Bachelor thesis: Possible Benefits With Combined Creatinine and cystatin C Test for Estimated GFR

Tobias Hedberg

June 5, 2015
1 Introduction

1.1 Background
When seeking medical care creatinine is used as an indicator for kidney malfunction in many different contexts and is measured as concentrations in blood tests. Creatinine levels are roughly proportional to the rate at which blood is filtered through the kidneys into primary urine. This rate can also be called glomerular filtration rate (GFR) and is used to predict the level of treatment needed in patients with kidney dysfunction. The benefits with creatinine is a widely available analysis were results return in a timely fashion. However, the level of creatinine is affected by factors in addition to GFR such as muscle mass, blood volume, sex and age [1]. Because of this, interest has been raised for a protein called cystatin C which provides a more exact estimate of GFR, but is costlier.
Because kidney decease is more common and relevant to patients of older age groups only patients with an age higher than 65 are being considered in this thesis. Additionally Sweden has an average life time expectancy which is rising [2], which creates an economical dilemma as well.

1.2 Aim
The goal of this thesis is to research possible benefits with a combined creatinine and cystatin C test compared to only a creatinine test when estimating a patient’s GFR.

1.3 Previous Research
A lot of research has been made on the kidneys and their functions and quite a large amount of models to estimate GFR have been presented, such as creatinine combined with cystatin C [3]. Not much about uncertainty of diagnosis in the diagnostic threshold areas of GFR has been researched, however. Placing a patient on the right side of a chronic kidney disease (CKD) grading line is essential to providing proper treatment, which is why this thesis has focused almost exclusively on this matter.
2 Medical Background

2.1 The Kidney

Normally a human has one pair of kidneys in their bodies. Located at the lowest pair of the ribs, next to the spine in the rear part of the abdominal cavity. They have the shape of a bean and size of a human fist. The kidneys play an important role in our body and make sure a lot of waste products and unwanted substances gets filtrated out of our body.

The kidney is provided with a steady flow of blood through one, or up to three, afferent (carrying inwards) arteries from the aorta. Their sizes can vary individually among the arteries in each kidney and as a note it can be added that this is important for deciding whether two patients are a match for transplantation or not. The kidneys works like a filtration unit in the body and consists of a network of capillaries that follows the afferent arteries and is called the glomerulus. Further on this blood is transferred through the organ in arteries of decreasing size. These carry blood into the smallest functioning unit of the kidney, the nephron. A partition of the blood, roughly equivalent to its plasma, exits the blood vessel into the Bowman capsule, which is the first part of the nephron. From here, the plasma, or primary urine, travels up and down the kidney in a tubule where it is continually concentrated through a series of ion channel and concentration gradient activities. The total accumulated rate at which plasma exits into the millions of Bowman capsules in both kidneys is the GFR. Blood moves on from the capsule after releasing some of its plasma in an efferent arteriole carrying it to a capillary network on the outside of the tubules where it supplies the nephrones with nutrients and oxygen. It then combines in bigger and bigger veins carrying it back to the heart, while the concentrated urine exits the kidneys into the ureters to the urinary bladder and onwards. [1]

2.2 Glomerular Filtration Rate (GFR)

The filtration of unwanted or high levels of different substances is situated in the above mentioned glomerulus. Filtration rate is measured in mL/min and in Figure [1] the values of the different stages of a kidney’s health is graphically described with red (0-15 mL/min), yellow (15-60 mL/min) and green (larger than 60 mL/min) [4]. Firstly, the red area represents patients with end stage renal decease, henceforth refereed to as ESRD. This is a serious diagnosis to get, you are likely to begin dialysis and the risk for death is severe. Secondly, the yellow area represents patients with chronic kidney decease, henceforth refereed to as CKD. This is a diagnosis where
the treatment usually is dietary changes and take an increasing number of medicines as decease progresses. Lastly, the green area represents healthy patients or patients with a temporarily slightly decreased kidney function.

