

J. Clin. Chem. Clin. Biochem.
Vol. 19, 1981, pp. 1153–1166

Comparison of Three Distribution-Free Procedures in the Establishment of Assigned Values in Control Sera¹⁾

The Establishment of Assigned Values in Control Sera, III

By H. Passing

Hoechst AG, POB 800320, D-6230 Frankfurt 80, Federal Republic of Germany

(Received February 23/July 15, 1981)

Summary: Distribution-free statistical procedures should be applied to the establishment of assigned values and uncertainty intervals in a control serum. The two problems, how to find an appropriate statistical evaluation procedure and how to find an optimized experimental design, are simultaneously dealt with here: Three distribution-free procedures are presented, each based on elimination of extreme values, and 60 designs are considered differing with respect to the number of reference laboratories, of independent series, and of single or double determinations. Using the data of the study described in part 1 of this series (Passing, H. et al. (1981) this j. 19, 1137–1144) we simulated these designs and the pertaining assigned values and uncertainty intervals given by each evaluation procedure.

From this study one evaluation procedure is shown to be superior to others. This optimized procedure has the following characteristics: Extreme values are eliminated so that the width of the uncertainty interval is as small as possible. The median of the remaining values is the assigned value. Moreover, 6 reference laboratories are shown to be appropriate.

Vergleich von drei verteilungsfreien Verfahren für die Ermittlung von Sollwerten in Kontrollseren. Ermittlung von Sollwerten in Kontrollseren, III.

Zusammenfassung: Für die Ermittlung von Sollwerten und Sollbereichen eines Kontrollserums sollten verteilungsfreie statistische Verfahren angewandt werden. Die beiden Probleme, wie man ein geeignetes statistisches Auswertungsverfahren und wie man einen optimierten Meßplan findet, werden hier simultan behandelt: Drei verteilungsfreie Verfahren werden dargestellt, von denen jedes auf der Elimination extremer Werte beruht, und 60 Meßpläne werden berücksichtigt, die sich in der Anzahl der Referenzlaboratorien, der unabhängigen Serien und in Einzel- oder Doppelbestimmungen unterscheiden. Indem wir die Daten der in Teil 1 dieser Reihe (Passing, H. et al. (1981) this j. 19, 1137–1144) beschriebenen Studie benutzten, simulierten wir diese Meßpläne und die von jedem Verfahren gegebenen zugehörigen Sollwerte und Sollbereiche.

Aus dieser Studie wird abgeleitet, daß ein Auswertungsverfahren anderen überlegen ist. Dieses optimierte Verfahren ist folgendermaßen charakterisiert: Extremwerte werden so eliminiert, daß der Sollbereich möglichst geringe Breite hat. Der Median der verbleibenden Werte ist der Sollwert. Außerdem wird gezeigt, daß 6 Referenzlaboratorien angemessen sind.

Introduction

In part 2 of this series (3) we discussed four statistical models which are partly applied to the establishment of assigned values of a control serum and their uncertainty intervals. They are based on assumptions concerning normal distribution and equality of precision of analytical values each in a characteristic manner. It has been shown that such assumptions will in general be erroneous

thus resulting in misleading or seriously wrong assigned values and uncertainty intervals. Therefore it has been concluded that a distribution-free procedure should be applied in order to have an evaluation algorithm which is universally appropriate to all constituents in all control sera. This consequence is also given in l. c. (4).

Furthermore, it has been discussed (3) that models based on differences in accuracy between the reference laboratories are necessarily inappropriate if they do not contain any variation of accuracy between customers. All

¹⁾ A preliminary report is given in l. c. (1).

such models based on an analysis of variance fail: They give a prediction interval for one analytical value of the customer assumed to be centered at the general mean of the reference laboratories, whereas the width of the interval does not cover any differences of accuracy. In fact, a theoretical model has not been developed up to now which covers differences of accuracy both between reference laboratories (being estimable) and between customers (being a priori unknown).

Consequently it seems reasonable to pool all the analytical values of the reference laboratories. In this paper we present three such procedures.

However, the choice of a statistical procedure may depend on the underlying experimental design i.e. the number of reference laboratories to be incorporated with the number of independent series and the number of determinations per series. Experimental designs proposed (4, 5) or mentioned (6) vary from 3 up to 10 laboratories and with respect to the number and the arrangement of values within laboratories. Conversely, an optimized design generally depends on the statistical procedure which is to be applied to evaluation. Therefore, it is necessary to elaborate an optimized design and an appropriate evaluation procedure simultaneously.

In this paper we show one evaluation procedure to be superior, and we determine the appropriate number of reference laboratories. The other details of an optimized design will be derived (7) by using the methods developed here. These two papers form a unit.

Materials and Methods

We used the experimental data of the study described in l.c. (2). The analyses were performed by laboratories of members of the Verband der Diagnostica- und Diagnosticgeräte-Hersteller (VDGH, Association of Diagnostics and Diagnostics Instrumentation Manufacturers). The study covered the constituents creatinine, glucose, urea, alanine and aspartate aminotransferase²), creatine kinase²), and γ -glutamyltransferase²). For each constituent we had 9 to 11 laboratories (see tab. 1) each having performed double determinations in approximately 18 independent series of the unknown sample, of the known control, and of a blind control. But we processed only those data which had met the respective validity criterion as given in l.c. (3). In doing so at least 15 series were found valid in each laboratory and each constituent. Therefore we confined ourselves to the first 15 valid series of each laboratory and constituent; data from further series were neglected (if any).

We considered the following statistical procedures for the evaluation of an arbitrary design consisting of I laboratories each with J independent series and K determinations per series, respectively. At first, these data are pooled and sorted

resulting in n values $x_1 \leq x_2 \leq \dots \leq x_n$ with $n = I \cdot J \cdot K$. As the uncertainty intervals are to contain at least 95% of the analytical results, up to 5% have to be eliminated. From n we derive $m = 5n/100$, which is rounded off. Then:

Procedure 1

m of the n values are eliminated so that the range of the remaining values is as short as possible. The extreme values of the remaining ones are the uncertainty limits. The arithmetic mean of these limits is the assigned value.

Procedure 2

m of the n values are eliminated so that the range of the remaining values is as short as possible. The extreme values of the remaining ones are the uncertainty limits (identical to procedure 1). The median of the remaining values is the assigned value.

Procedure 3

From m we derive $m^* = m/2$ which is rounded off. The m^* smallest values $x_1 \dots x_{m^*}$ and m^* greatest values $x_{n-m^*+1} \dots x_n$ are eliminated. The extremes of the remaining ones i.e. x_{m^*+1} and x_{n-m^*} are the uncertainty limits. The median of the remaining values is the assigned value.

Especially, for $n < 20$ no value is eliminated, and for $20 \leq n < 40$ exactly one value is eliminated in the procedures 1 and 2, but none in procedure 3.

Furthermore, we considered all designs consisting of

I = 3, 4, 5, 6, 8, or I_{\max} laboratories³),

J = 4, 5, 7, 10, or 15 series,

K = 1 or 2 determinations per series.

For fixed K a design with I laboratories and J series is characterized by I|J. If a design I|J covers at least as many laboratories and series as a design I'|J', that is if both $I \geq I'$ and $J \geq J'$ hold, then we call the design I|J larger than the design I'|J'. Conversely, the design I|J is called smaller than the design I'|J' if both $I \leq I'$ and $J \leq J'$ hold.

According to each design i.e. for each I, J, and K and for each of the 7 constituents mentioned above an experiment was simulated using our experimental data: I out of the I_{\max} laboratories were randomly selected by using a random number generator. The first J valid series of these laboratories were processed. If K = 1 the first values only of these series of the selected laboratories were selected, but if K = 2 both values were taken. The same random selection of I laboratories was used for all designs covering exactly I laboratories, for each fixed constituent. These designs differ with respect to J or K only. For different I however a new randomization was performed. Moreover, for each of the 7 constituents a new randomization was performed.

