SAS MACRO PROCEDURE FOR ANALYZING THE PROGRESSION OF CHRONIC RENAL FAILURE

Gretz,N. and Strauch, M.
Clinic of Nephrology, Klinikum Mannheim,
University of Heidelberg, Mannheim, FRG.

Abstract

Various kidney diseases lead to terminal renal failure. Some of these diseases seem to have a specific progression rate. Enormous difficulties, however, exist in analyzing progression rates. This is due to different starting points of the observation periods of the individual patients: some patients were seen early in the course of the disease, some late. Another problem results from the number of measurements of renal function per patient and unit of time. An approach is presented for calculating the median course of a disease after redefining the starting point of the observation period. Our macro plots the original data set, the overlaid individual courses after redefining their zeropoint and their median course. Furthermore the data of the median course are printed. The following SAS procedures were employed: GLM, GPLOT, GPRINT, PRINT, PRINTTO, REG, UNIVARIATE. Intermediate results are used extensively. We think that the availability of these results is a valuable feature of SAS and facilitates its application considerably.

Introduction

Some kidney diseases cause terminal renal failure. The rate of their progression to endstage renal disease, however, is unknown. It was supposed that each of these diseases has a specific progression rate. A widely used parameter for the assessment of renal function is the serum concentration of creatinine (SCR). As a renal disease progresses, this metabolic product accumulates following a hyperbolic function . This hyperbolic curve can be transformed into a straight line, using a reciprocal transformation. A reciprocal transformation, however, results in a straight line only, if there are not too many factors modifying the progression. Furthermore slopes can be compared only, if they are calculated from the same data range, but they cannot be used for prediction purposes.

This paper presents a SAS macro procedure for the assessment of disease specific progression rates, which will allow:

a) to test for homogeneity of the progression of the evaluated group of patients,
b) to compare different diseases, and
c) to predict the course of the disease in individual patients.

The output of the macro is exemplified by the data of 50 patients with polycystic kidney disease in whom 1085 SCR determinations had been performed.

Description of the SAS macro procedure

In this paragraph the programme and its output is described. The macro procedure is given in the appendix. The steps mentioned in this paragraph are referring to the flow sheet of the macro and the original programme (see appendix).

The programme handles the courses of the disease of individual patients separately. This is achieved by BY statements.

In each of the following MERGE statements the original data set or its modifications which contain several data lines per patient, were merged with data sets, containing a single line per patient. All these data sets contain as common variable the variable:PAT (patient). As the correspondence between the original data set and the data set containing the calculated data was not one-to-one, and the variable PAT appears in all data sets, the single data line of each patient was added repetitively to his original data. In the original data set this resulted in the availability of general/calculated information in each line of a patient.

The %MACRO statement specifies a list of macro variables, which are keyword parameters. The variables MINI and MAXI define the data range in which SCR data of patients have to be available. The parameters SCRL, SCRU, and NR specify the prechosen lower and upper part of the course and the minimal number of SCR determination of a section of the course, to which a curve is to be fitted. The

variable Z gives the x,y-coordinates of the zeropoint, on which the data were standardized.

First step: In this step the raw data: SCR and observation period in months are plotted (Fig. 1). Fig. 1 illustrates, when compared with the standardized data (Fig. 2), the problems caused by the different starting points of the observation period in our patients. One can imagine that there are a lot of hyperbolic curves with different zeropoints in this plot, but they are not ready accessable for any kind of evaluation. This plot clearly establishes that raw data are not helpful to achieve any of our aims: test of homogeneity, comparison of disease groups, and prediction.

Second step: Here the original data are screened for courses fulfilling criteria, which can be defined in the programme invoking the macro (see also %MACRO statement). At first the minimum and maximum of the SCR data of a patient are assessed and merged with original data set.

Then all courses, which satisfy the requirements for the data range per course, are selected for further analysis. The required data range is defined by the macro variables MIN and MAX. The remaining data are checked if in a defined data range (SCRL, SCRU) a certain number of observations (NR) is available. The variable NR can be specified according to the availability of data points and the need of accuracy of the fitted line. This step is crucial, as a) if there are less than 2 observations per course an error message will result and b) the accuracy of the fitted curve can be influenced.

The above steps result in two data sets available for further analysis: one (MIX) is the enhanced original data set, the other (MI) is a reduced data set containing parts of the individual courses, to which a curve is to be fitted.

