of ESKD varies widely according to ethnicity (USRDS Annual Report and ANZDATA annual report), while SES tends to be lower in ethnic minorities and marginalized groups. Persons with CKD coming from disadvantaged ethnic groups may be less likely to see doctors, specialists or receive adequate treatment, and therefore experience higher rates of unfavourable disease outcomes. However, studies of the role of SES in the relationship between ethnicity and ESKD amongst African Americans suggest SES factors explain only part of the observed excess ESKD risk, and that other factors must be involved (Perneger, 1995; Byrne, 1994). Moreover, the extent of the effect SES has on ESKD incidence has been shown to differ among white compared to African Americans, perhaps because African Americans do not experience the kidney-protective benefits of higher SES (Byrne, 1994). This is likely to relate to issues such as health care access, levels of insurance and other unconsidered risks such as environmental or genetic factors.

In our analysis, we therefore anticipated that ethnicity may be an effect-modifier. When we looked at the different ethnic groups in our study populations separately for the NHANES study there seems to be evidence, based on our analysis, that the association between SES & CKD is less powerful among Mexican Americans than other US ethnic groups.
Care needs to be taken in attributing the difference observed between InterAsia and the other surveys to the ethnic composition of the Thai population. The differences may in fact be caused by other factors specific to the Thai population, rather than their ethnicity.

Possible mechanisms behind these observations:

Several explanations for the association between SES and health have been put forward. The authors of the Whitehall study suggest the association between employment category and coronary and other diseases observed during their study is mediated by factors including deprivation in infancy and childhood, diet and nutrient intake, fewer leisure-time activities, lack of social support, housing and monetary difficulties as well as psychological characteristics such as lack of control over one's working life, numerous recent stressful life events and hostile behaviour. (Marmot MG, 1991).

Also See:

*Fored NDT 2003

*Cass Health and Place 2003 – urban disadvantage and delayed referral.

Limitations of our analysis:

Our data is cross-sectional and therefore comes with the inherent limitations of trying to draw conclusions about where individuals fit within the socioeconomic structure of society based on measures taken at a single time point, and representing only a selection of factors by which SES may be assessed. Household income and employment almost certainly change across the life course. Urban or rural area of residence may also change, although in our study population of adults over 35 reported level of education is likely to remain constant across the life course.

In our analysis we have only considered the impact of individual-level socioeconomic factors on kidney health, and therefore not accounted for context and the possibility that living in a poorer area exerts an independent effect on health.

It is hypothesized that the socioeconomic gradient in health is not simply the result of absolute material standards, but that there is an effect of relative deprivation mediated by psychosocial pathways (Marmot, 2001). In our analysis we constructed income and education as categorical variables, with income grouped into below and above median, and education grouped into 'less than average', 'average' and 'greater than average' for each country respectively. Our definition of these variables takes into account to some extent the potential significance of socioeconomic inequalities.

There is some danger of reverse causality, whereby poor physical and mental functioning as a result of CKD, particularly in its advanced stages, limit overall prospects for employment and income.

The results of this analysis are consistent with previously reported findings that individual-level SES or living in a low SES area is associated with CKD and ESKD (Merkin, 2005). *What does this study add? – looks at this question in a developed country and among more ethnic groups than previously reported.* Understanding the relationship between socio-economic status and CKD is important, not only because it identifies a sector of the community at elevated risk of CKD, but also because of its implications for prevention of CKD and ESKD. Individual-level SES is strongly associated with health care access, health insurance status, diet and numerous other factors which need to be considered in programs attempting addressing the burden of CKD around the world.

This is the largest analysis of the relationship between individual-level SES and CKD, and the first to look comparatively across a number of countries using nationally-representative data. Income, employment and education are associated with CKD risk, however the effects vary between different population groups and this requires further exploration. Our results suggest CKD prevention and management strategies should take account of the higher prevalence in low-income and poorly educated groups.

"Without understanding the social conditions that expose people to individually-based risk factors, interventions will fail more often than they should. This will occur because interventions will be targeted to behaviours that are resistant to change for unrecognized reasons…" "…some social conditions are fundamental causes of disease and as such cannot be effectively addressed by readjusting the individually-based mechanisms that appear to link them to disease in a given context…policymakers should require that all interventions seeking to change individual risk profiles contain an analysis of factors that put people at risk of risks. This will avoid the enactment of interventions aimed at changing behaviours that are powerfully influenced by factors left untouched by the intervention." (Link BG, 1995)

References

Adler NE, Boyce T, Chesney MA et al Socioeconomic status and health. The challenge of the gradient. Am Psychol 1994; 49:15-24

Davey Smith G, Shipley MJ, Rose G. The magnitude and causes of socio-economic differentials in mortality: further evidence from the Whitehall study. J Epidemiol Community Health, 1990; 44: 265-70

Byrne C. et al Race, Socioeconomic status, and the development of end-stage renal disease. Am J Kid Dis, 1994; 23(1):16-22;

Cass A et al. Social disadvantage and variation in the incidence of end-stage renal disease in Australian capital cities. Aust N Z J Public Health, 2001; 25(4):322-6

Coresh reference

Fored CM et al. Socio-economic status and chronic kidney failure: a population-based case-control study in Sweden. NDT, 2003; 18:82-88

Lynch JW et al. Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality, and acute myocardial infarction? Am J Epi 1996; 144:934-42;

Lynch JW. Social Position and health. Am Epidemiol, 1996; 6:21-3;
Lynch JW et al. Cumulative impact of sustained economic hardship on physical, cognitive, psychological, and social functioning. NEJM, 1997; 337(26):1889-1895;
Kunst AE et al Occupational class and cause specific mortality in middle aged men in 11 European
Link BG & Phelan J. Social conditions as fundamental causes of disease.J Health and Social Behav, 1995, Extra Issue: 80-94
countries: comparison of population-based studies. BMJ, 1998; 316:1636-1641
Marmot MG, Rose G, Shipley M, Hamilton PJS. Employment grade and coronary heart disease in British civil servants. J Epi Community Health 1978; 32:244-49;
Marmot MG, Davey Smith G, Stansfield S, Patel C, North F, Head J, White I, Brunner E, Feeney A. Health inequalities among British civil servants: the Whitehall II study. Lancet, 1991; 337; 1387-1393
Marmot M and Wilkinson RG. Psychosocial and material pathways in the relation between income and health: a response to Lynch et al. BMJ, 2001; 322:1233-1236
Merkin, S.S., Coresh, J., Diez Roux, A.V., Taylor H.A., Powe, N.R. Area Socioeconomic Status and Progressive CKD: The Atherosclerosis Risk in Communities (ARIC) Study. Am J Kidney Disease, 2005; 46(2):203-213
Pappas G, Queen S, Hadden W, Fisher G. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. NEJM, 1993; 329:103-9;
Perneger TV et al. Race and End-Stage Renal Disease: Socioeconomic Status and Access to Health Care as Mediating Factors. Archives of Internal Medicine, 1995; 155(11):1201-1208; Byrne C et al. Race, socioeconomic status, and the development of end-stage renal disease. AJKD, 1994; 23(1):16-22
Von Rossum et al. Employment grade differences in cause specific mortality. 25 year follow-up of civil servants from the first Whitehall study. J Epi Community Health 2000; 54:178-84
Young EW et al. Socioeconomic status and end-stage renal disease in the United States. Kidney Int, 1994; 45(3):907-11;
Reference for waist circumference
USRDS Annual Report and ANZDATA annual report