We computed the following quantities according to each design, each randomization, and each evaluation procedure:

- The assigned value. The assigned value is called the reference assigned value⁴), when $I = I_{\max}$, $J = 15$, and $K = 2$.
- The uncertainty interval of the assigned value, and
- its width. The width of the uncertainty interval of the reference assigned value is also used as a reference quantity.

Each design was simulated T_1 times by forming T_1 randomizations as described above, see table 1. In doing so it was taken care that all I_{\max} laboratories were met equally frequently namely IT_1/I_{\max} (being an integer) times, and that all randomizations were pairwise different. Thus a sample of size T_1 was

²) Enzymes: Glutamate-pyruvate-transaminase = alanine aminotransferase = L-alanine: 2-oxoglutarate aminotransferase EC 2.6.1.2; glutamate-oxalacetate-transaminase = aspartate aminotransferase = L-aspartate: 2-oxoglutarate aminotransferase EC 2.6.1.1; γ -glutamyltransferase = (5-glutamyl)-peptide: aminoacid 5-glutamyltransferase EC 2.3.2.2; creatine kinase = ATP: creatine N-phosphotransferase EC 2.7.3.2.

³) I_{\max} depends on the respective constituent and is identical to the number of laboratories underlying this study, see table 1.

⁴) The reference assigned value depends on the respective evaluation procedure.

Tab. 1. Number of replications of individual randomization steps.

Constituent	I _{max} ^{a)}	Number I of laboratories selected									I _{max}
		3	4	5	6	8	11	12	18	22	
Creatinine	9	6 ^{b)}	12 ^{c)}	9	18	9	18	6	11	9 ^{d)}	1
Glucose	9	6	12	9	16	9	18	6	12	9	1
Urea	9	6	12	9	18	9	17	6	11	9	1
Alanine aminotransferase	11	11	22	22	42	33	63	33	65	11	22
Aspartate aminotransferase	11	11	22	22	43	33	64	33	66	11	22
Creatine kinase	9	6	12	9	16	9	16	6	11	9	1
γ-Glutamyltransferase	9	6	12	9	18	9	17	6	12	9	1

- a) number of laboratories underlying this study.
- b) number T₁ of individual randomizations of the first as well as of the second set, respectively, corresponding to approx. 7% of the number of randomizations existing in total.
- c) number T_t of individual randomizations of the total set.
- d) number T₁ = T_t of individual randomizations of the first and of the total set corresponding to 100% of the number of randomizations existing in total.

drawn out of the population of randomizations existing in total and having size⁵⁾ $\binom{I_{max}}{I}$.

This complete process was performed independently even twice for each constituent and each chosen number of laboratories covering approximately 7% of the randomizations existing in total. This resulted in a first and a second set of randomizations. Both sets formed the total set of randomizations; yet, identical randomizations occurring accidentally both in the first and in the second set were considered only once⁶⁾. In particular if T₁ already corresponded to 100% the first set was identical to the total set of randomizations. The total set covered T_t randomizations.

To condense all these details, we calculated two objective quantities from the total set of randomizations for each constituent, each design, and each evaluation procedure:

- The median of the individual assigned values, and
- the median of the widths of the pertaining uncertainty intervals.

These objective quantities are judged by means of criteria concerning robustness and concordance. In order to visualize this judgement both objective quantities are plotted against I and J separately for K = 1 and K = 2. We perform two steps.

Step 1

In order to find an appropriate *evaluation procedure* the totality of results underlying each plot of each figure is scored as follows:

Score 2

The simulation results show high robustness i.e. they depend on the respective design as little as possible indicated by a small variation. Moreover, they are well concordant with the reference assigned value or the width of its uncertainty interval.

To be specific, for all designs larger⁷⁾ than 615, with one exception at most, the following condition is to hold: The median of

the individual assigned values differs from the reference assigned value by less than one in the last relevant decimal place given by the respective analytical method. As to the widths, the median of the widths of the pertaining uncertainty intervals differs from the width of the uncertainty interval of the reference assigned value by about 8% or less.

Score 1

The simulation results show sufficient robustness for most designs larger than 615 whereas their concordance may be less.

Score 0

The simulation results do not show sufficient robustness for designs larger than 615.

The scores are summed up over all constituents. An evaluation procedure yielding a distinctly larger sum of scores both for K = 1 and K = 2 is better than one with a smaller sum⁸⁾.

To summarize, each plot is judged as a whole. An appropriate score is attached to the sector of the plot defined by the design 615.

Step 2

Conversely, in order to find an *optimized design* we look for a sector defined by D* and having score 2 with the modification that 615 is replaced by D*. Such a design D* is said to have sufficient sample size. If in the case of K = 1 the assigned value of the design I_{max} 115 differs from the reference assigned value by at least one in the last relevant decimal place, a design D* is also said to have sufficient sample size if the following modified condition holds: The medians of the individual assigned values of all designs larger than D* differ from the median of the individual assigned values of D* by less than one on the last relevant decimal place for all but one. An analogous condition holds to the widths.

In other words, for each constituent, each K, and each objective quantity a design D* is said to have sufficient sample size if

- all larger designs give approximately the same result, with no more than one exception, and if secondarily
- the assigned value of D* or the width of its uncertainty interval is concordant with the reference assigned value or with the width of its uncertainty interval as closely as possible.

⁸⁾ Since the scores are given to the whole plot the selection of a procedure is uniformly valid for all designs larger than 615.

⁵⁾ The definition of the binomial coefficient is

$$\binom{a}{b} = \frac{a \cdot (a-1) \cdot \dots \cdot (a-b+1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot b}$$

see e.g. l.c. (8).

⁶⁾ Mathematically speaking, the total set is the union of the two sets.

⁷⁾ See the discussion for the distinction of 615.

There are always several designs D^* of sufficient sample size. For each constituent and each K , those D^* for which a smaller design of sufficient sample size does not exist are called minimal. It must be noted that there may exist several minimal D^* . Therefore, we try to find an optimized design by selection of a minimal design out of the set of all designs with sufficient sample size. Naturally, such a minimal design causes as little as possible experimental expenditure.

The following connection holds between step 1 and step 2: If the score 2 is attached in step 1 then there is a D^* equal to or smaller than 615. The converse need not be true.

Results

Figures 1 to 7 show the first results of the simulation; these are the medians of individual assigned values depending on the respective design.

The scale of the respective figures is given in table 2. Generally, the results obtained by procedure 1 show less robustness and less concordance than those given by the procedures 2 and 3. Therefore, the assigned values according to procedure 1 are less stable with

Tab. 2. Scale of figures 1 to 7.

Constituent	Scale %	Reference assigned value			Unit
		Pro- cedure 1	Pro- cedure 2	Pro- cedure 3	
Creatinine	1.0 ^{a)}	310	310	309	$\mu\text{mol/l}$
Glucose	1.0	12.2	12.2	12.2	mmol/l
Urea	1.0	12.9	13.1	13.1	mmol/l
Alanine amino- transferase	1.0	63.1	63.1	63.1	U/l
Aspartate amino- transferase	1.0	40.1	40.0	40.1	U/l
Creatine kinase	6.0	230	226	223	U/l
γ -Glutamyltrans- ferase	1.0	47.6	47.7	47.7	U/l

a) The distance between the lower boundary of the red net and the black basic line corresponds to 1.0% of the reference assigned value. Note that the black crossing points represent medians of T_t values, see table 1.

respect to precision and accuracy so that procedure 1 turns out to be less appropriate. Procedure 2 and procedure 3 only remain as alternatives.