Third step: In this step a straight line is fitted to a defined part of the SCR data of a patient. The data range to which the line is fitted is chosen by the variables SCRL and SCRU in the previous step. The data, to which a straight line is fitted, are not the original SCR values, but the reciprocally transformed data (RSCR). This is done due to pathophysiological considerations (1). It is important however, to make sure, that in each patient the line is fitted to the same part of the data set. Otherwise this approach is not valid, as fitting lines to different parts of the course in a single patient results in highly variable slopes of these lines (2).

For the curve fitting we use the SAS procedure REG. This procedure allows to specify an output data set containing statistics calculated for each patient. Thus it is possible to save the information on slope and intercept in an output data set and then to calculate a regression line for each patient .

Fourth step: Now the time scale of the course of the renal disease is changed. In each patient the time is calculated from the regression parameters: slope and intercept, in which he would pass through the chosen SCR value: Z. Then this calculated time point is set to zero and the time scale changed accordingly. Before doing that it is important to redefine the variable MONTHS of the output data set. This variable represents the slope of the regression line. If this variable is not redefined, the variable: MONTHS will have a different meaning in the output data set of the procedure REG and the modified original data set. In a MERGE statement this results in the deletion of the value of one of these two variables. We call the variable MONTHS, representing the slope of the regression line, SLOPE.

After the standardization of the zeropoint of the time scale, the original SCR values are plotted versus the modified time data (Fig. 2). A comparison of this plot (Fig.2) with the plot of the raw data (Fig. 1) demonstrates the effect of the standardization. In our example Fig. 2 still exhibits some degree of scattering.

Fifth step: In this step we calculate the median time, in which a single patient reaches successive SCR values. Before this is done all SCR values are rounded off. These rounded data are represented by the variable SCRR. Then the median is taken from the times with the same corresponding SCRR value. Thus the number of time and SCRR points available in each patient becomes equal. The median times and the SCRR data of each patient are outputted in the data set: NEW.

Sixth step: Now median (P50) and quartile (P25,P75) times for all patients together are calculated for successive SCRR values. We restrict this computation to SCR data below 11 mg/dl, as only then enough data are available. In addition to medians and

quartiles, the interquartile range (QRANGE), the fifth (P5) and the ninety-fifth percentile (P95) are estimated. All the above parameters are outputted into a data set.

Then this output data set is printed by using the SAS procedure GPRINT (Fig. 3). This step is not essential, but is very helpful, if you are interested in preparing tables for presentation. Slides can easily be prepared if in the GOPTION statement VSIZE is greater than 30. GPRINT can only be used after printing the above output data set and directing the output of the print procedure to the file FT20F001. It is reasonable to use the procedure: PRINT with the procedure FORMAT in order to improve the features of your output.

At last the median (P50) and quartile (P25,P75) times are plotted versus thier corresponding SCRR values (Fig. 4). The respective data points are connected with a straight line and the plots of the variables are overlaid.

Flow sheet of the macro

Step		Printed/plottet output
1	Raw data	Plot of the raw data(Fig.1)
2	Selection of patients fullfilling criteria for: lowest and highest SCR value, number of SCR determinations for a defined part of the course.	
3	Curve fitting	
4	Redefining the zeropoint of the course of each patient	Plot of the standardized data (Fig.2)
5	Calculation of the median time, when the individual patient reached certain SCR concentrations	
6	Calculation of the median and quartiles of the above median time in all patients	Print of the calculated data (Fig. 3) Plot of the median and quartile curves (Fig.4)

Applicability of the macro

Fig. 2 and 4 allow a visual examination for homogeneity. If the quartile curves are in parallel with the median curve, the course of the disease can be regarded as homogeneous. The same is true, if the distance between the quartile curves is constant. An additional test for homogeneity is a plot of the interquartile range (QRANGE) versus the respective SCR values.

This macro is also helpful to assess the influence of variables on the progression of renal diseases. If a variable is thought to be an influencing parameter, the original data are divided into groups according to this factor. Then the macro is used for each subgroup. The comparison of the outputs will show, quantitatively and visually, if an influence exists or not. In the same way the effects of therapeutic regimens can be tested.

The progression rates of patients with different renal diseases can be compared and common features detected. Medians and quartiles can be stored for each disease group and used to produce overlaid plots. This will easily visualize differences between groups, which can be quantified by using the printed output of the medians and quartiles.

The output of the presented macro procedure can also be used for prediction purposes in the individual patient, as it exhibits the median course of a disease and its variability

Acknowledgement

The authors are indebted to Mrs Wittich for her secretarial assistance and technical advice in preparing the manuscript.

