J Nephropharmacol. 2019; 8(2): e20.

http://www.jnephropharmacology.com DOI: 10.15171/npj.2019.20

Journal of Nephropharmacology

Statistical significance of prognostic factors on the progression of chronic kidney disease through simulation study

Alka Sabharwal[1], **Gurprit Grover**[2], **Shrawan Kumar**[1*]

[1]Department of Statistics, Kirori Mal College, University of Delhi, Delhi, India
[2]Department of Statistics, University of Delhi, Delhi, India

ARTICLEINFO

Article Type:
Original

Article History:
Received: 30 December 2018
Accepted: 14 February 2019
ePublished: 8 March 2019

Keywords:
Chronic kidney disease
Likelihood ratio test
Permutation test
Simulation
End-stage renal disease

ABSTRACT

Introduction: Chronic kidney disease (CKD) is a serious public health problem. It affects 10% to 16% of adults around the world. In India, the approximate prevalence is 800 per million population (pmp) and incidence of end-stage renal disease (ESRD) is 150 -200 pmp.
Objectives: To evaluate the differences in the impact of prognostic factors for the progression of CKD into higher stages using appropriate and robust tests.
Materials and Methods: Permutation test and likelihood ratio test were applied to ascertain the statistical significance of the prognostic factors for the progression of CKD into higher stages. The data consists of 100 non-hospitalized CKD patients of three stages namely stage 2, stage 3 and stage 4. A simulation study has been carried out to determine the power of permutation test and likelihood ratio test testing the significance of difference between the values of the parameters of the distribution of prognostic factors involved in the progression of CKD in stage 2, stage 3 and stage 4.
Results: Permutation test and likelihood ratio test based on our data set suggest that serum creatinine, urea, hemoglobin, albumin and age are the significant factors for the progression of CKD to higher stages.
Conclusion: Under various health conditions using simulation study, all the factors included in the study are responsible for the progression of the disease.

Implication for health policy/practice/research/medical education:
The present work helps in identifying the risk factors of CKD. It will help in implementing screening of at-risk populations which in turn will increase the early detection, initiate treatment of modifiable risk factors for ESRD, along with appropriate treatment of chronic kidney disease. The economic burden caused by the cost of renal replacement therapy /kidney transplant might be reduced by early detection of risk factors.
Please cite this paper as: Sabharwal A, Grover G, Kumar S. Statistical significance of prognostic factors on the progression of chronic kidney disease through simulation study. J Nephropharmacol. 2019;8(2):e20. DOI: 10.15171/npj.2019.20.

Original

Introduction

Chronic kidney disease (CKD) is a serious public health problem, which could lead to end-stage renal disease (ESRD) and increased cardiovascular morbidity and mortality (1). CKD affects 10% to 16% of adults around the world (2). CKD is usually asymptomatic until later stages and accurate prevalence data are lacking (3). The factors race, age, gender and serum creatinine are used for the computation of estimated glomerular filtration rate (eGFR) to determine the stages of CKD patient. Other prognostic factors such as urea, albumin, hemoglobin, body mass index along with concurrent diseases such as hypertension and diabetes are traditionally or nontraditionally are associated with CKD.

Hsu et al applied survival analysis to evaluate factors associated with time to an event of interest (e.g., ESRD and mortality) among CKD populations (4). Roy et al fitted three logistic regression models of progressive CKD. Age, sex, and race were taken as the predictors for the first model. The second model included GFR and proteinuria along with the predictors of first model. In the third model, they included other predictors; angiotensin II receptor blocker, an indicator for any history of cardiovascular disease, diabetes, educational level, systolic blood pressure, and body mass index (5).

In recent years, there is a considerable increase in

***Corresponding author**: Shrawan Kumar, Email; shrawan.kmc@gmail.com

research and applications related to permutation tests (6). It has been used in clinical trials, functional data analysis, spatial statistics, principal component analysis, shape analysis, survival analysis and many other fields (7-13). Galimberti et al have applied multivariate permutation test with simulations for the comparison of bone marrow transplantation and chemotherapy in the treatment of pediatric leukemia (14). Giancristofaro et al compared the sampling heterogeneity of a categorical variable X in two populations using permutation test (15). In practice, a large number of problems may also be solved effectively and usefully by applying traditional parametric methods (likelihood-based) or nonparametric methods (rank-based). Although in relatively mild conditions their permutation counterparts are generally asymptotically as good as the best ones (16).

Objectives

The main objective of this paper is a hypothesis testing, in which the null hypothesis specifies no difference between groups, is an important tool in the assessment of new medical interventions. We have used the parametric and non-parametric tests for assessing the role of prognostic factors in comparison of two populations for the progression of the disease. Firstly likelihood ratio test is applied to compare (i) stage 2 and stage 3 (ii) stage 3 and stage 4 with respect to each factor because of the applicability of the test in every distribution (continuous and discrete). To apply the likelihood ratio test, an appropriate distribution based on AIC value is fitted to the available data set. Secondly, we applied the permutation test for the same population as an alternative tool to see the consistency of the result obtained from the likelihood ratio test. The correspondence between parametric and permutation test results gives confidence that the results are stable and reliable(17) and are able to identify the role of prognostic factors in the progression of disease stage wise. The test has been applied to a data set of 100 CKD outpatients with varying health conditions. The paper intends to evaluate the differences in the impact of prognostic factors for the progression of CKD into higher stages using parametric and non parametric tests.

We aimed to determine the factors other than the parameters as age, gender, race, and serum creatinine, used in computation of eGFR to assess the health of CKD patients with respect to their stage using the results of appropriate tests.

Patients and Methods

Likelihood ratio test

The prognostic factors for the progression of CKD under consideration are age, sex, diabetes, hypertension, serum creatinine, urea, hemoglobin, body mass index and albumin. Let, there are three sets or groups of data consisting of n_1, n_2 and n_3 observations corresponding to three stages of CKD like, stage 2, stage 3 and stage 4 respectively. The observations are denoted by X_{ijk} , i=1,2,...,n_j; j=1,2,3; k=1,2,...,9. i indexes the patient number, j indexes the group number and k indexes the prognostic factor (1 age, 2 sex, 3 diabetes, 4 hypertension, 5 serum creatinine, 6 urea, 7 hemoglobin, 8 body mass index, 9 albumin) X_{ijk} indicates the value of the kth factor for the ith patient of the jth group. The best fitted distribution is obtained by AIC value of the distributions for each factor. The probability function and log likelihood function of X_{ijk}'s under various distribution are given as

i) Weibull distribution (γ, λ)

