Chapter 2

Methodology

This chapter describes the method used in the study including study design, population and sample, inclusion criteria, exclusion criteria, variables and conceptual framework, data collection and data management, measurements of renal function and statistical methods.

2.1 Study design

Patients with gout often have concurrent CKD and the relationship between two conditions is still unclear. Therefore a cross-sectional study was conducted to investigate the risk factors of CKD among gout patients in Nongjik Hospital, Pattani Province.

2.2 Population and sample

Population

The target population comprised all outpatients who had diagnosis of gout (ICD 10 coded as "M10") in Nongjik Hospital which is one of community hospitals in Pattani Province, Southern Thailand. The diagnosis of gout was based on symptoms and serum uric acid level.

Sample

The samples were 167 gout patients at outpatient department of Nongjik Hospital during January 2004 to December 2010.

2.3 Inclusion criteria

1. The outpatients who had diagnosis of gout.
2. The gout patients who had at least 3 serum creatinine laboratory results in study period.
3. The gout patients who have been treated with anti-gout regularly.

2.4 Exclusion criteria

1. The gout patients who had acute kidney disease.
2. The gout patient who had less than 3 serum creatinine laboratory results.
3. The gout patients lacked of follow-up for more than 1 year in study period.

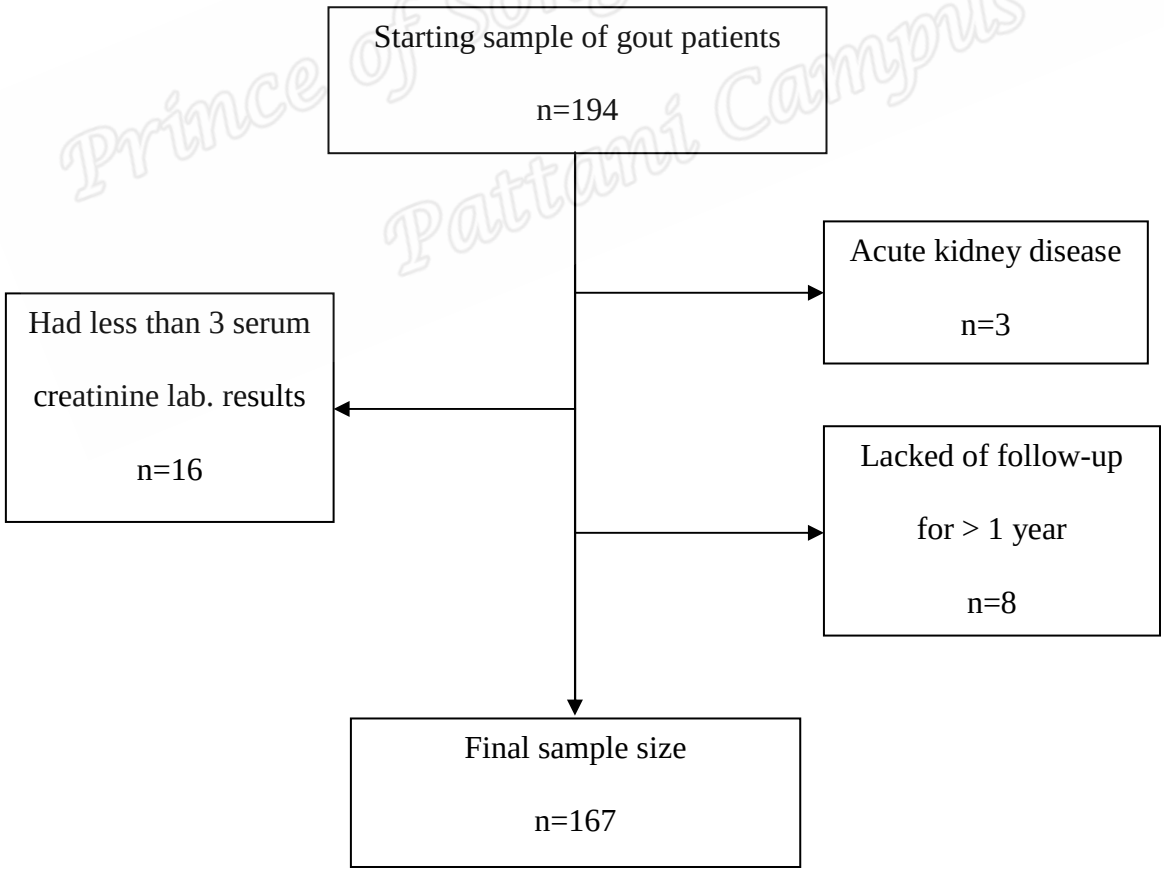

Figure 2.1: Diagram of gout patient selection

2.5 Variables and conceptual framework

