

Research on AIDS Therapy Based on Linear Interpolation Fitting

Yibo Wang¹, Zhiming Wang², Rui Jin³

¹Xi'an Gaoxin NO.1 High School, Xi'an, Shaanxi, 710075, China.

²Jilin University, Mathematics College, Changchun, Jilin, 130012, China.

³Jiangsu University of Science and Technology School of Science, Zhenjiang, Jiangsu, 212003, China.

Keywords: spss, AIDS therapy, nonlinear curve fitting, linear interpolation.

Abstract: This article uses the AIDS data published by the United States AIDS Medical Trial Institute, and uses linear interpolation to uniformly simulate the CD4 cell count and HIV concentration in the patient's weekly body. The effect of cell count is approximately a cubic curve. With reference to the HIV fitting curve, it was found that the patient should terminate the therapy at about 27 weeks. The respondent took different drug combinations and filled the data with uniform interpolation. CD4 cell count time series data, after SPSS fitting, Kruskal-Wallis H Test and median test method were used to fit a better curative effect scheme. Using only CD4 cell count as the standard, the efficacy of the four therapies is : The fourth treatment is the best, the third treatment is the worst, the second treatment is the worst, and the first treatment is the worst. Patients taking the third treatment should stop taking the medicine at 25 weeks, while the best taking the fourth treatment should stop at the 19 week.

1. Research background

The full name of AIDS in medicine is "Acquired Immune Deficiency Syndrome", or AIDS in English for short. It is caused by HIV (medical full name "Human Immunodeficiency Virus", in English for HIV) [1]. One of the worst plagues in society, it has consumed nearly 30 million lives in more than 20 years since its discovery in 1981. Medical inclusion criteria clinically rely mainly on physical examination and medical history, laboratory standards (CD4 <200 / mm³), Mainly rely on the CD4 cell count, if CD4 detection is not available, you can also detect the total lymphocyte count (TCL <1200 / mm³). CD4 cell count is better than total lymphocyte count, if conditions permit, should use CD4 cell count as much as possible ([2]).

The cd4 cells of the human immune system play an important role in resisting the invasion of HIV. When CD4 is lysed by HIV infection, its number will decrease sharply, and HIV will increase rapidly, leading to aids. The purpose of AIDS treatment is to minimize humans. The amount of HIV in the body, while producing more cd4, must at least effectively reduce the rate of cd4 reduction in order to improve the body's immune capacity. However, to date, there are only a few recognized drugs that can inhibit and delay its course, but no Special drugs for curing AIDS. Some current aids treatments not only have side effects on the human body, but also have high costs. Many countries and medical organizations are actively testing and looking for better aids treatments.

2. Research content

The data sources studied in this article are mainly based on two sets of data published by actg, the American AIDS Medical Trial Agency. The first set of data is more than 300 patients taking zidovudine, lamivudine, and indinavir at the same time. Every few weeks Cd4 cell count and hiv concentration tested (per μ l The amount in the blood); the second group is to randomly divide more than 1,300 patients into 4 groups, each group taking one of the following 4 treatments, taking CD4 concentration tested about every 8 weeks. 4 treatments The daily medicines are: 600 mg zidovudine or 400 mg deshydroxyglycoside, these two drugs are used on a monthly basis; 600 mg zidovudine

plus 2.25 mg zalcitabine; 600 mg zidovudine plus 400 mg Hydroxy glycosides; 600 mg zidovudine plus 400 mg dehydroxyglycoside, and then 400 mg nevirapine.

Based on two sets of data published by actg, the American AIDS Medical Trial Agency, predict the effect of continuing treatment, or determine the optimal treatment termination time.

Evaluate the advantages and disadvantages of the 4 therapies (only using cd4 as the standard), and predict the effect of continuing treatment for the better therapy, or determine the optimal treatment termination time.