Figure 1: GFR values representing different stages of the kidneys’ health. Red indicating 0-15 mL/min, yellow indicating 15-60 mL/min and lastly green indicating larger than 60 mL/min [4].

2.3 Creatinine

Creatinine is a waste product from the muscles in our bodies and has no role in our body. It is formed when creatinine phosphate, or phosphocreatine, is worn down. Phosphocreatine is quickly released during the first seconds when the body needs to move rapidly.

Creatinine levels are measured routinely by a blood test as an indicator of general health because early kidney decease lacks clear symptomatology and a great number of deceases and substances affect kidney function. Creatinine analysis has been proven a good method for estimating the care seeker’s kidney’s health, through GFR. As GFR decreases creatinine rises as a result of decreasing clearance. The level of creatinine in the blood is dependent on the constitution of one’s body, to be more precise, the muscle mass. A high value is stating that something might be wrong, but if you have an athletic body and thus a lot of muscle mass, this value might be just all right [7]. The price for a creatinine test is 9 SEK in the county of Skåne at the university hospitals [6].
2.4 Cystatin C

In contrast to creatinine cystatin C is not dependent on the muscle mass and thus more reliable. Cystatin C is a protein and can be found in all of the human body’s cells and its main function is to maintain and regulate the activity of some protein degrading enzymes [5]. However a cystatin C test is much more expensive than a creatinine test. The cost varies from county to county. As a reference the price for a cystatin C test is 24 SEK in the county of Skåne at the university hospitals [6].
3 Material

The material used in this thesis is collected from the research behind the article in Scandinavian Journal of Clinical and Laboratory Investigation [3]. They measured iohexol clearance (mL/min/1.73m²) in 857 patients and compared it to the formulas described further down in this thesis, equation (1) and (2). The data set consisted of both female (44%) and male (56%) patients from the cities Lund and Malmö, Sweden. Their age, sex, weight, length were also measured and included in the data set. The earlier work tried to figure out a formula without the weight parameter and therefore they came up with the formulas presented under the next segment in this thesis.
4 Design of Statistical Model

Human bodies are very complex and scientists still discover new areas to do further research on. To estimate and conclude different values in the body it would be far too complicated and expensive to measure all imaginable values to conclude a diagnosis for a potentially ill patient. This is why a medical doctor tries to narrow it down to just a couple of tests. As provided in the section above, a blood test is the fastest way to approximate the kidneys health status. Every test is only accurate to a certain point and this is why a statistical model should be considered. In this thesis the creatinine (pCr) will be indexed K and the combined test, with both creatinine and cystatin C (pCyC), will be indexed C.

In the formerly referred article in Scandinavian Journal of Clinical and Laboratory Investigation (SJCLI) two formulas for estimating GFR were presented with either creatinine (pCr) referred to in equation (1), or a combination of creatinine and cystatin C (pCyC) shown in equation (2).[3]

\[
GFR = e^{X - 0.0124 \cdot \text{age} - 0.339 \cdot \ln(\text{age}) - 0.226 \cdot (\text{if female})}
\]

where \(X\) is defined as

\[
X = \begin{cases} 
4.621 - 0.0112 \cdot \text{pCr} & \text{if pCr} < 150 \ \mu\text{mol/L} \\
8.17 + 0.0005 \cdot \text{pCr} - 1.07 \cdot \ln(\text{pCr}) & \text{if pCr} \geq 150 \ \mu\text{mol/L}
\end{cases}
\]

\[
GFR = e^{X - 0.718 \cdot \ln(p\text{CyC}) - 0.00452 \cdot \text{age} + 0.0549 \cdot \ln(\text{age}) - 0.153 \cdot (\text{if female})}
\]

where \(X\) is defined as

\[
X = \begin{cases} 
5.02 - 0.00716 \cdot \text{pCr} & \text{if pCr} < 150 \ \mu\text{mol/L} \\
7.07 + 0.0004 \cdot \text{pCr} - 0.636 \cdot \ln(\text{pCr}) & \text{if pCr} \geq 150 \ \mu\text{mol/L}
\end{cases}
\]

To start off, a model for creatinine will be formulated and explained first, followed by a combined model with both creatinine and cystatin C. The aim is to see if there are possible benefits of taking a second test (cystatin C), after the creatinine test. The second test might then give a better picture of the patients kidney status. Based on the results from the first (creatinine) test the goal with the following segments and sections are to conclude if one can decide whether the doctors decision will be better by doing a second test.