The variables of this study comprised seven determinants and one outcome variable.

Determinant variables

The seven determinants were BMI, presence of co-morbidity; HT, DM and dyslipidemia, thiazide use, anti-gout agent and serum uric acid.

BMI was defined as the weight in kilograms divided by the square of height in meters (kg/m^2) according to the World Health Organization (WHO) 2004. Patients who had BMI $\geq$ 23 kg/m^2 were defined as overweight.

The presence of co-morbidity; HT, DM and dyslipidemia was defined as taking anti-hypertensive, anti-diabetic and lipid lowering medications.

Thiazide use was defined as taking hydrochlorothiazide (HCTZ). It is used to treat high blood pressure and fluid retention. HCTZ also reduced excretion of uric acid by the kidney.

Anti-gout agent was defined as taking colchicine or allopurinol plus colchicine. Colchicine is a medication for acute gout. It relieves pain and reduces the inflammation of acute gouty arthritis. Allopurinol is a medication used to treat chronic gout. It is not effective for acute gout since it does not relieve pain or inflammation.

Serum creatinine and uric acid level were based on the last laboratory results. The serum uric acid concentration < 8 mg/dl was designated as the desirable goal for gout patients receiving uric acid reduction therapy in Nongjik Hospitol.

Outcome variable

The outcome of interest was CKD which was identified as a dichotomous variable; CKD or non-CKD.

Conceptual framework

Figure 2.2 shows a conceptual framework for summarize the variables of interest in the study.