3. Mathematical models for AIDS therapy

3.1 AIDS treatment effect model based on data linear interpolation

Among the test data of more than three hundred testees in the past year, due to the sampling test data, there are discontinuous features, and from the perspective of the week, the number of test times is too small to be suitable for further prediction. Therefore, linear interpolation was used to uniformly simulate the cd4 cell count and the concentration of hiv [3] in the body of each respondent each week. Here we assume that the concentrations of cd4 and hiv in the respondent's body are uniformly changed. , Average the cd4 and hiv content of all the respondents every week. In addition, due to some objective reasons, some data are distorted or have no statistical information value, such as the existence of -2, -1 weeks of data, some Respondents only checked the data once, and discarded 5% of data with no information value. Finally, vc ++ was used to do a linear fit to make up the difference. Then calculate the cd4 and hiv of all the respondents each week. The average value of the concentration, so we can get a more complete data from the time series perspective of the therapy. Finally, make a mean scatter plot, use spss to make the best model simulation curve, and get the therapy The effect on human cd4 cell count is approximately cubic curve, The effect on the concentration of human HIV can be better simulated by comprehensively using the s-curve and cubic curve.

After removing the singular data and the lack of ascent value from the cd4 count and hiv concentration data of the original 355 patients, they are sequentially from 1 to n_0 Renumbered as $1, 2, \dots, n_0$ We assume that k Week is at w_{ik} Times and w_{ik+1} Between times, the linear interpolation model is established as follows:

$$y_{ik} = x_{ik} + \frac{x_{ik} - x_{ik+1}}{w_{ik+1} - w_{ik}} \times (k - w_{ik}) \quad , \quad (i = 1, 2, \dots, 355 \quad k = 0, 1, \dots, n_i) \quad (1)$$

First calculate the average cd4 and hiv content of all the testees every week, and use spss to simulate the best fit curve equation of cd4 cell count change:

$$Y_1 = 100.34237 + 7.19105X - 0.18781X^2 + 0.00135X^3 \quad (2)$$

We use the curve fitting graph simulated by spss. From the intuitive and statistical inspection indicators, we can see that the effect of fitting and prediction is ideal.

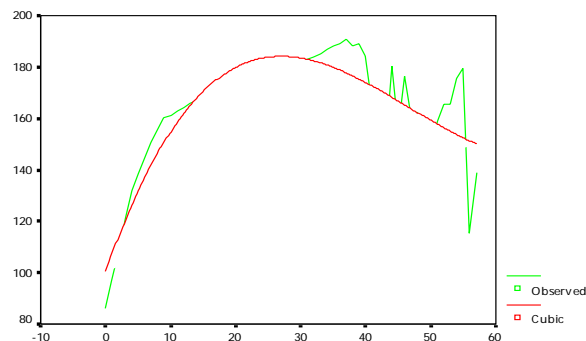


Figure 1 Curve fitting of cd4 cell count changes

The test statistic indicators are goodness of fit $R = 0.88317$, standard error Standard error = 11.117114, F test value: = 63.81568, significance level: Signif F = 0.0000 << 0.05. It can be seen that at 95% significance level Next, we consider the model to be reliable. The equation piecewise model for the variation of HIV concentration is constructed as:

$$Y_2 = \begin{cases} 4.14579 - 0.17861X + 0.00754X^2 - 0.0011X^3 & (X > 40) \\ \exp(1.07584 + 0.40011/X) & (0 \leq X \leq 40) \end{cases} \quad (3)$$

The consideration of using a piecewise function is to better simulate the original data, and at the same time, considering the actual situation, the human body will have an immune effect on the drug after taking it for a long time. At present, there is no drug in the world that can well inhibit it in the long run. The reduction of cd4 concentration in AIDS patients. Similarly, we made a simulation graph of the curve and the original value in the spss environment:

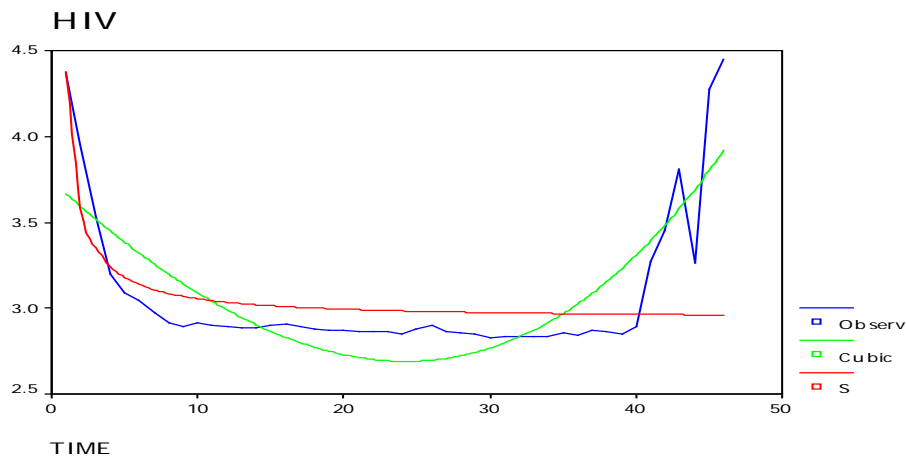


Figure 2 Curve fitting of cd4 cell count changes using a piecewise function

The test statistic index of the piecewise model is goodness of fit: = 0.8186, standard error: Standard Error = 0.25369, F test value: = 28.34473, significance level: Signif F = 0.0000 << 0.05. Visible at 95% At the confidence level of ., we accept the null hypothesis, and the model fits well.

Through the model of the content of the two indicators, it is easy to predict the effect of this therapy in continuing treatment, that is, just bring in the corresponding time value (in weeks), and the model can be used to solve it. At the same time, we can use graphics to solve the problem. It is clear that a peak in the cd4 cell count appeared around the twenty-seventh week, that is, the cd4 cell count showed a significant slump after the twenty-seventh week. It can be understood from the figure that the therapy was in the twentieth. After seven weeks, the effect on the human body began to weaken significantly, and it can be considered that the drug should be stopped. At the same time, the change in the concentration of HIV was also considered to support the view that the concentration of HIV at this time reached almost the lowest point. According to the experience in the current medical industry Most of them use cd4 cell counts to predict the status of AIDS. Here we also follow this principle. In this issue, we should observe the cd4 cell counts and take into account the cost of HIV and AIDS. We believe that Medication should be stopped around the 27th week.

3.2 Data linear interpolation model

For more than 1,300 respondents, they were randomly divided into four groups to take different forty weeks of follow-up test efficacy. After studying the data in Annex II, considering that the data given is relatively scattered, we use the idea of Question 1, we think This data also has discontinuities, and from the perspective of the number of detections, the data is not complete enough to predict. For this reason, we use the same method for data processing and use linear interpolation to fill the data from the zeroth week to the last week. Also here we assume that the change during the treatment is uniform. In this way, we can get the weekly time series data of the four therapies, so as to calculate the

weekly average of all the respondents, using spss through multiple simulations, After fitting a better model for each therapy, you can use this model to predict the efficacy of each therapy separately.

Following the previous analysis idea, first calculate the average cd4 content of all the testees each week. The equation of the optimal curve of the change in cd4 cell count in patients treated by therapy one with spss simulation is:

$$Z_1 = 2.99245 - 0.00712X - 8.70746E - 05X^2 - 2.46608E - 06X^3 \quad (4)$$

We use the curve fitting graph simulated by spss. From the intuitive and statistical test indicators, we can see that the fitting and prediction results are very good.

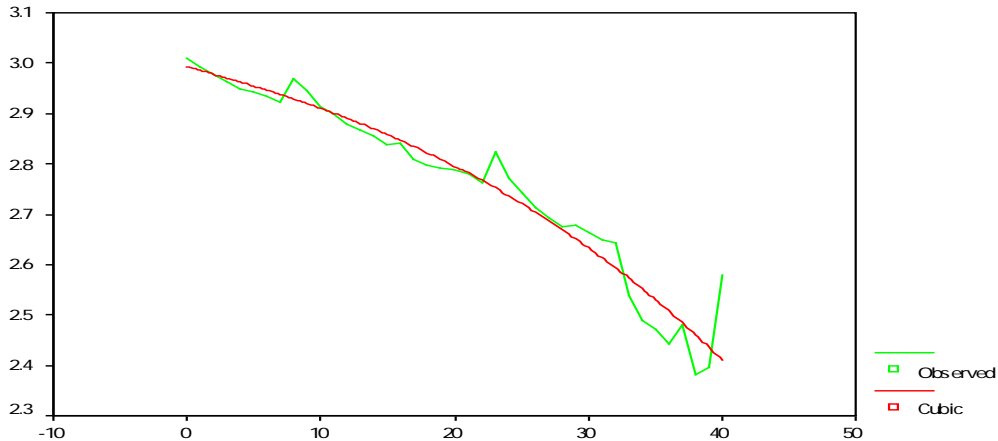


Figure 3 Curve Fitting for AIDS Patients in Treatment One

The statistical index of the therapy-model test is goodness of fit: = 0.992447, standard error: Standard Error = 0.04724, F test value: = 217.19730, significance level: Signif F = 0.000 << 0.05.