$$f_j(x_{ijk}) = \lambda_j\gamma_j\ (\lambda_j x_{ijk})^{\gamma_j-1}\ \exp(-(\lambda_j x_{ijk})^{\gamma_j})\ ;$$

i= 1,2,...,n_j; j=1,2,3; k=1;

$$\lambda_j > 0,\ \gamma_j > 0,\ x_{ijk} > 0.$$

$$L_j(\gamma_j,\lambda_j) = n_j(\log\gamma_j + \gamma_j\log\lambda_j) + (\gamma_j\ -1)\sum_{i=1}^{n_j}\log x_{ijk} - \sum_{i=1}^{n_j}(\lambda_j x_{ijk})^{\gamma_j};$$

j=1,2,3 ; k=1

ii) Lognormal distribution

$$f_j(x_{ijk}) = \frac{1}{x_{ijk}\sigma_j\sqrt{2\pi}}\exp\left[\frac{-1}{2\sigma_j^2}(\log x_{ijk} - \mu_j)^2\right]$$; i=1,2,...,n_j; j=1,2,3; k= 5,6,8;

$$\mu_j > 0,\ \sigma_j > 0,\ x_{ijk} > 0.$$

$$L_j(\mu_j,\sigma_j) = \frac{-1}{2\sigma_j^2}\left[\frac{n_j}{(\sigma_j\sqrt{2\pi})} - \frac{\log\mu_j}{(\sigma_j\sqrt{2\pi})}\sum_{i=1}^{n_j}\frac{1}{x_{ijk}}\right]$$

iii) Gamma distribution

$$f_j(x_{ijk}) = \frac{\lambda_j^{\gamma_j}}{\Gamma\gamma_j}x_{ijk}^{\gamma_j-1}e^{-\lambda_j x_{ijk}};\ \text{i=1,2,...,n}_j;\ j=1,2,3;$$

$$k=7;\lambda_j>0,\gamma_j>0,\ x_{ijk}>0.$$

$$L_j(\gamma_j,\lambda_j) = n_j\gamma_j\log\lambda_j - n_j\log\Gamma\gamma_j + \sum_{i=1}^{n_j}(\gamma_j-1)\log x_{ijk} - \lambda_j\sum_{i=1}^{n_j}x_{ijk}$$

iv) Normal distribution

$$f_j(x_{ijk}\mid\mu_j,\sigma_j) = \frac{1}{\sigma_j\sqrt{2\pi}}\exp\left[\frac{-1}{2\sigma_j^2}(x_{ijk}-\mu_j)^2\right];$$

i=1,2,...,n_j; $j=1,2,3; k=9; \mu_j>0; \sigma_j>0; x_{ijk}>0.$

$$L_j(\mu_j,\sigma_j) = -n_j\log(\sigma_j\sqrt{2\pi}) - \sum_{i=1}^{n_j}\frac{1}{2\sigma_j^2}(x_{ijk}-\mu_j)^2$$

v) Bernoulli distribution

$$f_j(x_{ijk}) = p_j^{x_{ijk}}(1-p_j)^{1-x_{ijk}};\ i=1,2,...,n_j;$$

$$j=1,2,3; k=2,3,4; p_j>0; x_{ijk}>0.$$

$$L_j(p_j) = \sum_{i=1}^{n_j}x_{ijk}\log(p_j) + (n_j - \sum_{i=1}^{n_j}x_{ijk})\log(1-p_j)$$

We aimed to compare the two populations (i) group 1 and group 2 and (ii) group 2 and group 3 under each variable. One of the methods to accomplish the task is likelihood ratio test. In likelihood ratio test, the maximum likelihood estimates of unknown parameters are obtained by solving the likelihood equations simultaneously for each population. Likelihood equations are formed by partially differentiating the log likelihood function with respect to unknown parameters and are equated to zero. We also obtain the maximum likelihood estimates of unknown parameters of the combined population under null hypothesis.

For example the two Weibull populations corresponding to two groups (group 1 and group 2) can be compared using likelihood ratio test as;

$H_o: \lambda_1 = \lambda_2 = \lambda$ and $H_o: \gamma_1 = \gamma_2 = \gamma$. Where γ and λ are unknown.

The log likelihood of the combined group under H_0 is given by

$$L(\gamma,\lambda) = (n_1+n_2)(\log\gamma + \gamma\log\lambda) +$$
$$(\gamma\text{ -1})(\sum_{i=1}^{n_1}\log x_{i1k} + \sum_{i=1}^{n_2}\log x_{i2k})$$
$$-(\sum_{i=1}^{n_1}(\lambda x_{i1k})^{\gamma} + \sum_{i=1}^{n_2}(\lambda x_{i2k})^{\gamma})$$

The maximum likelihood estimates of unknown parameters are $\tilde{\lambda}_1$ and $\tilde{\gamma}_1$ for group 1, $\tilde{\lambda}_2$ and $\tilde{\gamma}_2$ for group 2 and $\tilde{\lambda}$ and $\tilde{\gamma}$ for combined group.

We compute the test statistic $X_L = 2(L_1(\tilde{\gamma}_1,\tilde{\lambda}_1) + L_2(\tilde{\gamma}_2,\tilde{\lambda}_2) - L(\tilde{\gamma},\tilde{\lambda}))$ and it follows chi-square distribution with two degrees of freedom.

Similarly the log likelihood of combined group under H_0 for lognormal distribution, gamma distribution, normal distribution and Bernoulli distribution are given by

$$L(\mu,\sigma) = \frac{-1}{2\sigma^2}\left[\frac{(n_1+n_2)}{(\sigma\sqrt{2\pi})} - \frac{\log\mu}{(\sigma\sqrt{2\pi})}(\sum_{i=1}^{n_1}\frac{1}{x_{i1k}} + \sum_{i=1}^{n_2}\frac{1}{x_{i2k}}\right]$$

$$L(\gamma,\lambda) = (n_1+n_2)\gamma\log\lambda - (n_1+n_2)\log$$
$$\overline{\gamma} + (\sum_{i=1}^{n_1}(\gamma-1)\log x_{i1k} + \sum_{i=1}^{n_2}(\gamma-1)\log x_{i2k})$$
$$-\lambda(\sum_{i=1}^{n_1}x_{i1k} + \sum_{i=1}^{n_2}x_{i2k})$$

$$L(\mu,\sigma) = -(n_1+n_2)\log(\sigma\sqrt{2\pi}) -$$
$$\sum_{i=1}^{n_1}\frac{1}{2\sigma^2}(x_{i1k}-\mu)^2 - \sum_{i=1}^{n_2}\frac{1}{2\sigma^2}(x_{i2k}-\mu)^2$$

$$L(p) = \sum_{i=1}^{n_1}x_{i1k}\log(p) + \sum_{i=1}^{n_2}x_{i2k}\log(p) +$$
$$((n_1 - \sum_{i=1}^{n_1}x_{i1k}) + (n_2 - \sum_{i=1}^{n_2}x_{i2k}))\log(1-p)$$

The likelihood test for the comparison of two populations under lognormal, gamma, normal and Bernoulli distributions can be applied as explained in Weibull distribution.