The simplest model would be a direct relationship between GFR and creatinine, like the one in equation (3) where \(f(K)\) is equal to equation (1).

\[
GFR = f(K)
\]
In Figure 2, the relationship between GFR is estimated by the level of creatinine with the formula in equation (1). If a comparison is made with Figure 3 where the actual values of the GFR and their correlated creatinine levels are presented, a clear relationship can be seen. In Figure 3, the blue markers represent female patients and the orange markers represent male patients.

Figure 2: GFR model from SJCLI[3].

Figure 3: GFR versus creatinine. The blue markers represent female patients and the orange markers represent male patients.

Since the body is such a complex structure it is more true to life to have a relationship where the relationship is as in equation (4), where the value of GFR is an expected value. Because this is not a linear proportionality a deterministic prediction of exact GFR cannot be made. This is reflected in
a clinical context by the fact that more weight is attached to a significant change of creatinine levels in subsequent analyses than is given to a single analysis alone.

\[ \langle \text{GFR} \rangle = f(K) \]  

(4)

With the body being as complex as it is and the model having a level of uncertainty the \( \xi \)-term is added to the formula, as seen in equation (5). This is a term consisting of the standard deviation multiplied with \( \epsilon \), which is a random variable and it is \( N(0, \sigma) \) distributed.

\[ \text{GFR} = f(K) + \xi_K \]  

(5)

The properties of \( \xi_K \) is as displayed below. In Figure 4 the normal probability plot for \( \xi_K \) can be seen and a good approximation is that \( \xi_K \) is taken from a normal distribution.

\[ \langle \xi_K \rangle = 0 \quad \langle \xi_K^2 \rangle = \sigma_K^2 \]  

(6)
When actual GFR and GFR estimated by the formula in equation (1) is plotted against each other the conclusion can be drawn that they have a heteroscedastic relationship and this is shown in Figure 5. This means that the variation is increasing with larger values of GFR. By doing a logarithmic transformation of these functions the variation is now homoscedastic and more suitable to make further studies on. The new representation of the relationship is presented in Figure 6.
The function for the correlation between the logarithmic forms of \( f(K) \) and GFR is then estimated, displayed as equation (7). This estimation of the parameters and variables, and all to follow, is further explained in section 5.

\[
\ln GFR = \alpha_K + \beta_K \ln(f(K)) + \xi_K
\]  

(7)

The above procedure is then made in the very same way as above, but with the only difference in that now the function \( f(K) \) is replaced by \( f(C) \), which is equation (2). This combined test is now represented by the final equation (8). The newly introduced variables \( \alpha_C, \beta_C \) and \( \sigma_C \) are determined in the same procedure as for creatinine and will be further explained how they were estimated in section 5.

\[
\ln GFR = \alpha_C + \beta_C f(C) + \xi_C
\]  

(8)

The benefit with a combined test is that the test’s accuracy will increase significantly since it is based on two separate test, especially since the cystatin C is independent of body mass. Since \( f(C) \) uses creatinine in its formula they are also dependent on each other. This is verified by plotting \( \xi_K \) and \( \xi_C \) against each other and this is shown in Figure 7.

Since these are both normally distributed these form a bi-variate distribution with three variables; \( \sigma_K \) the standard variation of creatinine, \( \sigma_C \) the standard variation of the combined test and \( \rho \) the correlation coefficient for the bivariate distribution.
5 Estimations of Model Parameters

Why is a parametrical model necessary; why could not the data be used directly? The reason for this is the number of data points in the provided data. Since the given number of patients is definite (857 patients), a model is the only suitable solution. The limits to what can be described by the model is the validity of the original model, in this case the $f(K)$ and $f(C)$. And it has been proven very satisfactory in the article in SJCLI [3].