References

1. Mitch, W.E.; Walser,M.:
A simple method of estimating progression of chronic renal failure
Lancet 2, 1326-1328 (1976)

2. Gretz,N.; Manz,F.; Strauch,M.:
Predictability of the progression of chronic renal failure
Kidney International 24, suppl 15, S2-S5 (1983)

Appendix

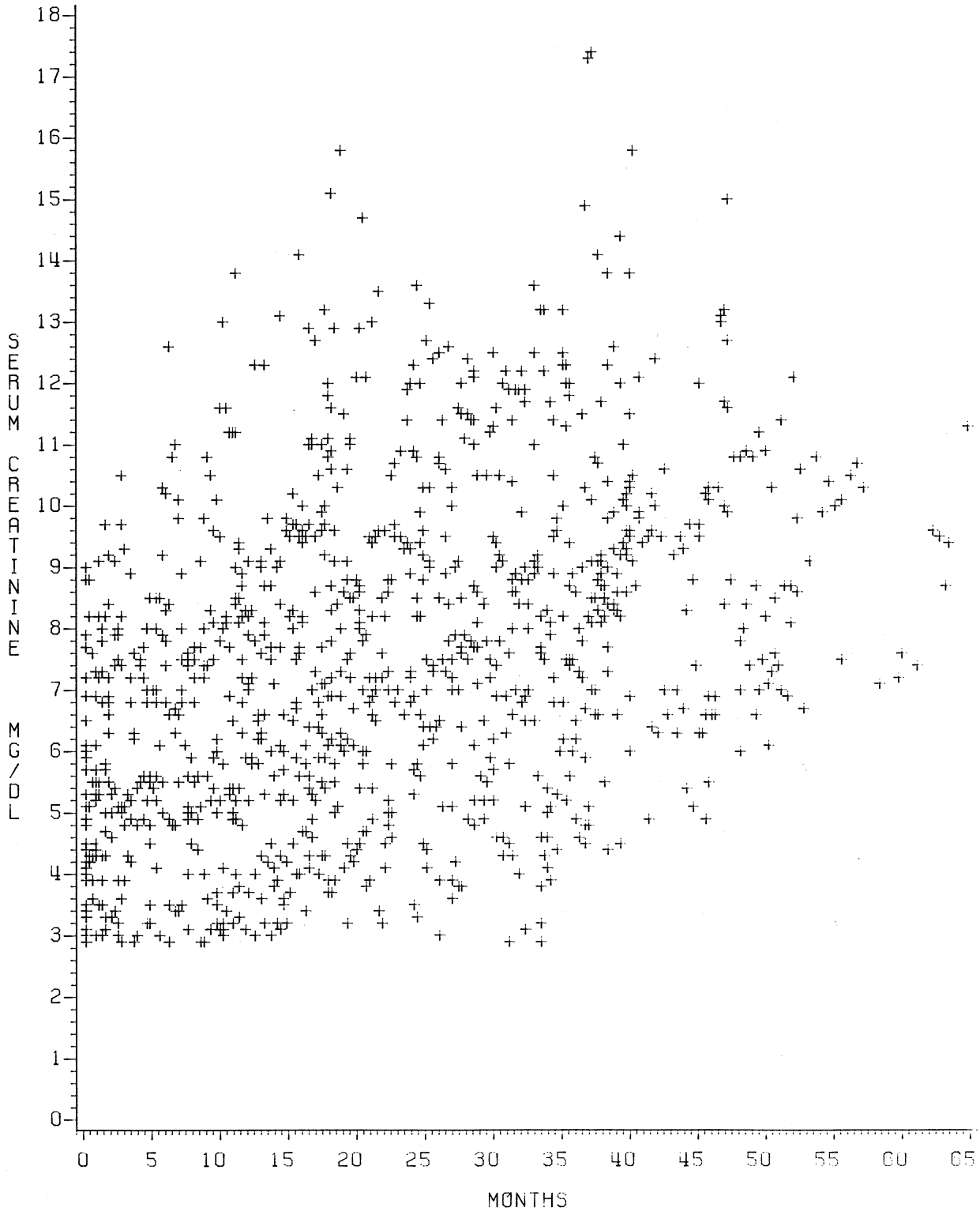
```
%MACRO CURVE2(MINI=,MAXI=,SCRL=,SCRU=,NR=,Z=);

TITLE2 .H=3       RAW DATA;

         *-----------------------------*
         |                             |
         |          RAW DATA           |                    1. STEP
         |                             |
         *-----------------------------*;

PROC SORT;BY PAT;

LABEL SCR='SERUM CREATININE   MG/DL'
      MONTHS='OBSERVATION PERIOD  MONTHS';

PROC GPLOT; PLOT SCR * MONTHS;

         *-----------------------------*
         |                             |
         |   PATIENT SELECTION         |                    2. STEP
         |                             |
         *-----------------------------*;

PROC UNIVARIATE NOPRINT; BY PAT ; VAR SCR;
     OUTPUT OUT=MINMAX  MIN=MIN MAX=MAX;

DATA MIX;   MERGE ORIGINAL MINMAX; BY PAT ;
            IF MIN<&MINI AND MAX>&MAXI;

DATA MIXED; SET MIX;IF SCR>&SCRL AND SCR<&SCRU;

PROC UNIVARIATE NOPRINT; BY PAT ; VAR SCR;
     OUTPUT OUT=COUNTED  N=N;

DATA MI; MERGE MIXED COUNTED; BY PAT ;
RSCR=1/SCR;
IF N>&NR;

         *-----------------------------*
         |                             |
         |      CURVE FITTING          |                    3. STEP
         |                             |
         *-----------------------------*;

PROC REG OUTEST=AB NOPRINT; BY PAT;
     ID  MONTHS;
     MODEL RSCR=MONTHS;
```

```
        *-----------------------------*
        |                             |
        |    STANDARDIZED CURVES      |                    4. STEP
        |                             |
        *-----------------------------*;