Permutation test

Our objective is to compare the two populations (i) group 1 and group 2 and (ii) group 2 and group 3 under each variable. When a parametric form of underlying distribution is not specified, permutation test also known as randomization test is widely used as practice in nonparametric statistics because of minimal assumptions and flexibility of the test statistic. Permutation tests can be applied to continuous, ordinal or categorical data from normal as well as non-normal distribution. Permutation methods can still be applied even when a parametric statistical method fails (18). There is a relaxation in choosing a test statistic.

Let there are n_1 observations of the prognostic factor corresponding to one stage and n_2 observations corresponding to other stage. The null hypothesis is that there is no significant difference between the effects of prognostic factor in two different stages. The only assumption in permutation test is that the distribution of prognostic factor under the null hypothesis is same in both the stages and being tested. Under the null hypothesis any particular permutation of the observations between the two stages has the same probability to occur as any other permutation. The steps involved for permutation test are summarized as follows:

Step 1: Compute the difference between the mean of the observed data corresponding to two stages and denote it by D_{obs}.

Step 2: Form a vector of $n_1 + n_2$ observations.

Step 3: Permute $n_1 + n_2$ observations between the two stages and assign n_2 observations to one stage and n_2 observations to other stage.

Step 4: Consider all possible permutations $\frac{(n_1+n_2)!}{n_1!n_2!}$ for small sample sizes and a pre determined number of random sample of permutations for large sample sizes.

Step 5: Calculate the difference between the mean of one stage and other stage for each permutation under consideration and denote it by D.

Step 6: For right tailed test P value is computed as the proportion of D greater than or equal to D_{obs}. The mathematical formula for computing P value is

$$P = \frac{Number\ of\ D's \geq D_{obs}}{\binom{n_1+n_2}{n_1}}$$

Step 7: Reject the null hypothesis at α level of significance if P value is less than or equal to α.

Depending upon the situation, one can choose the test statistic. Instead of difference of means the other test statistics can be difference of medians.

Ethical issues

The research followed the tenets of the Declaration of Helsinki.

Results

Data description

The data consists of 100 non-hospitalized CKD patients. This data comprise of three stages of CKD patients namely stage 2, stage 3 and stage 4 having 20, 37 and 43 number of patients respectively. The data corresponds to the variables age, sex, diabetes, hypertension, serum creatinine (SrCr), urea, hemoglobin (Hb), body mass index (BMI) and albumin (Alb) for each stage. Stage wise number of male patients and female patients along with diabetes and hypertension status are shown in Table 1. Table 2 shows the descriptive statistics corresponding to continuous variables for each stage.

Application of likelihood ratio test

To apply the likelihood ratio test we have computed the AIC values of all the distributions to select an appropriate distribution for each stage. For the variable serum creatinine, AIC values of normal, gamma, Weibull and lognormal distributions are 20.89593, 19.23063, 22.71633 and 18.67922 respectively for stage 2 of CKD patients. AIC value of lognormal distribution is found to be minimum. The same result holds for the stage 3 and stage 4 of CKD patients for the variable serum creatinine.

Table 1. Stage wise distribution of patients of chronic kidney disease with respect to sex, diabetes and hypertension

Stage	Sex		Diabetes		Hypertension	
	Male	Female	No	Yes	No	Yes
2	13	7	11	9	15	5
3	21	16	15	22	18	19
4	24	19	12	31	10	33

Hence, lognormal distribution is selected to represent the population of serum creatinine for stage 2, stage 3 and stage 4 of CKD patients. Table 3 shows the stage wise AIC value of each variable for every fitted distribution.

The paper intends to compare the two populations (i) stage 2 population and stage 3 population and (ii) stage 3 population and stage 4 population under each variable. Firstly we have compared stage 2 population and stage 3 population. Out of 57 cases, there were 20 patients in stage 2 and 37 patients were in stage 3. Available data has been used to test the equality of two stages under each variable. To find the likelihood ratio test statistic, and to test the hypothesis of equality of two populations, we have used the maximum likelihood estimators of unknown parameters of the distribution under stage 2 and stage 3 and combined groups. The respective log-likelihood values are computed and are shown in Table 4. The value of test statistic χ^2 for testing the scale parameters for stage 2 and stage 3 for serum creatinine is 22.0635 and corresponding P value is less than 0.001.

Secondly we have compared the stage 3 population and stage 4 populations. Out of 80 patients, there were 37 patients were in stage 3 and 43 patients were in stage 4. The maximum likelihood estimators of unknown parameters of the distribution under stage 3 and stage 4 and combined groups are used to compute the respective log-likelihood values under each variable. The value of test statistic χ^2 for testing the scale parameters for stage 3 and stage 4 for serum creatinine is 64.8697 and corresponding P value is less than 0.001. The value of the maximum likelihood estimates of the unknown parameters, the log likelihood hood values, test statistic under null hypothesis and P value of the test for the first and second case for each variable are summarized in Table 4 and Table 5.

Permutation test

There are two broad approaches to compare the

Table 2. Descriptive statistics (stage wise) of variables age, BMI, Hb, urea, SrCr and albumin

Stage		Age	BMI	Hb	Urea	SrCr	Alb
	N	20	20	20	20	20	20
	Mean	38.35	23.74	12.35	54.70	1.685	3.88
2	SD	15.069	5.059	1.478	15.051	0.3787	0.701
	Minimum	18	17	10	38	1.2	2
	Maximum	67	35	17	86	2.5	6
	N	37	37	37	37	37	37
	Mean	47.49	23.93	10.83	76.95	2.473	3.50
3	SD	13.291	5.133	1.800	20.576	0.7475	0.708
	Minimum	22	12	7	39	1.3	2
	Maximum	72	36	14	145	4.1	5
	N	43	43	43	43	43	43
	Mean	53.35	23.98	9.23	113.28	4.195	3.14
4	SD	12.958	4.230	2.036	23.807	0.9097	0.527
	Minimum	25	16	6	66	2.3	2
	Maximum	78	33	13	158	6.2	5

Table 3. AIC values of the fitted distribution stage wise for the prognostic factors and selected distribution for SrCr, urea, alb, Hb, BMI and age