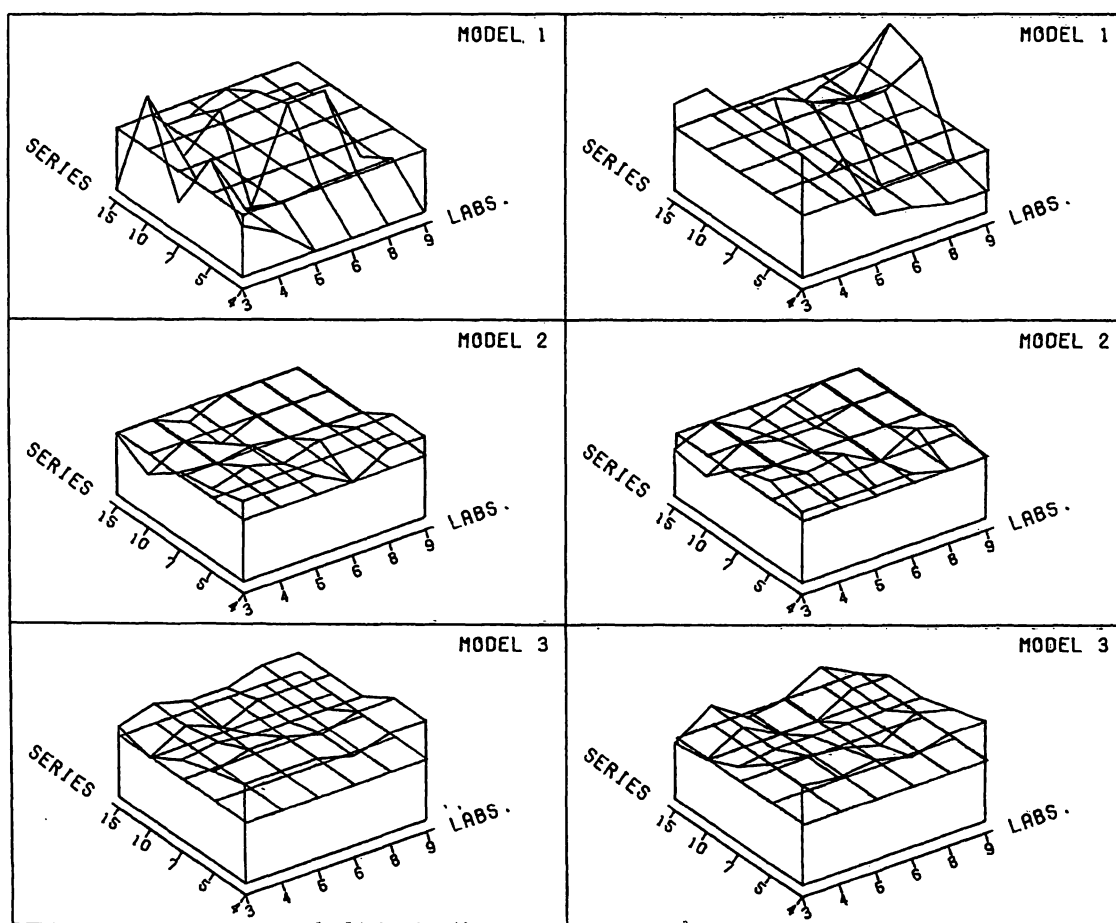


Fig. 1. Creatinine: Medians of individual assigned values.

The left column of plots is based on the first values only ($K = 1$), the right one on both values ($K = 2$) of the double determinations, and the lines correspond to the three evaluation procedures. The medians computed from the total set of individual randomization steps are plotted three-dimensionally in black against the respective design i.e. the number of laboratories and of series. Extreme medians are truncated at \pm scale, see table 2. The red net gives the reference assigned value as a reference coordinate. Therefore, for $K = 2$ the black and the red point coincide in the upper edge.

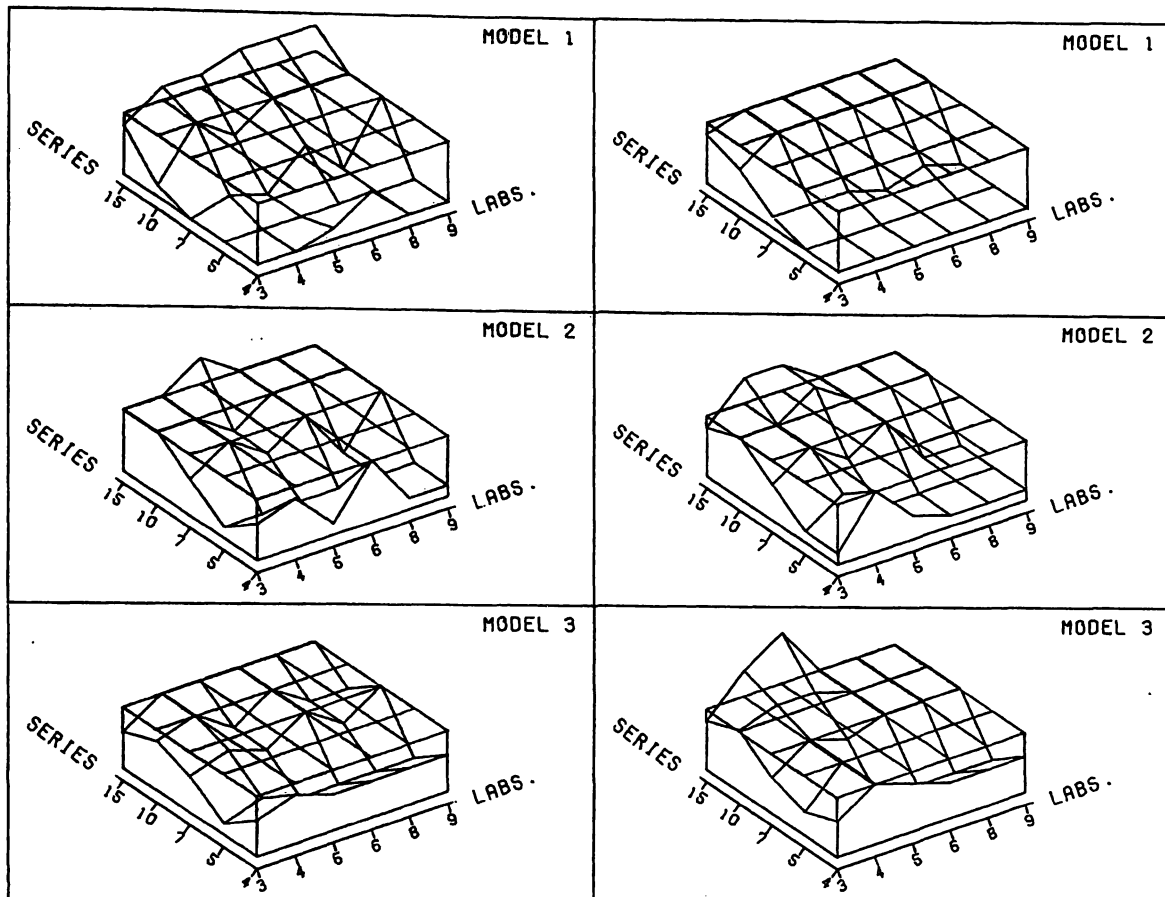


Fig. 2. Glucose: Medians of individual assigned values. For details see figure 1.