DATA B; SET AB; SLOPE=MONTHS; W6=((1/&Z)-INTERCEP)/SLOPE;
PROC SORT; BY PAT;

DATA ALL; MERGE MIX B; BY PAT;
MONTHS=MONTHS-W6;
SCRR=ROUND(SCR,1.0);

LABEL MONTHS='STANDARDIZED MONTHS'
      SCRR='ROUNDED SERUM CREATININE   MG/DL';

PROC SORT; BY PAT SCRR;

TITLE2 .H=2 STANDARDIZED DATA;
TITLE3 .H=2 DATA RANGE:&MINI - &MAXI   ZEROPOINT:0/&Z;
TITLE4 .H=2 APPROXIMATION:&SCRL - &SCRU   NUMBER OF OBS.:&NR;

PROC GPLOT; PLOT SCR*MONTHS;

        *-----------------------------*
        |                             |
        |  MEDIAN/PAT AND INTERVAL    |                    5. STEP
        |                             |
        *-----------------------------*;


PROC UNIVARIATE NOPRINT; BY PAT SCRR; VAR MONTHS;
     OUTPUT OUT=NEW MEDIAN=MEDIAN;

PROC SORT; BY SCRR;

        *-----------------------------*
        |                             |
        |   MEDIAN OF THE MEDIANS     |                    6. STEP
        |                             |
        *-----------------------------*;

DATA RESULTS; SET NEW; IF SCRR<11;

PROC UNIVARIATE NOPRINT ; BY SCRR; VAR MEDIAN;
     OUTPUT OUT=NEWEST MEDIAN=P50
     Q1=P25 Q3=P75 QRANGE=QRANGE P5=P5 P95=P95;


        *-----------------------------*
        |                             |
        | MEDIAN AND QUARTILE CURVES  |
        |                             |
        *-----------------------------*;

PROC PRINTTO UNIT=20    ;
PROC PRINT;FORMAT P50 P25 P75 QRANGE P5 P95 9.;
TITLE4  ; TITLE3  ; TITLE2   ;

PROC GPRINT DD=FT20F001;
TITLE2 .H=2  MEDIAN AND QUARTILE CURVES;
TITLE3 .H=2  DATA RANGE:&MINI - &MAXI   ZEROPOINT:0/&Z;
TITLE4 .H=2  APPROXIMATION:&SCRL - &SCRU   NUMBER OF OBS.:&NR;

PROC GPLOT; PLOT   SCRR*P50=2 SCRR*P25=1 SCR*P75=3 /
     OVERLAY VAXIS=3 TO 10 BY 1 HAXIS=-35 TO 30 BY 5;
     SYMBOL1 L=20 I=JOIN C=BLACK V=STAR W=4;
     SYMBOL2 L=1 I=JOIN C=BLACK V=STAR W=8;
     SYMBOL3 L=20 I=JOIN C=BLACK V=STAR W=4;

%MEND CURVE2;
```