Variable	Stage	AIC value				Selected distribution
		Normal	Gamma	Weibull	Lognormal	
SrCr	Stage-2	20.89593	19.23063	22.21633	18.67922	Lognormal
	Stage-3	86.44799	84.0101	86.42375	84.00627	
	Stage-4	116.8801	118.549	120.2608	116.514	
Urea	Stage-2	168.1898	165.2769	169.109	164.2254	Lognormal
	Stage-3	331.7742	328.0411	334.4291	327.5492	
	Stage-4	401.9335	399.9594	401.9335	397.6355	
Alb	Stage-2	45.52367	45.90034	47.00497	46.49061	normal
	Stage-3	81.4023	83.88597	81.27593	85.2095	
	Stage-4	66.27714	66.27714	77.03078	69.84866	
Hb	Stage-2	75.3517	74.06572	80.20777	74.36572	Gamma
	Stage-3	152.3803	151.3157	151.309	153.2351	
	Stage-4	186.1776	185.8815	186.7835	186.4983	
BMI	Stage-2	124.5779	122.9229	126.2421	122.3701	Lognormal
	Stage-3	229.8226	229.3306	230.4627	229.4794	
	Stage-4	249.0484	247.1683	252.5318	246.6791	
Age	Stage-2	168.2379	166.5566	166.5727	167.1403	Weibull
	Stage-3	299.4311	300.8825	298.4062	302.7371	
	Stage-4	345.3216	346.8957	344.7474	348.8308	

Table 4. Maximum likelihood estimates of parameters and log likelihood values of fitted distribution stage wise, test statistic and *P* value for SrCr, urea, Alb, Hb, BMI and age

Variable name	Stages	Estimate of Parameters	-2* Loglikelihood value	Test statistic	*P* value
SrCr Lognormal(logμ, logσ)	2	(0.49882 , 0.21208)	14.6792		
	3	(0.86058 , 0.30169)	80.0063		
	com 1 (2 and 3)	(0.73365,0.32353)	116.7490	22.0635	0.00002
	4	(1.40909 , 0.22851)	112.5140		
	com 2 (3 and 4)	(1.15541,0.38073)	257.3900	64.8697	<0.00001
Urea Lognormal(logμ, logσ)	2	(3.96902, 0.25099)	160.2254		
	3	(4.30986, 0.25756)	323.5492		
	com 1 (2 and 3)	(4.19027, 0.30269)	503.2158	19.4412	0.00006
	4	(4.70617, 0.22358)	393.6355		
	com 2 (3 and 4)	(4.52288, 0.31080)	763.7111	46.5264	<0.00001
Alb Normal(μ, σ)	2	(3.87500, 0.68328)	41.9003		
	3	(3.49730, 0.69806)	77.2759		
	com 1 (2 and 3)	(3.62982, 0.71597)	125.6090	6.4328	0.0401
	4	(3.13953, 0.52035)	62.2771		
	com 2 (3 and 4)	(3.30500, 0.63461)	154.2700	14.7169	0.00064
Hb Gamma(ϒ, λ)	2	(76.42817, 0.161524)	70.3855		
	3	(36.18532, 0.299245)	147.4957		
	com 1 (2 and 3)	(37.40508, 0.30373)	231.3401	13.4589	0.00120
	4	(20.53472, 0.44937)	181.8815		
	com 2 (3 and 4)	(21.96165, 0.45392)	345.3441	15.9669	0.00034
BMI Lognormal(logμ, logσ)	2	(3.14674, 0.20060)	118.3701		
	3	(3.15165, 0.22090)	226.4794		
	com 1 (2 and 3)	(3.14993, 0.21401)	345.0926	0.2431	0.88555
	4	(3.16225, 0.17216)	242.6791		
	com 2 (3and 4)	(3.15735, 0.19628)	471.6933	2.5348	0.28156
Age weibull(ϒ, λ)	2	(2.85667, 43.18375)	162.6727		
	3	(4.16469, 52.39949)	294.4062		
	com 1 (2 and 3)	(3.49119, 49.36167)	463.7557	6.6768	0.03549
	4	(4.74862, 58.34459)	340.7474		
	com 2 (3and 4)	(4.36044, 55.67909)	642.7780	7.6224	0.0221

Table 5. Maximum likelihood estimates of parameters and log likelihood values of fitted distribution stage wise, test statistic and *P* value for categorical prognostic factors diabetes, sex and hypertension

Variable	Stage	Estimate of parameter	-2* Log likelihood value	Test Statistic	*P* value
Diabetes Binom(1,p)	2	0.45	27.5256		
	3	0.5946	49.9606		
	Com 1 (2 and 3)	0.5439	78.5796	1.0935	0.2957
	4	0.7209	50.9182		
	Com 2 (3 and 4)	0.6625	102.2981	1.4191	0.2335
Sex Binom(1,p)	2	0.35	25.8979		
	3	0.4324	50.6151		
	Com 1 (2 and 3)	0.4035	76.8826	0.3696	0.5432
	4	0.4419	59.0279		
	Com 2 (3 and 4)	0.04375	109.6503	0.0072	0.9324
Hypertension (HTN) Binom(1,p)	2	0.25	20.0161		
	3	0.5135	49.9606		
	Com 1 (2 and 3)	0.4210	72.5626	2.5860	0.1078
	4	0.7674	46.642		
	Com 2 (3 and 4)	0.65	107.6818	11.0793	0.0009

populations among two groups: non-parametric and parametric. The objective of the paper is to examine the progression of disease stage wise under the influence of various prognostic factors. The permutation test as a non parametric test has been applied to compare the two populations (i) sage 2 and stage 3 populations and (ii) stage 3 and stage 4 populations for each variable.

Firstly we have compared the two populations of stage 2 and stage 3 groups. There are 20 observations under stage 2 and 37 observations under stage 3. Mean difference between the observations of stage 2 and stage 3 are obtained. We have also computed the mean difference between two sets of observations after permuting the 57 observations into two stages consisting of 20 observations in stage 2 and 37 observations in stage 3. The value of test statistic using the mean differences of original sets of data and permuted sets of data under two groups is computed for each variable and corresponding *P* value of the test is obtained. The mean difference of original sets of observations, the value of test statistic and *P* value are shown in Table 6. The similar procedure has been adopted for comparison of stage 3 population and stage 4 population. The mean difference of original sets of observations, the value of test statistic and *P* value for permutation test are shown in Table 6 and Table 7.