GFR depends of the parameters of $f(K)$, which are levels of creatinine, age and sex. The likelihood function ($L$) is proportional to the conditional probability of any given GFR provided the values of said parameters. This leads to the proportionality in equation (9):

$$L \propto P(GFR_i \mid K_i, age_i, sex_i)$$ (9)

Using the Maximum Likelihood method to estimate the parameters of this statistical model, the product of all probabilities the total probability is calculated according to equation (10).

$$L = \prod_{i=1}^{N} P(GFR_i \mid K_i, age_i, sex_i)$$ (10)

Furthermore the natural logarithm of $L$ is taken and this results in equation (11).
\[ l = \ln(L) = \sum_{i=1}^{N} \ln(P(GFR_i \mid K_i, \text{age}_i, \text{sex}_i)) \] (11)

To estimate the parameters introduced \((\alpha_K, \alpha_C, \beta_K, \beta_C, \sigma_K, \sigma_C \text{ and } \rho)\) in the previous section a scheme for estimating all the parameters was made. A good way to optimize the parameters without creating a biased estimator is to use the Nelder-Mead method [7]. This method works great if a good initial value is provided. These values were calculated in a five-step scheme presented below.

- **\(\alpha_K\) and \(\beta_K\),**
  - By plotting \(\ln(\text{GFR})\) against \(\ln(f(K))\) and taking the mean value of \(\alpha_K\) and \(\beta_K\) based on the function in equation (7).

- **\(\sigma_K\) and thereby \(\xi_K\),**
  - By subtracting and thus calculate the residual from: \(\ln(\text{GFR}) - \alpha_K - \beta_K \cdot \ln(f(K)) = \text{res}_K\)

- **\(\alpha_C\) and \(\beta_C\),**
  - By plotting \(\ln(\text{GFR})\) against \(\ln(f(C))\) and taking the mean value of \(\alpha_C\) and \(\beta_C\) based on the function in equation (8).

- **\(\sigma_C\) and thereby \(\xi_C\),**
  - By subtracting and thus calculate the residual from: \(\ln(\text{GFR}) - \alpha_C - \beta_C \cdot \ln(f(C)) = \text{res}_C\)

- **Correlation coefficient \(\rho\).**
  - By plotting the residual from creatinine and the residual from cystatin C and taking the mean value of the coefficient.

After the five steps explained above an initial value was provided and put into this method, the Nelder-Mead method, which is also a built-in function in MATLAB and named *fminsearch* [7].
6 Optimisation of Decision

Since the cost of the treatment from one diagnosis to another varies a lot, the decision was made to eliminate the cost from the analysis and only take into consideration when someone might receive wrong treatment based on the (GFR) test results. If the patient receives the right treatment it might be at a cost for the state or the hospital but the cost is strictly relevant to the diagnosis and not an unnecessary additional cost where the money could be better spent on something else.

The qualitative goal with optimisation of the decision is to help the doctors determine the right diagnosis or treatment when suspecting kidney malfunction by testing the GFR. The quantitative goal with this section is to minimize the probability of an incorrect diagnosis and therefore maximize the probability of giving the right treatment.

A creatinine test costs less than a cystatin C analysis and the creatinine test is done as a routine when suspecting kidney malfunction and the cystatin C is used when a more precise and more accurate value of the GFR is desired. Consequently one can ask the question; when is it necessary or beneficial to do another test (cystatin C)? This will be covered and explained in this section and in the sections to follow.

As mentioned in section 2, three kinds of diagnosis are to chose from, ESRD, CKD and healthy. To optimize the decision three conditional probabilities were created as seen in equation (12a), (12b) and (12c). These probabilities determines how probable it is to have one of the three diagnoses given the test result from the creatinine test.