It can be seen that at the significance level of 95%, we think the model is reliable. The optimal value in the effective interval appears when x is about twenty-six weeks, so we predict that patients taking therapy one are best in For about 26 weeks, stop taking this drug combination.

Repeat the above-mentioned operation process, first calculate the average CD4 content of all the testees every week (see the excel table in the appendix), the equation of the optimal curve of the change of CD4 cell count in the patients treated by therapy two simulated with SPSS for:

$$Z_2 = 3.015932 - 0.001708X - 0.000212X^2 - 1.76209968E - 06 X^3 \quad (5)$$

We use the curve fitting graph simulated by spss. From the intuitive and statistical test indicators, we can see that the fitting and prediction results are very good.

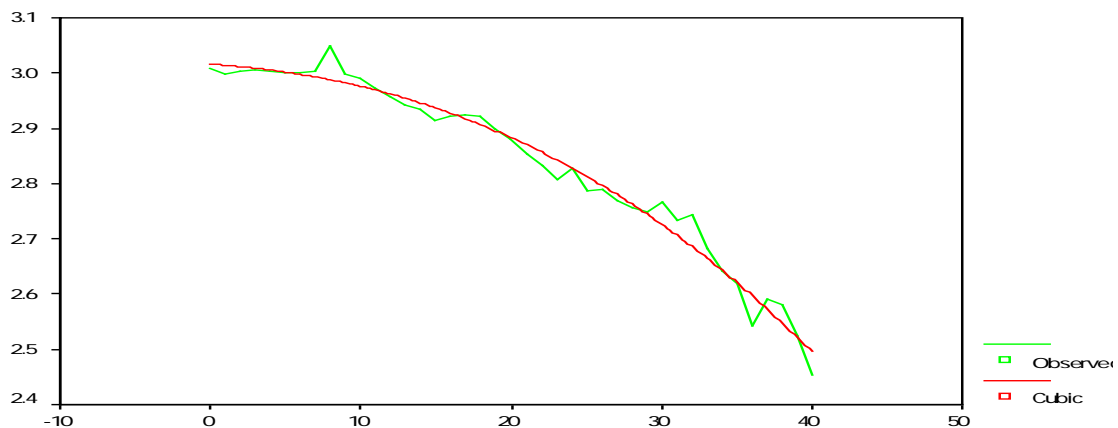


Figure 4 Curve Fitting for AIDS Patients Performing Treatment II

The statistical index of the second model test for the therapy is goodness of fit: = 0.97781, standard error: Standard Error = 0.02376, F test value: = 588.46970, significance level: Signif F = 0.000 <<

0.05. It can be seen at the 95% confidence level Next, it can be concluded by the value of the upper statistic that we accept the null hypothesis.

Repeat the above-mentioned operation process, first calculate the average CD4 content of all the testees every week (see the excel table in the appendix). The equation of the optimal curve for the change of CD4 cell count in patients treated with therapy three using SPSS is :

$$Z_2 = 2.972103 + 0.032937X - 0.002161X^2 + 3.31715957E - 05 X^3 \quad (6)$$

We use the curve fitting graph simulated by spss. From the intuitive and statistical inspection indicators, we can see the effect of fitting and prediction.