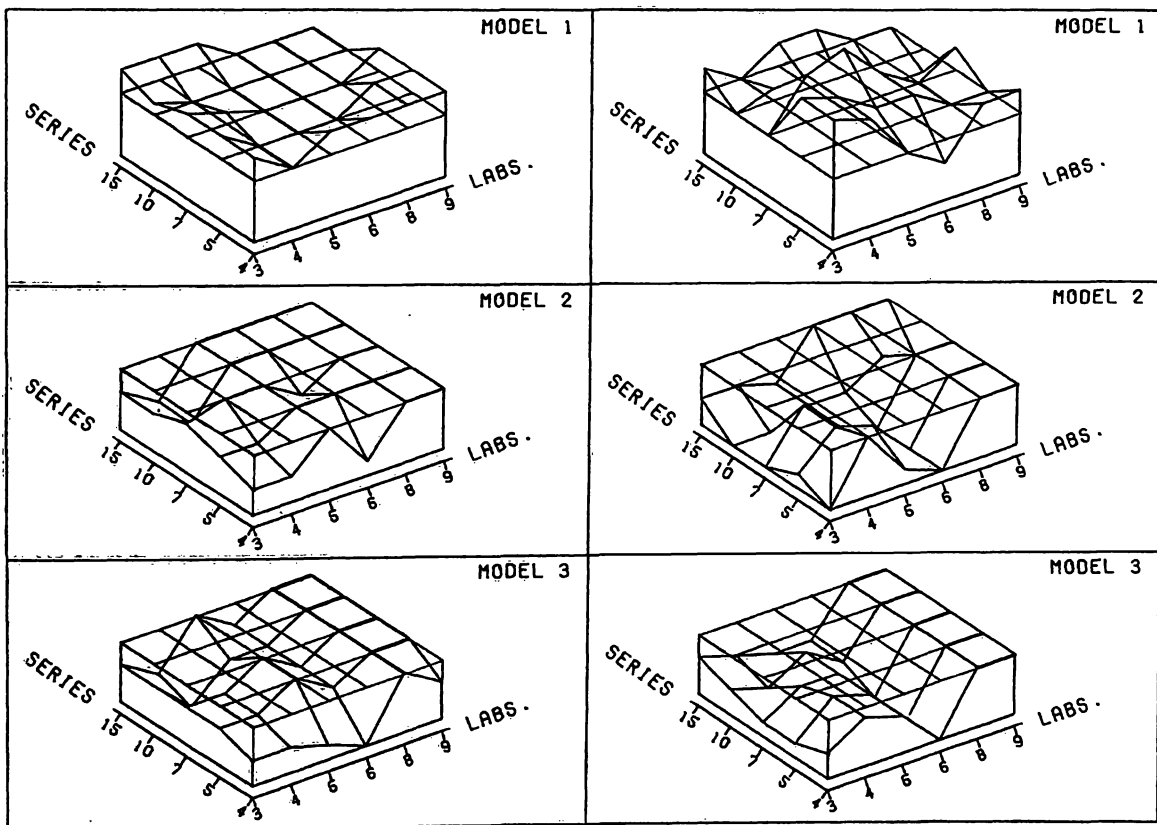


Fig. 3. Urea: Medians of individual assigned values. For details see figure 1.

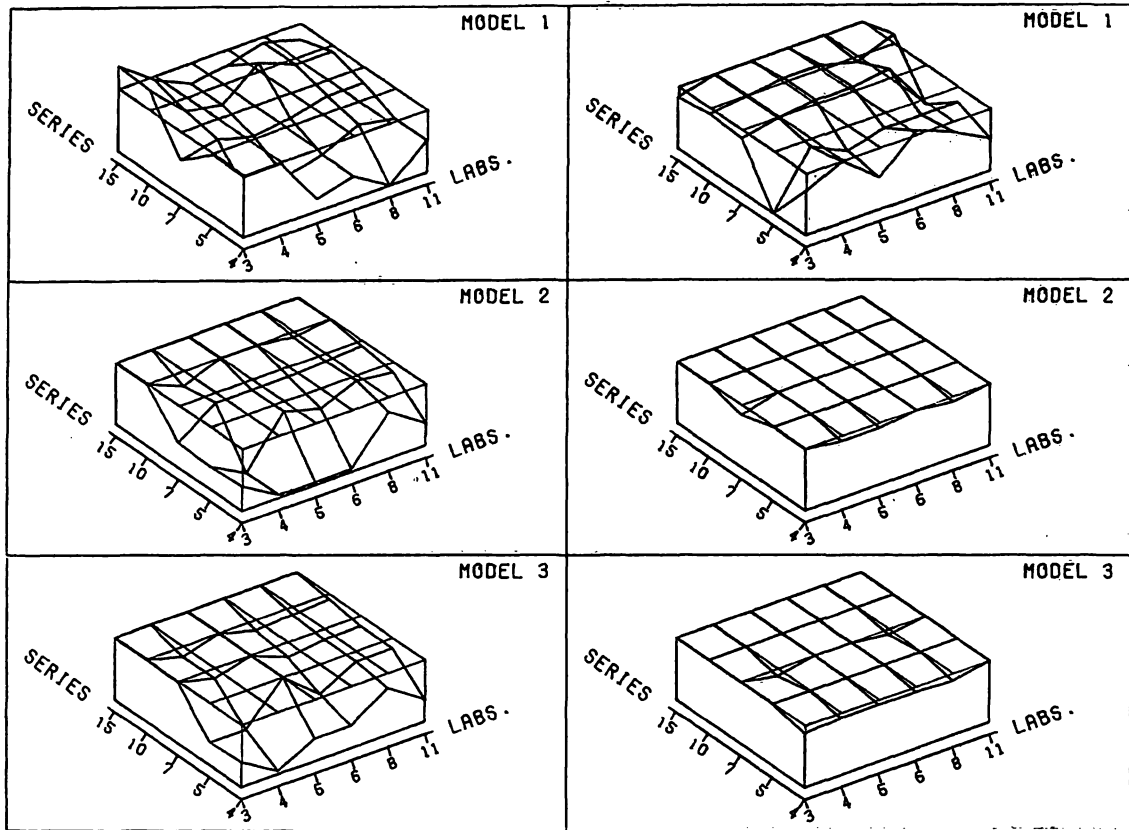


Fig. 4. Alanine aminotransferase: Medians of individual assigned values. For details see figure 1.

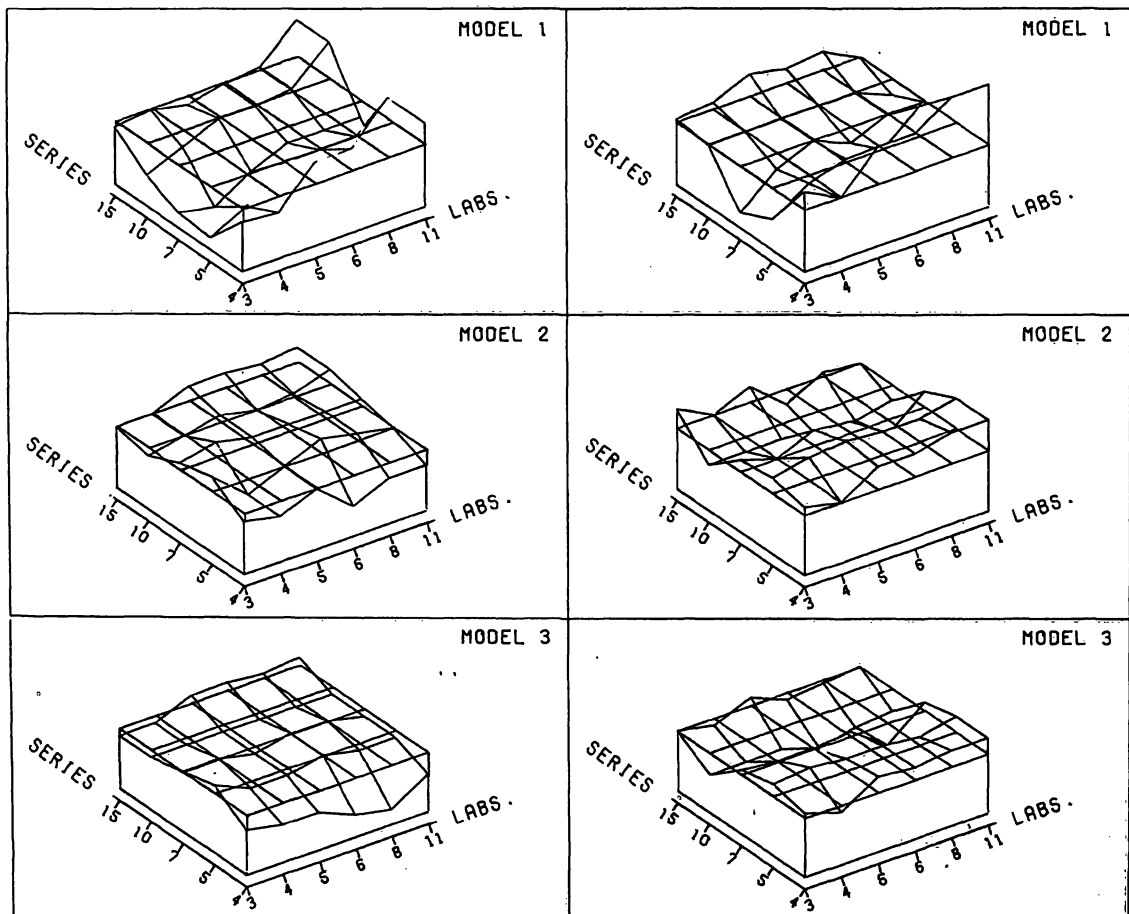


Fig. 5. Aspartate aminotransferase: Medians of individual assigned values. For details see figure 1.