Simulation study

A simulation study has been carried out to determine the power of permutation test and likelihood ratio test testing the significance of difference between the values of the parameters of the distribution of prognostic factors involved in the progression of CKD in stage 2, stage 3 and stage 4. Firstly, we have compared the scale parameters of the fitted distribution for stage 2 and stage 3. We have simulated a sample of 500 observations 200 times with the values of the parameters of the fitted distribution for each prognostic factor of stage 2 of the disease. Another set of 500 observations are also generated 200 times with an increment s1 in the value of the parameter of the underlying distribution. We have taken ten different values of s1 in each case. Permutation test and likelihood test are applied and power of the test is computed each time. The same procedure is applied for comparison of the scale parameters of the fitted distribution for stage 3 and stage 4. The results of the simulation study showing the distribution of the prognostic factor, values of the parameters, value of the increment (s1), power of the permutation test and likelihood ratio test involving (i) stage 2 and stage 3 and (ii) stage 3 and stage 4 are summarized in Table 8 and Table 9.

Power curves of the permutation test and likelihood

Table 6. Mean difference, test statistic, p-value for the prognostic factors age, BMI, Hb, urea, SrCr and Alb

	Stage 2 and Stage 3			Stage 3 and Stage 4		
Variable	Mean difference	Test statistic	*P* value	Mean difference	Test statistic	*P* value
Age	-9.1365	-2.2719	0.02309	-5.862351	-1.9573	0.05032
BMI	-0.1867	-0.1329	0.8943	-0.048874	-0.0469	0.9625
Hb	1.51527	2.98	0.002883	1.601823	3.4339	0.000595
Urea	-22.246	-3.722	0.0001976	-36.33312	-5.6357	<0.00001
SrCr	-0.788	-3.8222	0.0001323	-1.722376	-6.3977	<.00001
Alb	0.3777	1.884	0.05956	0.3577624	2.4983	0.01248

Table 7. Mean difference and *P* value for the categorical variables diabetes, sex and hypertension

	Stage 2 and Stage 3		Stage 3 and Stage 4	
Variable	**Mean difference**	***P* value**	**Mean difference**	***P* value**
Diabetes	-0.1446	0.4427	-0.1226	0.3399
Sex	-0.0824	0.7514	-0.0094	1
Hypertension	-0.2054	0.1992	-0.3620	0.002

Table 8. Power of the permutation test and likelihood ratio test associated with stage 2 and stage 3.

Variable	**Power**									
Serum creatinine lnorm(logμ=p1,logσ=p2) lnorm(logμ=p3,logσ=p2) p1=0.4988 p2=0.2569,p3=p1+s1	Increment (s1)	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	Power-permutation	0.11	0.25	0.43	0.685	0.835	0.935	0.995	0.995	1	1
	Power-likelihood	0.095	0.185	0.415	0.635	0.79	0.895	0.91	0.98	0.995	1
Urea lnorm(logμ=p1,logσ=p2) lnorm(logμ=p3,logσ=p2) p1=3.969 p2=0.2543, p3=p1+s1	Increment(s1)	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	Power-permutation	0.115	0.24	0.455	0.69	0.86	0.965	0.995	1	1	1
	Power-likelihood	0.085	0.185	0.35	0.61	0.79	0.925	0.98	0.99	1	1
Body mass index lnorm(logμ=p1,logσ=p2) lnorm(logμ=p3,logσ=p2) p1=3.1467 p2=0.2108, p3=p1+s1	Increment	0.01	0.02	0.025	0.03	0.035	0.04	0.045	0.05	0.055	0.06
	Power-permutation	0.1	0.275	0.49	0.645	0.745	0.87	0.925	0.965	0.975	0.99
	Power-likelihood	0.09	0.195	0.425	0.545	0.635	0.775	0.855	0.935	0.965	0.98
Hemoglobin Gamma(ϒ=p1,λ=p2) Gamma(ϒ=p1,λ=p3) p1=56.3068 p2=6.191 p3=p2-s1	Decrement(s1)	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24
	Power-permutation	0.13	0.325	0.455	0.625	0.73	0.91	0.92	0.98	0.985	0.995
	Power-likelihood	0.1	0.3	0.33	0.495	0.64	0.845	0.875	0.945	0.97	0.98
Albumin Normal(μ=p1,σ=p2) Normal(μ=p3,σ=p2) p1=3.875 p2=0.6907, p3=p1-s1	Decrement(s1)	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
	Power-permutation	0.06	0.18	0.275	0.43	0.655	0.8	0.865	0.95	0.98	1
	Power-likelihood	0.065	0.125	0.21	0.35	0.55	0.685	0.79	0.89	0.965	0.98
Age Weibull(ϒ=p1,λ=p2) Weibull(ϒ=p1,λ=p3) p1=3.5107 p2=43.1838 p3=p2+s1	Increment	0.1	0.2	0.5	1	1.5	2	2.5	3	3.5	4
	Power-permutation	0.045	0.05	0.125	0.215	0.415	0.645	0.81	0.905	0.97	0.99
	Power-likelihood	0.06	0.065	0.11	0.225	0.35	0.595	0.805	0.885	0.975	0.995
Diabetes binom(1,p1) binom(1,p2) p1=0.45, p2=p1+s1	Increment	0.03	0.04	0.05	0.08	0.09	0.1	0.11	0.12	0.13	0.14
	Power-permutation	0.155	0.19	0.39	0.69	0.805	0.855	0.945	0.96	0.98	0.995
	Power-likelihood	0.07	0.16	0.395	0.69	0.76	0.805	0.88	0.91	0.92	0.945
Sex binom(1,p1) binom(1,p2) p1=0.35 p2=p1+s1	Increment	0.02	0.04	0.06	0.08	0.1	0.11	0.12	0.13	0.14	0.15
	Power-permutation	0.06	0.285	0.475	0.725	0.885	0.955	0.97	0.99	0.995	1
	Power-likelihood	0.005	0.25	0.515	0.72	0.865	0.905	0.93	0.97	0.97	0.985
Hypertension binom(1,p1) binom(1,p2) p1=0.25 p2=p1+s1	Increment	0.02	0.04	0.06	0.08	0.1	0.11	0.12	0.13	0.14	0.15
	Power-permutation	0.1	0.215	0.5	0.78	0.93	0.935	0.965	0.99	0.995	1
	Power-likelihood	0.01	0.205	0.51	0.775	0.9	0.91	0.92	0.955	0.965	0.975

Table 9. Power of the permutation test and likelihood ratio test associated with stage 2 and stage 3