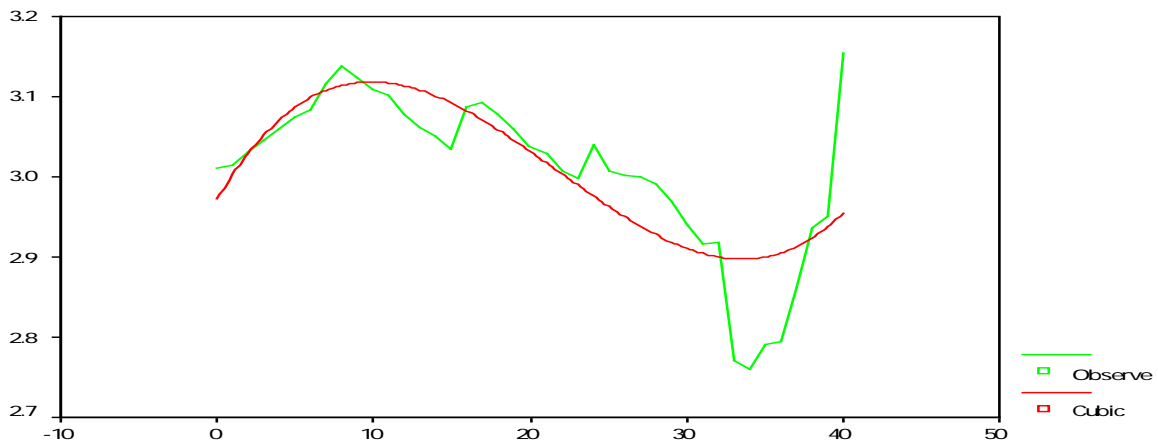


Figure 5 Curve Fitting for AIDS Patients Performing Treatment III

The statistical test of the three models of therapy is the goodness of fit: = 0.81013, standard error: Standard Error = 0.06047, F test value: = 23.55132, significance level: Signif F = 0.000 << 0.05. It can be seen at the 95% confidence level, It can be concluded from the above statistics that we accept the null hypothesis.

The best quality within the effective range appears at about nineteen weeks of x, so we predict that patients taking therapy three, preferably around nineteen weeks, stop taking this drug combination.

Repeat the above-mentioned operation process, first calculate the average cd4 content of all the testees each week (see the appendix). The equation is:

$$Z_2 = 2.875674 + 0.064755X - 0.003298 X^2 + 4.32029407E - 05 X^3 \quad (7)$$

We use the curve fitting graph simulated by spss. From the intuitive and statistical inspection indicators, we can see that the fitting and prediction results are very good:

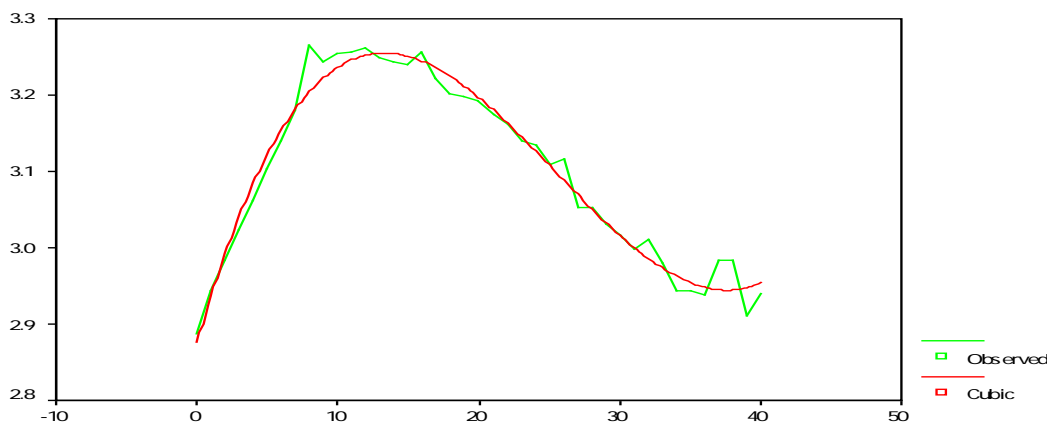


Figure 5 Curve Fitting for AIDS Patients Performing Treatment III

The statistical index of the four model tests for the therapy is the goodness of fit: = 0.98713 Standard error: Standard Error = 0.01996, F test value: = 469.99144, significance level: Signif F =

0.000 << 0.05. It can be seen at the 95% confidence level According to the value of the above statistics, we can conclude that we accept the null hypothesis. The best quality in the effective interval appears when X is about twenty-five weeks, so we predict that patients taking therapy four, preferably at twenty-five For about a week, stop taking the drug combination.

Summarizing the above four prediction models of curative effect, all can use the cubic curve to have a better simulation. We can see a clear trend from the graph, that is, at the beginning of taking the four drug combinations, the effect is better. In particular, the third and fourth drug combinations are used. The long-term effects are not particularly good. But we also noticed that the third and fourth therapies have a more obvious upward trend in cd4 cell counts after nearly forty weeks. In particular, after carefully matching the original data, we found that this rule does exist. For this, we temporarily think that the third and fourth treatments still have some effect after long-term use.

3.3 Comprehensive comparison of the efficacy of the four therapies

Since these data are obtained in a randomly grouped population, we assume that these data samples are independent, so we can use the non-parametric test of multiple independent samples in statistics to compare the four data samples Here, we use multiple independent samples in SPSS software to compare rank sum test functions, including median test and Kruskal-Wallis H Test, and give a statistical analysis of the relative efficacy of each therapy. .