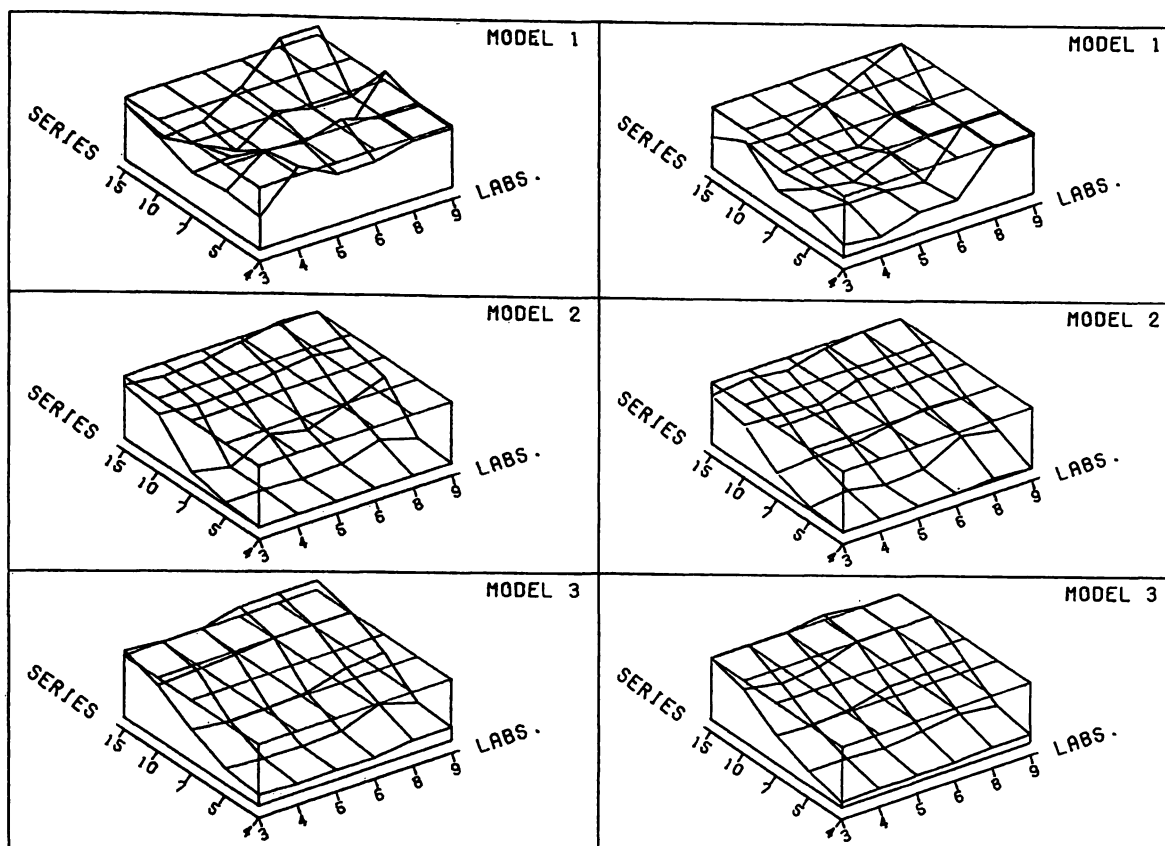


Fig. 6. Creatine kinase: Medians of individual assigned values. For details see figure 1.

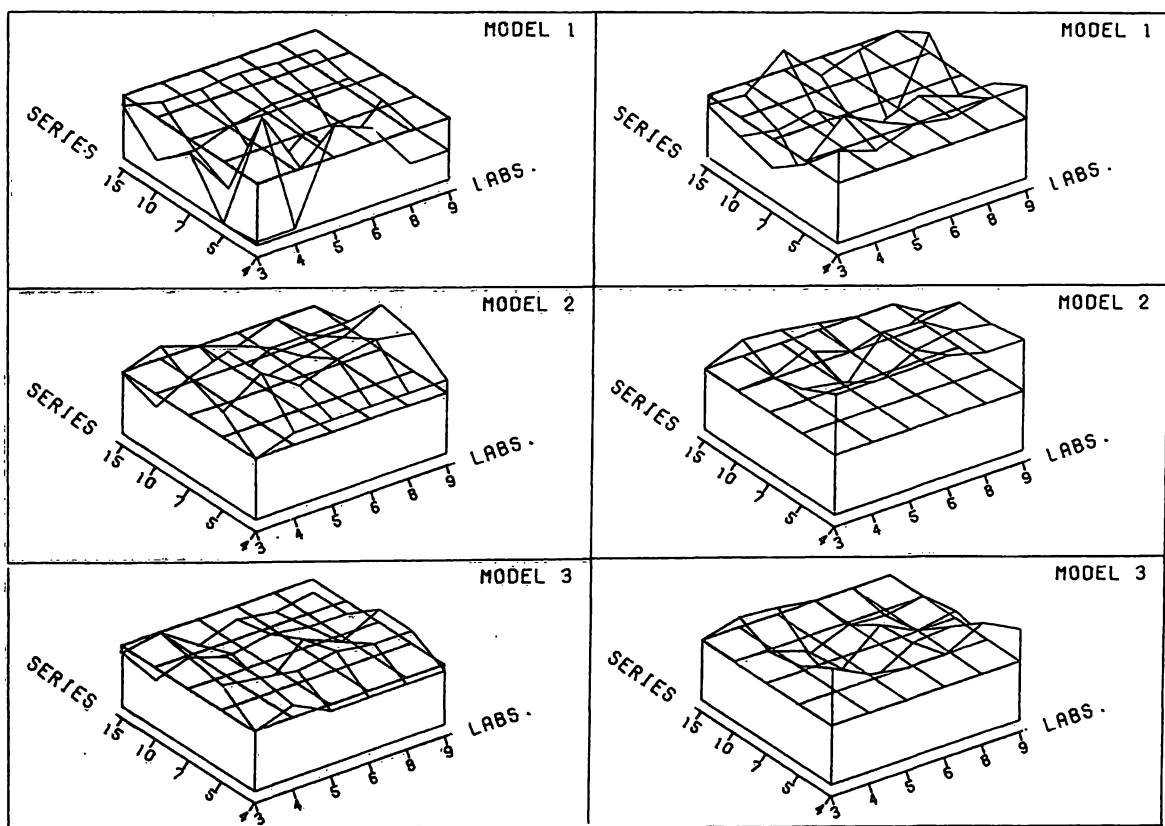


Fig. 7. γ -Glutamyltransferase: Medians of individual assigned values. For details see figure 1.

The judgement of procedure 2 is now described in detail in the case of creatinine. Table 3 contains the medians of individual assigned values when using the first values only. As the reference assigned value is 310 $\mu\text{mol/l}$ three designs (6|5 and 8|5 and 9|5) differ by 1 $\mu\text{mol/l}$ from the reference assigned value whereas the other designs larger than 6|5 have a corresponding median. Therefore, the score 2 cannot be attached whereas the score 1 is appropriate.