Variable	Power										
Serum creatinine lnorm(logμ=p1,logσ=p2) lnorm(logμ=p3,logσ=p2) p1=0.8606 p2=0.0.2406, p3=p1+s1	Increment (s1)	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	Power-permutation	0.085	0.275	0.435	0.645	0.835	0.935	0.99	0.995	1	1
	Power-likelihood	0.065	0.24	0.31	0.565	0.79	0.905	0.98	0.995	1	1
Urea lnorm(logμ=p1,logσ=p2) lnorm(logμ=p3,logσ=p2) p1=4.3099 p2=0.2543, p3=p1+s1	Increment(s1)	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	Power-permutation	0.085	0.215	0.505	0.71	0.9	0.96	0.995	1	1	1
	Power-likelihood	0.09	0.18	0.4	0.59	0.86	0.91	0.995	1	1	1
Body mass index lnorm(logμ=p1,logσ=p2) lnorm(logμ=p3,logσ=p2) p1=3.1517 p2=0.1966, p3=p1+s1	Increment	0.01	0.02	0.025	0.03	0.035	0.04	0.045	0.05	0.055	0.06
	Power-permutation	0.075	0.175	0.51	0.495	0.78	0.795	0.935	0.96	0.99	0.995
	Power-likelihood	0.06	0.17	0.445	0.43	0.705	0.675	0.895	0.9	0.965	0.98
Hemoglobin Gamma(ϒ=p1,λ=p2) Gamma(ϒ=p1,λ=p3) p1=28.36 p2=3.3417 p3=p2-s1	Decrement(s1)	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
	Power-permutation	0.085	0.14	0.275	0.535	0.745	0.9	0.95	0.98	0.995	1
	Power-likelihood	0.075	0.095	0.235	0.435	0.63	0.82	0.915	0.955	0.995	0.995
Albumin Normal(μ=p1,σ=p2) Normal(μ=p3,σ=p2) p1=3.4973 p2=0.6093 p3=p1-s1	Decrement(s1)	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2
	Power-permutation	0.08	0.18	0.34	0.515	0.735	0.87	0.95	0.995	0.995	1
	Power-likelihood	0.065	0.17	0.25	0.4	0.625	0.79	0.9	0.975	0.99	1
Age Weibull(ϒ=p1,λ=p2) Weibull(ϒ=p1,λ=p3) p1=4.4567 p2=52.3995 p3=p2+s1	Increment	0.1	0.2	0.5	1	1.5	2	2.5	3	3.5	4
	Power-permutation	0.06	0.085	0.1	0.21	0.44	0.605	0.805	0.94	0.97	0.99
	Power-likelihood	0.065	0.065	0.085	0.19	0.425	0.605	0.845	0.96	0.995	0.99
Diabetes binom(1,p1) binom(1,p2) p1=0.5946, p2=p1+s1	Increment	0.03	0.04	0.05	0.08	0.09	0.1	0.11	0.12	0.13	0.14
	Power-permutation	0.15	0.22	0.37	0.44	0.825	0.915	0.96	0.975	0.99	0.995
	Power-likelihood	0.06	0.195	0.38	0.465	0.8	0.89	0.93	0.94	0.965	0.965
Sex binom(1,p1) binom(1,p2) p1=4324 p2=p1+s1	Increment	0.02	0.04	0.06	0.08	0.1	0.11	0.12	0.13	0.14	0.15
	Power-permutation	0.105	0.24	0.55	0.7	0.91	0.94	0.96	0.99	1	1
	Power-likelihood	0	0.21	0.55	0.685	0.855	0.89	0.91	0.965	0.96	0.985
Hypertension binom(1,p1) binom(1,p2) p1=0.5135 p2=p1+s1	Increment	0.02	0.04	0.06	0.08	0.1	0.11	0.12	0.13	0.14	0.15
	Power-permutation	0.1	0.265	0.45	0.71	0.895	0.935	0.97	0.995	0.995	1
	Power-likelihood	0.005	0.215	0.475	0.71	0.85	0.87	0.925	0.965	0.975	0.975

ratio test have been plotted for different values of the increment (s1) in the values of the parameters are shown in Figure 1 and Figure 2, for each variable.

Discussion

CKD is recognized as having changed from a subspecialty issue to a global health concern (1). Identification of factors predisposing an individual to CKD is important not only in terms of personal but also from community health point of view as some risk factors can be modified and thereby preventing or slow downing the progression of disease to end stage renal disease. Economic burden of disease can be reduced by early intervention

According to NKF, eGFR is computed on the basis of age, sex, and race and also serum creatinine. Likelihood ratio test though very commonly used test but is highly