We have obtained the average of CD4 cell counts for all patients from 0 to 40 weeks for each therapy. Since these data are obtained in a randomly grouped population, we assume that these data samples are independent, so we The non-parametric test of multiple independent samples in statistics can be used to compare the four data samples. We use the multiple independent samples included in SPSS11.5 software to compare rank and test functions. The details include: Digit test and Kruskal-Wallis H Test.

Median test

Table 1 Median test effect of different therapies

		therapy			
		Therapy 1	Therapy 2	Therapy 3	Therapy 4
CD4	> Median	5	12	31	34
	<= Median	36	29	10	7

It is clear from the frequency table 1 of the statistical results that there are 36 less than the median in therapy one and only five more than the median in therapy one, and only seven less than the median in therapy four, and There are 34 numbers greater than the median. It can be seen that the effect of treatment four is better. From the table, it can also be roughly seen that the methods one to four are gradually better.

Table 2 Test statistics indicators

	CD4
N	164
Median	2.960295
Chi-Square	59.024(a)
df	3
Asymp. Sig.	.000

a 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 20.5.

b Grouping Variable: METHOD

From the values of the Chi-Square test statistics in Table 2, we can see that at the 95% confidence level, we believe that there is a significant difference in the efficacy of the four therapies.

Kruskal-Wallis H Test

Table 3 Kruskal-Wallis H Test effect of different therapies

	METHOD	N	Mean Rank
CD4	1	41	41.80
	2	41	57.10
	3	41	105.73
	4	41	125.37
	Total	164	—

It can be seen from Table 3 that the efficacy of the first to fourth treatments is getting better and better, and the gap is obvious. It has reached a unified conclusion with the median test.

Table 4 Test statistics indicators

	CD4
Chi-Square	85.065
df	3
Asymp. Sig.	0.000

a Kruskal Wallis Test

b Grouping Variable: METHOD

From the Chi-Square values in Table 4, etc., at a significance level of 95%, we can consider that there are obvious differences between the four therapies. In summary, the four therapies are based on the CD4 cell count only. The curative effect is as follows: therapy four is the best, therapy three is the third, therapy two is the worse, and therapy one is the worst. And the four kinds of therapies all have better short-term effects and the long-term effects will decrease, but the three and four therapies are slightly better.

3.3 Optimal therapy for different age groups

We guess that the best treatment method may be different in different age groups. So we group ages to find out the best treatment method for that age group in different age groups. We refer to the previous AIDS Age distribution statistics, the overall profile of ACTG's published data, and the person's own physiological characteristics, the data are divided into three segments according to age: 25 years old and before, 25 to 40 years old, and over 40 years old. These data are divided into groups according to therapy. Four groups, counting the average CD4 cell counts of all the respondents at each detected moment (a total of six data). Similarly, since these data are obtained in randomly grouped populations, we assume that these data samples are independent. Yes, we can still use non-parametric tests of multiple independent samples, mainly using Kruskal-Wallis H Test and median test to compare these data samples. The comparison results are as follows:

(1) the first age group (25 years old and before)

Kruskal-Wallis H Test

Table 6 Efficacy of four therapies for patients of the first age

	First age group (<25)	Number	Average rank
cd4 mean	1	6	16.83
	2	6	4.33
	3	6	14.00
	4	6	14.83
	Total	24	—

Table 7 Test statistics level

	cd4 mean
Chi-Square	11.180
df	3
Asymp. Sig.	0.011

a Kruskal Wallis Test

b Grouping Variable: first age

According to the Kruskal Wallis Test data in Tables 6 and 7, it can be seen that the significance level = 0.011 < 0.05, which means that there are significant differences between the four treatments, and it is known by the average rank. Good therapy is the first, i.e. the first drug combination is suitable for adolescent patients.

Median test

Table 8 Efficacy of four therapies for patients in the first age group

		First age			
		1	2	3	4
cd4 mean	> Median	5	0	3	4
	<= Median	1	6	3	2

Table 9 Test statistics level

	cd4 mean
N	24
Median	2.582567
Chi-Square	9.333(a)
df	3
Asymp. Sig.	0.025

a 8 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 3.0.

b Grouping Variable: first age

According to the data of Table 8 and Table 9 median test method, it can be seen that the significance level = 0.025 < 0.05, it can be considered that there are significant differences in the four treatments, and there are more than the median number. The best treatment for patients of one age group is the first, that is, the first drug combination is suitable for adolescent patients. The test results are consistent with the Kruskal Wallis Test.