Probabilities for different diagnoses:

\[ P(ESRD \mid f(K)), \]  \hspace{1cm} (12a)  
\[ P(CKD \mid f(K)), \]  \hspace{1cm} (12b)  
\[ P(Healthy \mid f(K)). \]  \hspace{1cm} (12c)

With these probabilities the right diagnosis and treatment accordingly is the one with the largest probability. By plotting the probability for wrongly treating a patient against GFR based on creatinine test gives the results as shown i Figure 8.
Figure 8: Probability of misdiagnosing patients as a function of creatinine analysis results (GFR).

7 Monte-Carlo Analysis on Test Results

If the probability for misdiagnosing the patient is called $P_{\text{wrong}}$ and the probability of doing so, after an extra test, the combined cystatin C test, is called $P_{\text{wrong}}^\ast$. The relationships between these test will be as in equation (13) since cystatin C tests are more precise\cite{8}.

$$P_{\text{wrong}} \geq P_{\text{wrong}}^\ast$$ (13)

This does not mean that every decision based on the latter test are better than a decision made from only the first test, this is just the probability for giving the patient the wrong treatment based on their actual kidney’s health. When is it beneficial for a doctor to prescribe the combined test? If the difference between the probabilities is very small there is no need for an extra test, since it may not be worth the extra cost.

With the now known value of the estimated GFR ($f(K)$) the bivariate distribution for $\xi_C$\cite{9} as displayed in equation (14) was used to calculate the probability for a value of $f(C)$. This was done by using equation (7) to calculate $\ln(\text{GFR})$ and by inserting this into equation (8) and thus the only term left unknown is $f(C)$, which is then calculated. $\epsilon$ is drawn from a standard normal distribution.
\[ \xi_C = \rho \sigma_C \frac{\xi_K}{\sigma_K} + \sqrt{1 - \rho^2} \sigma_C \epsilon \]  

(14)

By sampling from the GFR distribution (creatinine-based) where every sample now has its unique, known GFR value it is possible to use a Monte-Carlo method to sample from the distribution, and by using the method explained in section \[ f(K) \] is replaced by \[ f(C) \]. The fraction of incorrect decisions is estimated for each case. This new probability is then plotted against GFR, as shown in Figure 9.

Figure 9: Probability plot for wrongly treating patients based on a creatinine (blue) and a combined (green) test.

In this figure one can see that the probability for wrongly diagnosing patients is very much the same for creatinine (marked as blue in the figure) and the combined test (marked as green in the figure), except for values close to the borders between the different kidney’ statuses. In these areas the probabilities for mistreating patients are significantly reduced. Notice that the values are only estimates since they are constructed by sampling, but the estimation error is rather small, as can be seen by the statistical noise in the graph.
8 Discussion and Conclusions

This thesis was constructed from the models in equation (1) and equation (2). This relationship is given as a formula relating measurements to the GFR value, but does not account for uncertainty in the estimation, hence the goal was to explicitly account for statistical variations.

The approach was changed from a cost perspective on estimating GFR to the perspective of minimizing the probabilities of misdiagnosing a patient, based on the results from the blood test. To know the exact GFR is seldom of greater importance to a doctor, what’s of great value is to determine the correct diagnosis. The relevant quantity to minimize is the cost the incorrect decisions might give rise to. However, since such measure are very hard to estimate, it was therefore decided to instead focus on fraction of correct decision, since such variable is easy to compute in the model.

To determine if it is effective to also measure cystatin C, the fraction of correct decision concerning treatment of the patient was compared using a Monte-Carlo procedure. The conclusions from this report there are potential benefits of using this clinically in different hospitals in Sweden since it might be very cost effective in using cystatin C tests only when applicable from Figure 9.

8.1 Caveats

The raw data of 857 patients and their age, sex, cystatin C, creatinine and their actual GFR value was received from the writers of the article in Scandinavian Journal of Clinical and Laboratory Investigation. This data, used in the current thesis, is not completely representative for the entire population of Sweden, since patients with kidney dysfunction is overly representative.
References