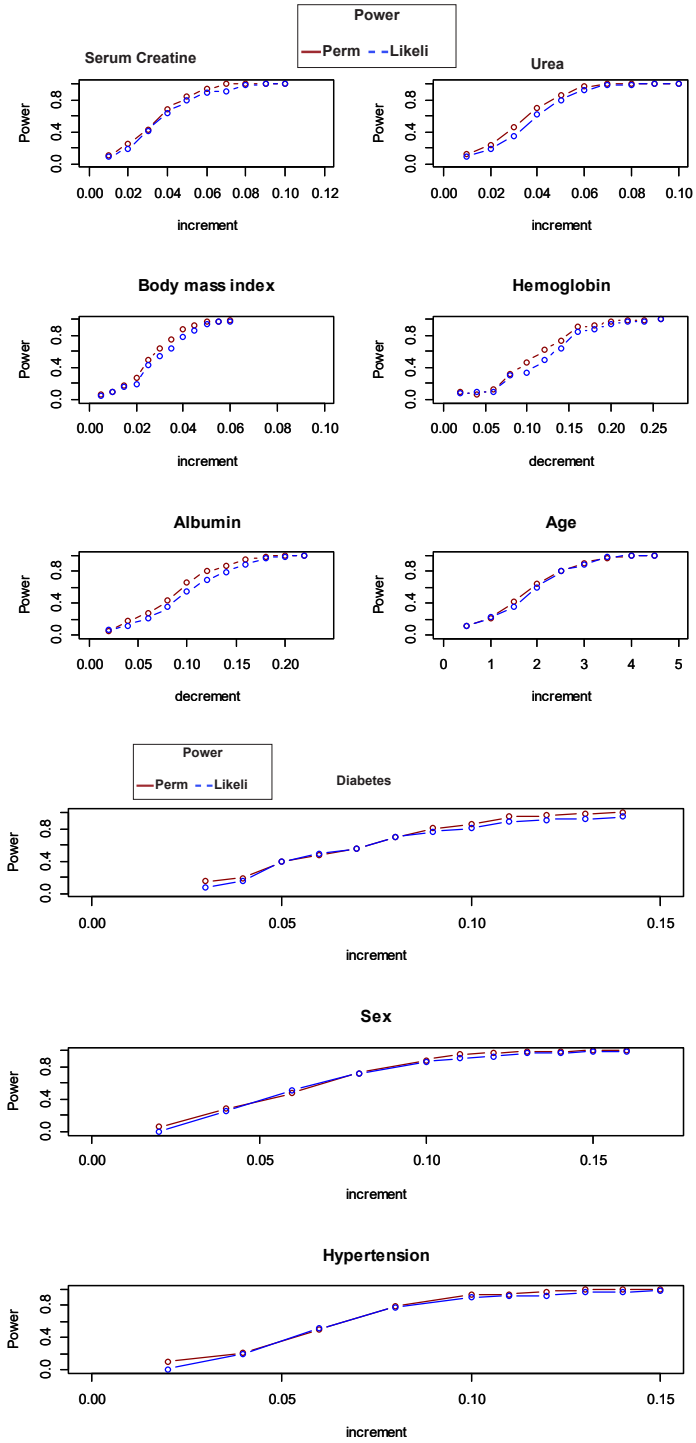

Figure 1. Power curves associated with permutation test and likelihood ratio test for stage 2 and stage 3 for the prognostic factors serum creatinine, urea, body mass index, hemoglobin, albumin, age, diabetes, sex and hypertension.

recommended and is applicable for both continuous and discrete distribution. In practice, parametric methods require a set of stringent assumptions which are difficult to justify and are quite unrealistic and unclear and seldom met. In such situations, nonparametric tests are more appropriate than parametric counterparts. The relative risk of efficiency of nonparametric method is very small under the applicability of both the tests (19). The permutation tests are based on more realistic foundations, are intrinsically robust with credible inferences. Permutation tests are successful in many cases where parametric tests fail. As a rule, the test involving fewer assumptions and

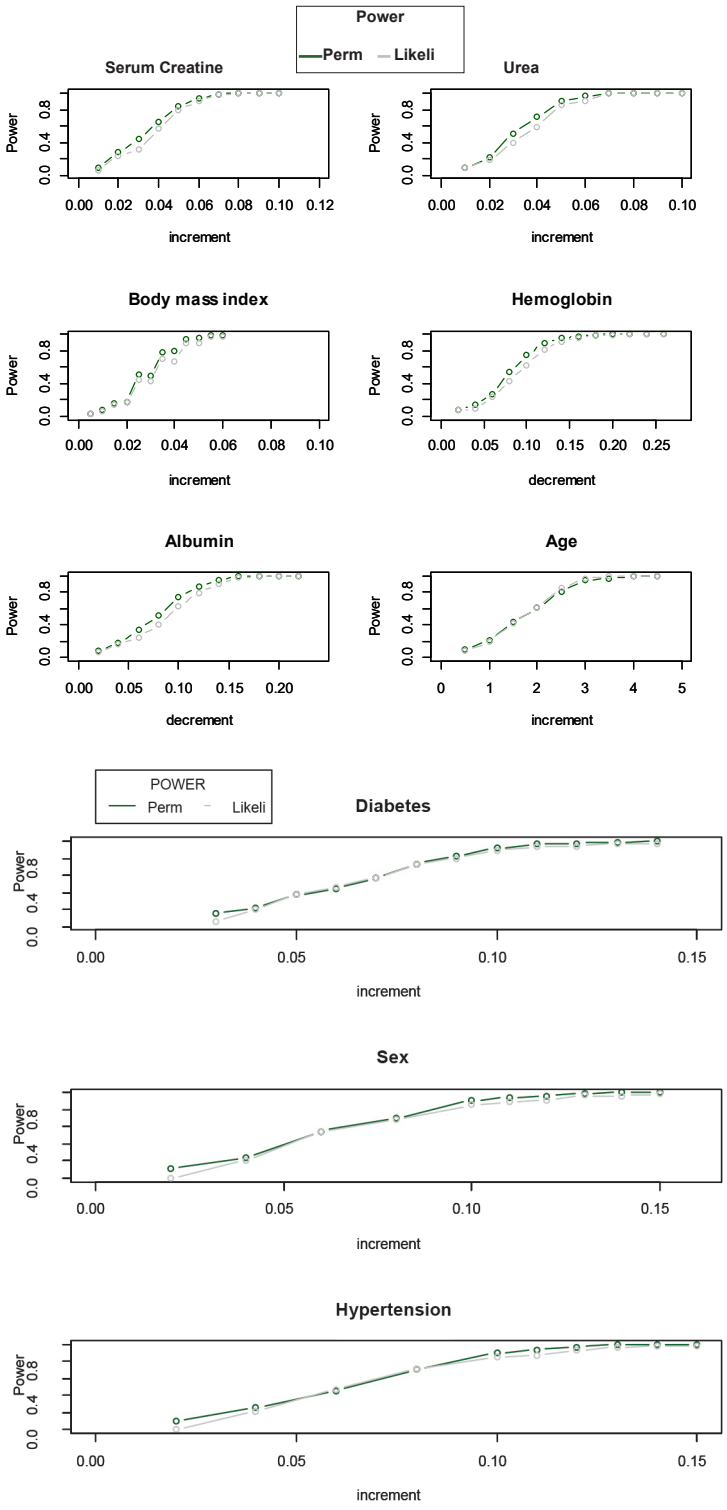

Figure 2. Power curves associated with permutation test and likelihood ratio test for stage 3 and stage 4 for the prognostic factors serum creatinine, urea, body mass index, hemoglobin, albumin, age, diabetes, sex and hypertension.

limitations will have wider applications. Permutation tests require minimum conditions and provide flexibility in selection of test statistics and are independent of distribution. They are widely used test. Permutation test falls under the conditional method of inference and can be treated a test of an exact nonparametric nature in a conditional context. Under the hypothesis of no difference for underlying population distribution, the condition is being made on the pooled observed data in the form of a set of sufficient statistics (20).

Permutation test and likelihood ratio tests suggest that serum creatinine and age are significant factors for the

progression of CKD in both from (i) stage 2 to stage 3 and (ii) stage 3 to stage 4. The result has been obtained from our data set and is consistent with the simulation study as well. The power of the permutation test and likelihood ratio test is 1 from the simulation data and has been shown in Table 8 and Table 9.

Sex of a patient is a non significant factor for the progression of the disease from (i) stage 2 to stage 3 and (ii) stage 3 to stage 4 as revealed by permutation test and likelihood ratio test applied on our data set. However, power table of simulation study indicates that the power of both permutation test and likelihood ratio test is approximately 0.72 for the comparison of stage 2 and stage 3 and is approximately 0.1 for the stage 3 and stage 4. The power table of simulation study also suggests that if the difference between parameter is 0.13 for the comparison of both (i) stage 2 and stage 3 and (ii) stage 3 and stage 4 then sex of a patient is a significant factor for the progression of CKD and power of the test in such cases is approximately 0.99.