(2) Second age group (25 to 40 years)

Kruskal Wallis Test

Table 10 Efficacy of four therapies for patients in the second age group

	Second age	Number	Average rank
cd4 mean	1	6	8.33
	2	6	9.67
	3	6	17.50
	4	6	14.50
	Total	24	—

Table 11 Test statistics level

	cd4 mean
df	3
Asymp. Sig.	0.089
	—

a Kruskal Wallis Test

b Grouping Variable: Second age group

According to the Kruskal Wallis Test data in Tables 10 and 11, it can be seen that the significance level is slightly greater than 0.05, that is, it can be considered that there are differences between the four treatments. It is known from the average rank that it is best for patients in the second age group. The third therapy is the third drug combination, which is suitable for middle-aged patients.

(3) the third age group (over 40 years old)

Kruskal Wallis Test

Table 12 Efficacy of the four therapies for patients in the third age group

	Third age	Number	Average rank
cd4 mean	1	6	4.50
	2	6	14.00
	3	6	11.83
	4	6	19.67
	Total	24	—

Table 13 Test statistics level

	cd4 mean
Chi-Square	14.167
df	3
Asymp. Sig.	0.003

a Kruskal Wallis Test

b Grouping Variable: third age

According to the Kruskal Wallis Test data in Tables 12 and 13, it can be seen that the significance level = 0.003 < 0.05, which means that there are significant differences between the four treatments. According to the average rank, it is the most important for patients in the third age group. Good therapy is the fourth, that is, the fourth drug combination is suitable for elderly patients.

Median test

Table 14: The efficacy of the four therapies for patients in the third age group

		Third age			
		1	2	3	4
cd4 mean	> Median	0	4	3	5
	<= Median	6	2	3	1

Table 15: Test statistics levels

	cd4 mean
N	24
Median	3.004913
Chi-Square	9.333(a)
df	3
Asymp. Sig.	0.025

a 8 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 3.0.

b Grouping Variable: third age

From the data of the median test method, it can be seen that the significance level = 0.025 < 0.05, that is to say, there are significant differences between the four therapies. It is known from the number greater than the median that it is best for patients of the third age. The fourth therapy is the fourth drug combination, which is suitable for elderly patients. The test results are consistent with the Kruskal Wallis Test.

Using the Kruskal-Wallis H Test and the median test, we can conclude that the first therapy is more suitable for adolescent patients in the first age group; the third therapy is more suitable for middle-aged patients in the second age group; and The fourth treatment is best for elderly patients in the third age group.

4. Summary

Due to the discontinuous characteristics of the sampling test data, and from a weekly perspective, the number of tests is too small to make further predictions, so linear interpolation fitting was used to simulate each respondent uniformly Cd4 cell count and hiv concentration in the body every week (and take care to exclude some singular data), and then average the cd4 cell count and hiv concentration of all the respondents each week, so that you can get a It is relatively complete data in terms of time series. With these data, with the help of computer's powerful fitting function, we can get the curve of the effect of therapy on human cd4 cell count and HIV concentration, and then predict the treatment effect and determine the optimal treatment termination time.

For the study of the efficacy of therapies, we first classify the data according to the treatment method. The treatment of the data under each treatment method follows the treatment method of problem 1. First, we establish a model of the effect of each treatment method on the cd4 cell count, and then use statistics. The non-parametric test method of multiple independent samples uses the independent ranks of multiple independent samples in the spss11.5 software to compare the rank sum test function to complete the effect comparison work, and then determines the optimal treatment termination time, which has practical significance.

References

- [1] Zhu Zheng, Tang Qi, Lu Hongzhou. Current status and challenges of pre-exposure prevention research [j]. Chinese Journal of AIDS, 2019, 25 (12): 1307-1310.
- [2] Hu Yunquan et al. Fundamentals and Applications of Operations Research (Fourth Edition) [d]. Higher Education Press. 2004.4 (4): 308-340
- [3] Liu Hui, Sun Bo, Zeng Fanyong, Xie Shuibo. Quick estimation of total runoff control rate and design rainfall based on linear interpolation [j]. Water Resources Planning and Design, 2019 (10): 41-43 + 148