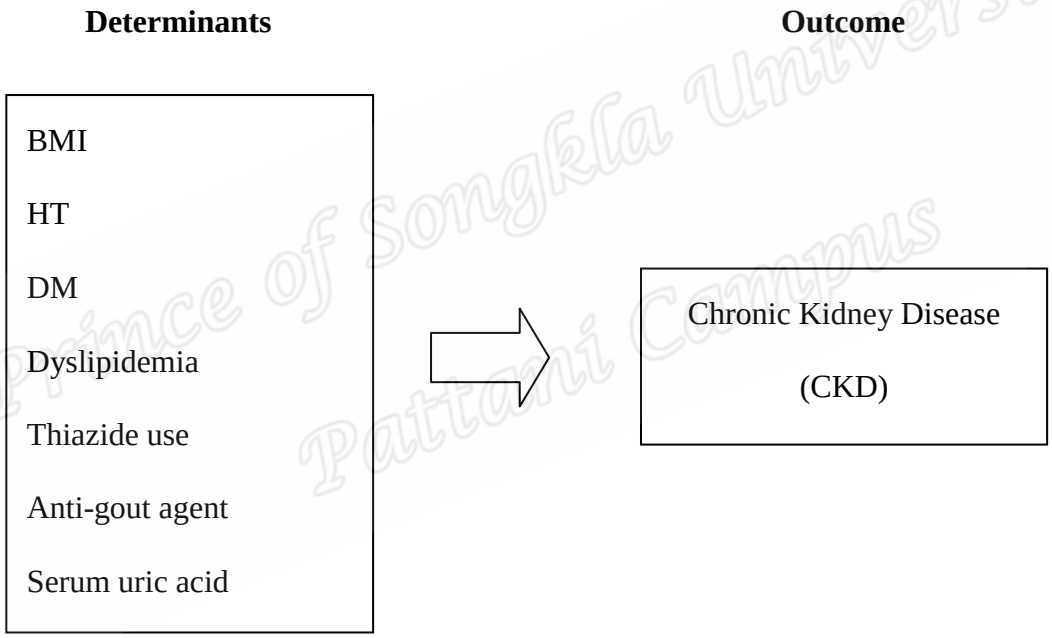

Figure 2.2: Conceptual framework showing variables in the study

2.6 Data collection

The secondary data were collected from Nongjik Hospital database on 194 patients with gout at outpatient department from January 2004 to December 2010. The data included gender, age, weight (kg), height (cm), present of HT, DM, dyslipidemia, thiazide use, serum creatinine (mg/dl), anti-gout drugs and serum uric acid (mg/dl).

2.7 Measurements of renal function

According to the NKF/K-DOQI guideline, CKD is defined as either kidney damage or eGFR less than 60 ml/min/1.73 m^2 for three months or more. For this study CKD was defined as eGFR< 60 ml/min/1.73 m^2. eGFR was calculated to determine renal function with the simplified Modification of Diet in Renal Disease (MDRD) equation defined as follows:

$$eGFR = 186 \times serum\ creatinine^{-1.154} \times age^{-0.203} \times 0.742(if\ the\ individual\ is\ female) \times 1.21(if\ the\ individual\ is\ African\ American)$$

In this equation, eGFR is expressed as ml/min/1.73 m^2, serum creatinine is expressed as mg/dl and age is expressed as years. A body surface area of 1.73 m^2 is the normal average adult surface area. The race is defined as African American or non African American. This is due to higher average muscle mass and creatinine generation rate in African Americans.

The stage of CKD was based on staging described by the NKF KDOQI as follows:

All patients with eGFR ≥ 90 ml/min/1.73 m^2 are defined as having no CKD. Patients with eGFR 60-89 ml/min/1.73 m^2 are assigned to the stage 2 (mild disease) category, eGFR 30-59 ml/min/1.73 m^2 are assigned to the stage 3 (moderate disease) category, eGFR 15-29 ml/min/1.73 m^2 are assigned to the stage 4 (severe disease) category and eGFR < 15 ml/min/1.73 m^2 are assigned to the stage 5 (Kidney failure) category.

eCrCl was calculated by using Cockcroft and Gault equation defined as follows:

$$eCrCl = \frac{(140 - age) \times weight(kg) \times 0.85(if\ the\ individual\ is\ female)}{72 \times serum\ creatinine}$$

Dose adjustment of gout medications according to eCrCl was based on published guidelines (Table 2.1).

Table 2.1: Dose adjustment of anti-gouts according to eCrCl

Drug	eCrCl (ml/min)	Dose
Allopurinol	0	100 mg 3 times per week
	10	100 mg alternate days
	20	100 mg daily
	40	150 mg daily
	60	200 mg daily
	80	250 mg daily
	100	300 mg daily
	120	350 mg daily
	140	400 mg daily
Colchicine	< 10	Avoid use
	< 50	reduce dose by half

2.8 Statistical methods

All risk factors are categorical variables. Pearson's chi-square test was used to assess the associations between the determinant variables and outcome. Multivariate analyses were performed to investigate any independent associations between the

predictors and outcome by using logistic regression and not significant variables were eliminated from the model by backward method.

Descriptive statistics

The outcome and all determinants were categorical variables which were summarized by using percentages.

Univariate analysis

Pearson's chi-square test and 95% confidence intervals for odds ratio were used to assess the association between the determinant variables and outcome. The formulas based on contingency tables are as follows (*X* is a determinant of interest, *Y* is outcome).