We have not considered the factor race of the patient as data of the CKD patients have been collected from the India only.

Apart from the factors considered for the computation of eGFR, some other factors are also responsible for the progression of the disease into severe stages. In our study the factors urea. Hemoglobin and albumin are also coming out to be significant factors for the progression of disease in both (i) stage 2 to stage 3 and (ii) stage 3 to stage 4 by both permutation test and likelihood ratio test. Simulation study also supports this fact as the power associated with the tests is almost unity.

Permutation test and likelihood ratio test based on our data set suggest that body mass index is not a significant factor for the progression of disease towards severity. The simulation study suggests the powers of the permutation test and likelihood ratio test corresponding to the estimated scale parameters of the data set are 0.1 and 0.09 respectively for the comparison of stage 2 and stage 3 and are 0.075 and 0.06 for stage 3 and stage 4. However simulation study also suggests that if the difference between the scale parameters are .06 for (i) stage 2 and stage 3 and (ii) stage 3 and stage 4 then body mass index will be a significant factor the progression of the disease and the power of the test in this case will be approximately 0.99.

Diabetes is a non significant factor as suggested by both permutation test and likelihood ratio test based on the data set for both (i) stage 2 and stage 3 and (ii) stage 3 and stage 4. However the simulation study result is totally in contrast with the data set result. The simulation study based on power of the permutation test and likelihood ratio test suggests that diabetes is a significant factor for the progression of disease from (i) stage 2 to stage 3 and (ii) stage 3 to stage 4. One of the possible reason can be data set is too small and is a single sample where as in simulation study the data set is quite large and the number of sample sets are also very large.

Hyper tension is coming out to be a significant factor for the progression of disease from stage 3 to stage 4 but is a non significant factor for the progression from stage 2 to stage 3. However, according to simulation study hyper tension is a significant factor for the progression of disease. The power associated with the permutation test and likelihood ratio test is almost one.

Conclusion

There exists a statistical significant relationship between the prognostic factors serum creatinine, body mass index, urea, hemoglobin, albumin, age, sex, diabetes, hypertension and progression of CKD to higher stages.

Limitations of the study

The total number of subjects under the present study is small which further reduces the size of the sample in each stage of CKD. Most of the patients are from Delhi and its surrounding areas only. We suggest to investigate the other possible risk factors such as family history, socio-economic status, smoking and nephrotoxins (alcohol and recreational drugs).

Authors' contribution

All the authors jointly conceived and designed the analysis. AS and SK have collected the data and performed the analysis. SK has drafted the final manuscript. All the authors read, revised and approved the final manuscript.

Conflicts of interest

The authors declare no conflict of interest.

Ethical considerations

Ethical issues (including plagiarism, data fabrication, double publication) have been completely observed by the authors.

Funding/Support

The present research work is not funded or supported by any agency or an organization.

References

1. Kazancioğlu R. Risk factors for chronic kidney disease: an update. Kidney Int Suppl (2011). 2013;3:368-371. doi: 10.1038/kisup.2013.79.
2. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, et al. Prevalence of chronic kidney disease in the United States. JAMA. 2007;298:2038-47. doi: 10.1001/jama.298.17.2038.
3. Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, et al. Global Prevalence of Chronic Kidney Disease - A Systematic Review and Meta-Analysis. PLoS One. 2016;11:e0158765. doi: 10.1371/journal.pone.0158765.
4. Hsu JY, Roy JA, Xie D, Yang W, Shou H, Anderson AH, et

al. Statistical methods for cohort studies of CKD: survival analysis in the setting of competing risks. Clin J Am Soc Nephrol. 2017;12:1181-9. doi: 10.2215/CJN.10301016.
5. Roy J, Shou H, Xie D, Hsu JY, Yang W, Anderson AH, et al. Statistical methods for cohort studies of CKD: prediction modeling. Clin J Am Soc Nephrol. 2017;12:1010-1017. doi: 10.2215/CJN.06210616.
6. David HA.The beginnings of randomization tests. Am Stat. 2008; 62:70-2. doi: 10.1198/000313008X269576.
7. Agresti A, Klingenberg B. Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. Journal of the Royal Statistical Society: Series C (Applied Statistics).2005; 54: 691-706.
8. Tang L, Duan N, Klap R, Asarnow, JR, Belin, TR. Applying permutation tests with adjustment for covariates and attrition weights to randomized trials of health-services interventions. Stat Med. 2009; 28: 65-74. doi: 10.1002/sim.3453.
9. Cox DD, Lee JS. Pointwise testing with functional data using the Westfall–Young randomization method. Biometrika. 2008;95:621-34. doi: 10.1093/biomet/asn021.
10. Downer R. A permutation alternative and other test procedures for spatially correlated data in one-way Anova. J Stat Comput Simul. 2002;72:747-57. doi: 10.1080/00949650214268.
11. Dray S. On the number of principal components: A test of dimensionality based on measurements of similarity between matrices. Comput Stat Data Anal. 2008;52:2228-37. doi:10.1016/j.csda.2007.07.015.
12. Brombin C, Salmaso L. Multi-aspect permutation tests in shape analysis with small sample size. Comput Stat Data Anal. 2009;53:3921-31. doi: 10.1016/j.csda.2009.05.010.
13. Abd-Elfattah EF, Butler RW. The weighted log-rank class of permutation tests: p-values and confidence intervals using saddlepoint methods. Biometrika. 2007;94:543-51.
14. Galimberti S, Valsecchi MG. Multivariate permutation test to compare survival curves for matched data. BMC Med Res Methodol. 2013;13:16. doi: 10.1186/1471-2288-13-16.
15. Giancristofaro RA, Bonnini S. Permutation tests for heterogeneity comparisons in presence of categorical variables with application to university evaluation. Metodoloski Zvezki. 2007;4:21.
16. Hoeffding W. The large-sample power of tests based on permutations of observations. The Annals of Mathematical Statistics. 1952;23:169-92.
17. Ottoboni K, Lewis F, Salmaso L. An empirical comparison of parametric and permutation tests for regression analysis of randomized experiments. Stat Biopharm Res. 2018. doi: 10.1080/19466315.2018.1458648 .
18. Good P. Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer Science & Business Media; 2013.
19. Lehmann EL. Parametric versus nonparametrics: two alternative methodologies. J Nonparametr Stat. 2009;21:397-405. doi: 10.1080/10485250902842727.
20. Pesarin F, Salmaso L. Permutation Tests for Complex Data: Theory, Applications, and Software. West Sussex, UK: John Wiley & Sons, Ltd; 2010.