The results of judgement of all plots according to procedure 2 and procedure 3 are given in table 4. It should be stressed that the plots of each figure are based on the same randomizations. Therefore, for fixed I, differences between plots are not caused at random, but they are completely due to the different evaluation procedures, and to K.

The score 2 is not assigned in certain cases. In doing so the constituents are weighed. Procedures 2 and 3 cannot be discriminated here, neither for K = 1 nor for K = 2. Moreover, the respective reference assigned values are well concordant with each other, see table 2.

Figures 8 to 14 show the other results of the simulation that are the medians of widths of the pertaining uncertainty intervals, depending on the respective design.

The scale of these figures is given in table 5.

The results of the judgement of procedures 2 and 3 both for K = 1 and K = 2 are given in table 6.

Tab. 3. Medians of individual assigned values for creatinine (according to procedure 2 by use of the first values only), in $\mu\text{mol/l}$.

Number of series selected	Number of laboratories selected					
	3	4	5	6	8	9
15	310	310	309	310	310	310
10	309	310	309	310	310	310
7	310.3	310	310.3	310	310	310
5	310.5	311	310.5	311	311	311
4	311	311	311	310	311	311

The results of all designs covering at least 6 laboratories and 5 series are framed.

Tab. 4. Scores of plots of figures 1 to 7a).

Constituent	Procedure 2		Procedure 3	
	K = 1 ^b	K = 2 ^b	K = 1	K = 2
Creatinine	1	2	0	0
Glucose	0	0	0	1
Urea	2	0	2	0
Alanine amino-transferase	1	2	0	2
Aspartate amino-transferase	1	0	2	0
Creatine kinase	0	0	0	1
γ -Glutamyltransferase	0	0	0	1
Total	5	4	4	5

a) The meaning of the scores is given in Materials and Methods.
b) Using first values only (K = 1) or both values (K = 2) of the double determinations.

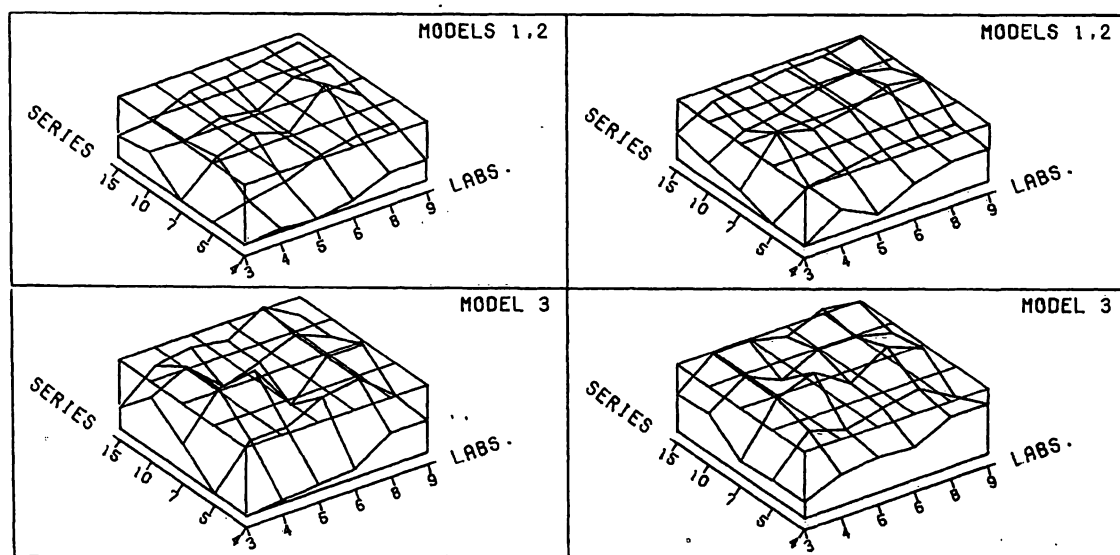


Fig. 8. Creatinine: Medians of widths of uncertainty intervals of assigned values.

For details see figure 1, the scale is given in table 5. However, the red net gives the width of the uncertainty interval of the reference assigned value according to the respective procedure so that for K = 2 the black and the red point coincide in the upper edge. Its height is identical for plots within lines but not necessarily within columns. The width according to procedure 3 may be higher than according to procedure 2 but not smaller.

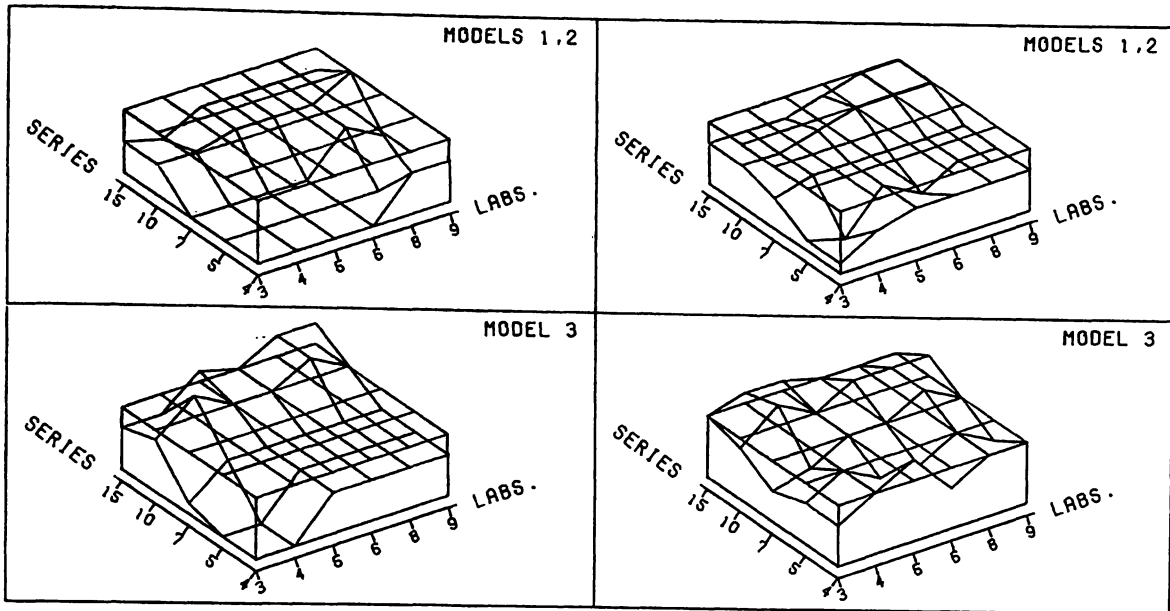


Fig. 9. Glucose: Medians of widths of uncertainty intervals of assigned values. For details see figure 8.

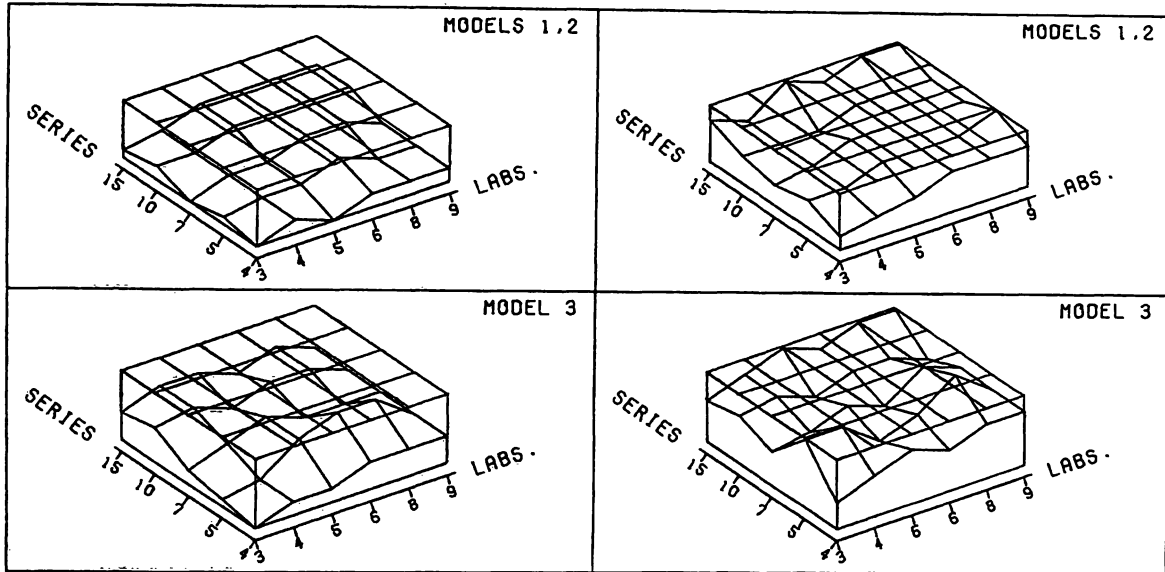


Fig. 10. Urea: Medians of widths of uncertainty intervals of assigned values. For details see figure 8.