2 × 2 table

X is the independent variable, *Y* is the dependent variable. The odds ratio is a measure of the strength of an association between two binary variables (McNeil, 2006). So both the outcome and the determinant are dichotomous. For example, *X* is HT (coded as present and absent) and *Y* is CKD (coded as non-CKD and CKD). The data may be presented as a table with four cells (called a 2 × 2 contingency table) as follows.

HT	Outcome		Total
	Non-CKD	CKD	
Present	a	b	a + b
Absent	c	d	c + d
Total	a + c	b + d	a + b + c + d

The estimate the odds ratio is

$$OR = \frac{a \times d}{b \times c} \tag{2.1}$$

One method of testing the null hypothesis of no association between present HT and CKD is using the z-statistic $z = \ln(OR) / SE$, where SE is the standard error of the natural logarithm of the odds ratio (McNeil, 1996). An asymptotic formula for this standard error is given by

$$\text{SE } (\ln \text{OR}) = \sqrt{\frac{1}{\text{a}} + \frac{1}{\text{b}} + \frac{1}{\text{c}} + \frac{1}{\text{d}}} \tag{2.2}$$

A 95% confidence inteval (95% CI) for the odds ratio is given by

$$95\%\ CI = OR \times \exp(\pm 1.96 \times SE[\ln OR]) \tag{2.3}$$

Pearson's chi-square test

Pearson's chi-square statistic for independence is defined as

$$\chi^2 = \frac{(ad - bc)^2 n}{(a + b)(c + d)(a + c)(b + d)} \tag{2.4}$$

The *p-value* is the probability that a chi-square distribution with 1 degree of freedom exceeds this statistic. If the *p-value* is less than a pre-defined significant level, then the null hypothesis is rejected.

Logistic regression

Multiple logistic regression analysis was used for modeling the association between several independent variables and CKD. Logistic regression is a method of analysis that gives a particularly simple presentation for the logarithm of the odds ratio describing the association of a binary outcome with factors. It automatically provides estimates of odds ratios and confidence intervals for specific combinations of the risk factor (McNeil, 1996). For a set of predictor variables $x_1, x_2, \ldots, x_k$ and a binary outcome *Y* the logistic regression model takes the following form:

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \sum_{i=1}^{k} \beta_i \, x_i \qquad (2.5)$$

Where P denotes the probability of occurrence of the specified outcome; x are the set of independent variables, α is the constant coefficients, β is the set of regression coefficients and *k* is the number of predictor variables. In this studies the outcome (Y) is CKD (coded as '0 = non-CKD' or '1 = CKD'). The probability of the 'CKD' category (Y = 1) can be expressed as

$$P[Y = 1] = \frac{\exp\left(\alpha + \sum_{i=1}^{p} \beta_i \, x_i\right)}{1 + \exp\left(\alpha + \sum_{i=1}^{p} \beta_i \, x_i\right)} \qquad (2.6)$$

Using the logistic regression model for the data arising from a two-by-two table, we suppose x_i = 1 or 0, that is, the values of determinant x are taken to be 1 (exposure) and 0 (no exposure). Thus the logistic regression model can be written as

$$\ln\left\{\frac{P(Y=1|X=1)}{1-P(Y=1|X=1)}\right\} = \alpha + \beta x \qquad (2.7)$$

$$\ln\left\{\frac{P(Y=1|X=0)}{1-P(Y=1|X=0)}\right\} = \alpha \qquad (2.8)$$

The equations (2.7) and (2.8) are the logarithms of the odds for the outcome given the exposure (x = 1) and non-exposure (x = 0), respectively. After exponentiation each equation, the odds for exposed and non-exposed groups can be written as exp (α + β) and exp (α), respectively. Thus the odds ratio has the simple formula

$$OR = \frac{\exp(\alpha+\beta)}{\exp(\alpha)} = \exp(\beta) \qquad (2.9)$$

Comparing models

Checking the adequate of the original logistic model after dropping a subset of variables from the model will use a value of the likelihood function of the model (Woodward, 2005). Let L_{sub} and L_{full} represent the maximum values of the likelihood function under the sub model and under the full model respectively. Standard likelihood theory indicates that if the sub model is adequate, the difference $2\log L_{full} - 2\log L_{sub}$ will have a distribution that is approximately χ^2_d where *d* is number of variables dropped, if a dropping variable is a categorical variable with multiple categories, *d* is number of the categories. This difference can be expressed as a difference of deviances:

$2\log L_{full} - 2\log L_{sub}$ = deviance of sub model - deviance of full model (2.10)

This difference represents the increase in the deviance after dropping the *d* terms from the model. This difference will always be positive, if the increase is small then dropping the extra terms will not increase the deviance by very much and so the variables can be dropped to get a simple model. The difference in the deviance has approximately a χ_d^2 distribution if the dropped variables are not needed in the model. Thus we can calculate a *p-value* to test the hypothesis that the dropped variables are not needed by comparing this difference in the deviance to a χ_d^2 distribution. A small *p-value* provides evidence against the sub model. That is a small *p-value* indicates that we should not drop all of *d* of the variables from the model.

2.9 Flow of data processing

After obtained the raw data, the records were stored in an excel file. Then cleaning and checking processes were done by R Package Program. Pearson's chi-square test was used to assess the association between outcome and each determinant and used the logistic regression analysis to assess the association between outcome and all determinants. The graphical and statistical analyses were performed using R program. The flow diagram for data processing was summarized in Figure 2.3.

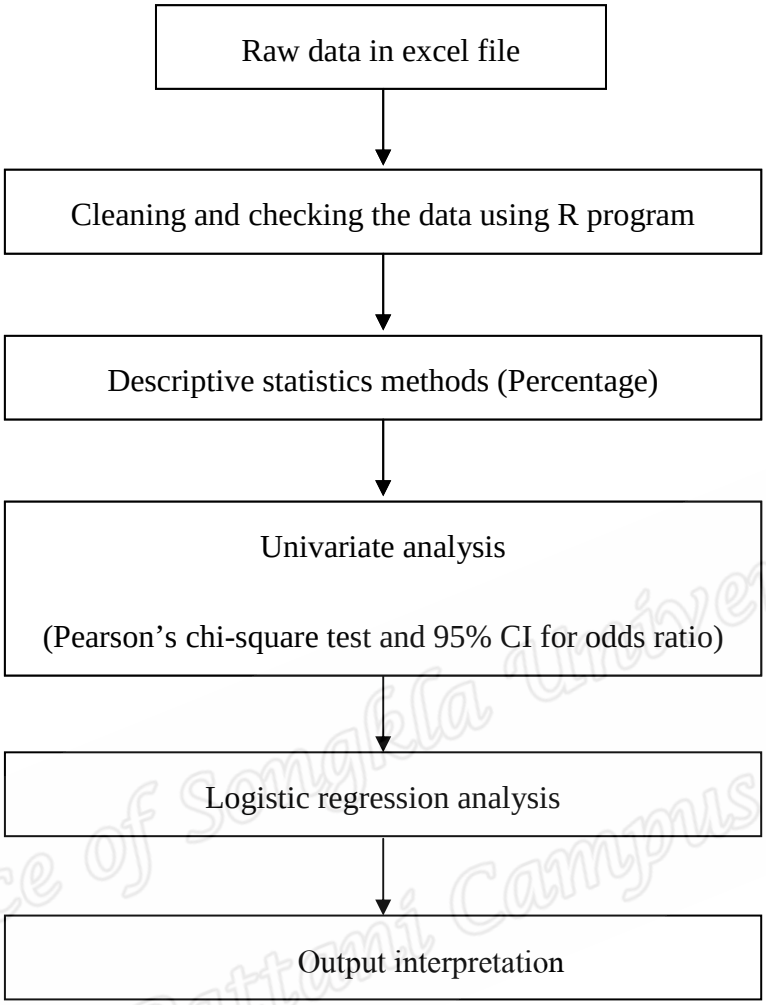

Figure 2.3: Flow diagram for data processing