POLYCYSTIC KIDNEY DISEASE

Fig.1

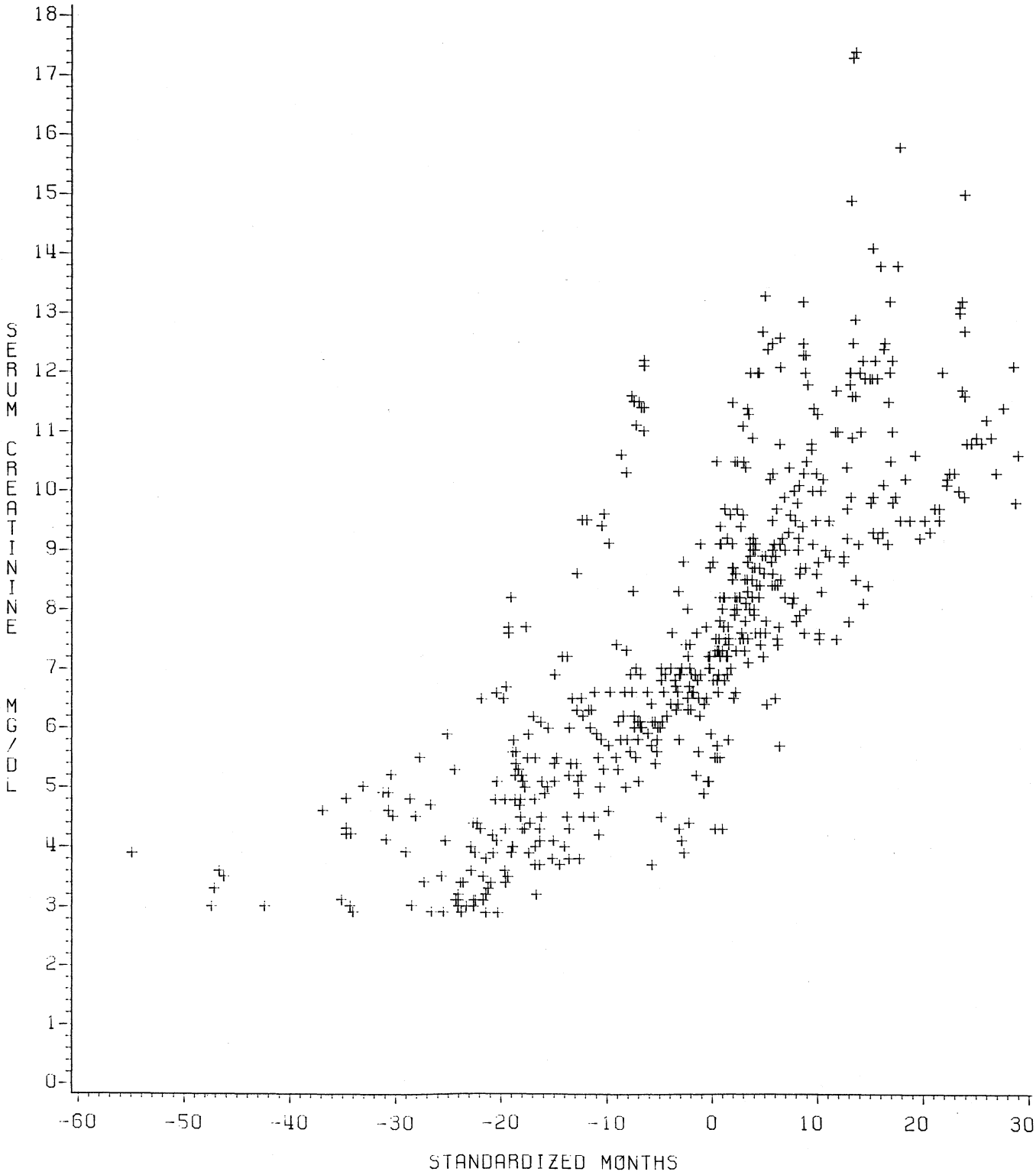
RAW DATA



POLYCYSTIC KIDNEY DISEASE

Fig. 2

STANDARDIZED DATA
DATA RANGE: 5 - 10 ZERØPØINT·0/7
APPROXIMATION: 6 - 8 NUMBER ØF ØBS. . 2



POLYCYSTIC KIDNEY DISEASE

Fig. 3

MEDIAN AND QUARTILE CURVES
DATA RANGE:5 - 10 ZEROPOINT:0/7
APPROXIMATION:6 - 8 NUMBER OF OBS.:2

POLYCYSTIC KIDNEY DISEASE

OBS	SCRR	P50	P25	P75	QRANGE
1	3	-24	-30	-21	9
2	4	-19	-28	-16	12
3	5	-14	-18	-7	11
4	6	-6	-11	-3	8
5	7	-1	-3	1	5
6	8	3	1	5	4
7	9	6	2	10	8
8	10	8	3	12	9

POLYCYSTIC KIDNEY DISEASE

Fig. 4

MEDIAN AND QUARTILE CURVES
DATA RANGE:5 - 10 ZEROPOINT:0/7
APPROXIMATION:6 - 8 NUMBER OF OBS.:2