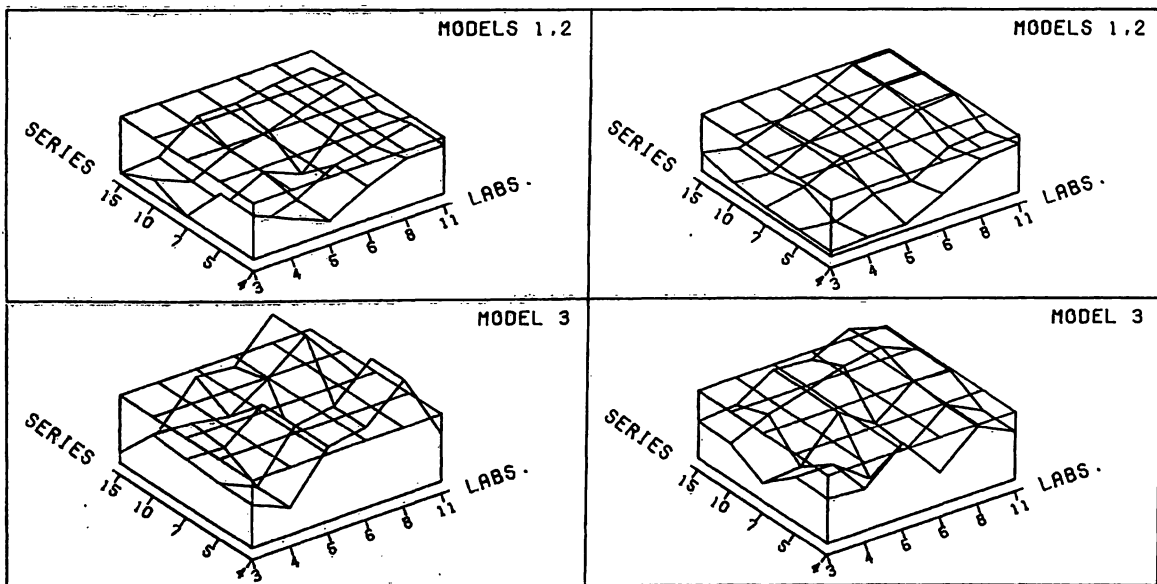


Fig. 11. Alanine aminotransferase: Medians of widths of uncertainty intervals of assigned values. For details see figure 8.

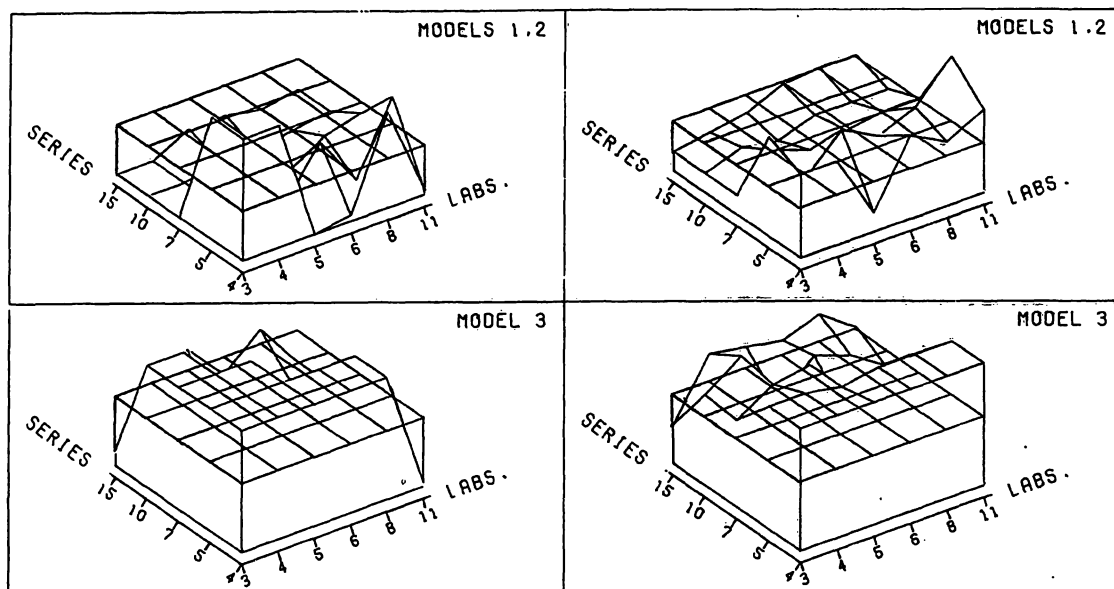


Fig. 12. Aspartate aminotransferase: Medians of widths of uncertainty intervals of assigned values. For details see figure 8.

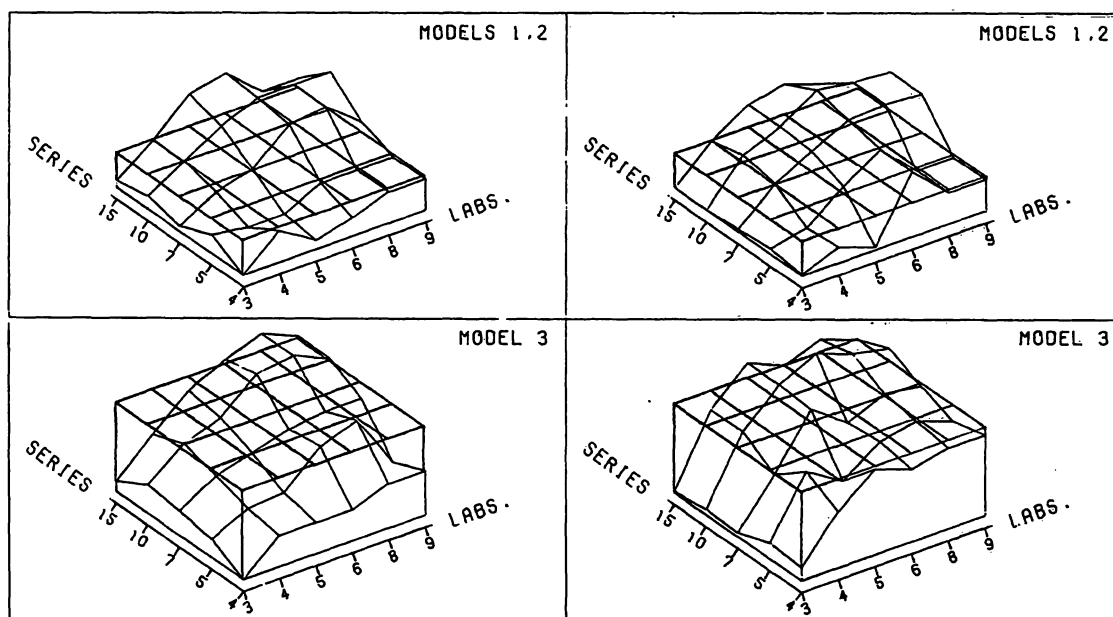


Fig. 13. Creatine kinase: Medians of widths of uncertainty intervals of assigned values. For details see figure 8.

Tab. 5. Scale of figures 8 to 14.

Constituent	Δ ^{b)}	Scale ^{b)}	Width ^{a)}		Unit
			Pro- cedure 2	Pro- cedure 3	
Creatinine	40	20.5	59	62	$\mu\text{mol/l}$
Glucose	1.1	0.3	1.4	1.4	mmol/l
Urea	1.3	0.45	1.7	1.8	mmol/l
Alanine amino- transferase	7.0	2.0	8.8	9.2	U/l
Aspartate amino- transferase	4.0	0.95	4.8	5.1	U/l
Creatine kinase	64	23.5	77	98	U/l
γ -Glutamyl- transferase	3.8	1.55	4.9	5.8	U/l

a) width of uncertainty interval of reference assigned value.
 b) the plotted quantity is $(\text{width} - \Delta) / \text{scale}$ giving the same scale to both procedures. Note that the black crossing points represent medians of T_i values, see table 1.

Tab. 6. Scores of plots of figures 8 to 14^{a)}.

Constituent	Procedure 2		Procedure 3	
	K = 1	K = 2	K = 1	K = 2
Creatinine	1	2	1	2
Glucose	1	2	1	2
Urea	1	2	0	0
Alanine amino- transferase	1	2	0	1
Aspartate amino- transferase	1	2	0	0
Creatine kinase	0	0	1	1
γ -Glutamyl- transferase	1	0	0	0
Total	6	10	3	6

a) see table 4 for explanations.

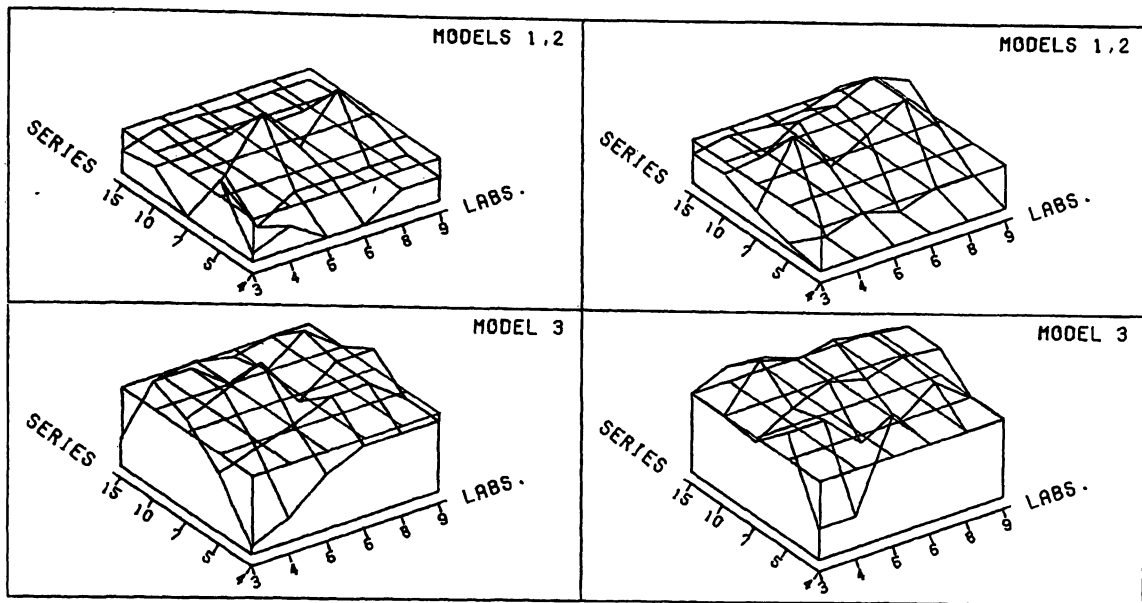


Fig. 14. γ -Glutamyltransferase: Medians of widths of uncertainty intervals of assigned values. For details see figure 8.

Therefore, procedure 2 is superior to procedure 3 for both K. Consequently, procedure 2 is selected from our three procedures, and hereafter we restrict ourselves to it when looking for an optimized design.

Table 7 is derived from figures 1 to 14 and gives minimal designs of sufficient sample size in the respective case according to procedure 2.

In particular, in the case of creatinine all designs larger than $D^* = 6 | 7$ give a median of individual assigned

values of $310 \mu\text{mol/l}$ corresponding to the reference assigned value when using the first values only (cf. tab. 3). Additionally, $D^* = 3 | 15$ has also sufficient sample size though the design $5 | 15$ being larger than $3 | 15$ differs by $1 \mu\text{mol/l}$ from the reference assigned value. But by our criterion one such exception is admitted.

It should be stressed that for fixed I and for each constituent the same randomizations are used. Therefore, variations between results of different designs may be

Tab. 7. Minimal designs D^* of sufficient sample size according to procedure 2.

Constituent	Deviation ^{a)}		Width ^{b)}	
	K = 1 ^{c)}	K = 2 ^{c)}	K = 1	K = 2
Creatinine	6 7 ^{d)} or 3 15	4 7 or 6 4 ^{**e)}	5 10 or 8 5	4 7 or 5 5 ^{**} or 6 4 ^{**}
Glucose	3 7	3 10	3 10	3 10 or 4 7 or 5 4 ^{**}
Urea	5 4 ^{**e)}	5 15 or 8 4	3 10 or 4 4 ^{*f)}	5 5 ^{**} or 6 4 ^{**}
Alanine aminotransferase	3 10 or 8 7	3 4 ^{**}	6 4 [*]	6 5 ^{**}
Aspartate aminotransferase	4 7 or 5 5 ^{*f)}	6 15 or 11 10	3 10 or 4 7	4 5 ^{**}
Creatine kinase	3 10	5 10	6 15	6 15
γ -Glutamyltransferase	6 10	6 10	4 10 or 6 5 [*] or 8 4	5 15

a) deviation from reference assigned value.

b) width of uncertainty interval.

c) see table 4 for explanations.

d) 6 laboratories with 7 independent series each.

e) design is smaller than 6 | 5 (see text for definition) in accordance with score 2 in tables 3 or 5.

f) design is smaller than 6 | 5. The score 1 in tables 4 or 6 is caused by worse correspondence with the red net.

influenced at random only in cases of differing numbers of laboratories.

No design is distinguished here either in the case of $K = 1$ or of $K = 2$. It is a well known fact (3) however that generally the variation between laboratories is larger than between series which is again larger than the variation within series. Finding the number of laboratories is therefore the most sensitive and thus the most important step when searching for an appropriate design. That is why this step must be performed at first.

From table 7 it can be stated that there does exist a design of sufficient sample size requiring 6 laboratories at most. This statement holds for both objective quantities and both cases of K for every constituent.

Discussion

The T_t randomization steps are performed using the given $I_{\max} = 9$ or 11 laboratories repeatedly so that the results of all randomization steps are not independent. Table 8 shows the expected number of joint laboratories if I laboratories are selected from I_{\max} laboratories. This number increases rapidly if I is enlarged. This fact causes an increasing dependence of the individual results. The range of the distribution decreases systematically if I is enlarged whereas its median is expected to be estimated correctly for all I . That is why we studied the medians only, whereas the ranges of respective distributions were neglected.

When establishing an assigned value in practice however a single value is produced for each constituent, and not a median of a set of values. But a single value has a larger variation than a median. Both variations differ by $\sqrt{T_t}$ at most⁹⁾ (see tab. 1). This factor depends on the number I of laboratories selected, whereas it is constant for all numbers J of series, both K i.e. single

Tab. 8. Expected number of joint laboratories if I from I_{\max} laboratories are selected repeatedly.

I	I_{\max}	
	9	11
3	1.0	0.8
4	1.8	1.4
5	2.8	2.3
6	4.0	3.3
7	5.4	4.4
8	7.0	5.8

⁹⁾ All results obtained are simulated from a pool of I_{\max} laboratories. Therefore the results are not independent and the variation of a median need not follow the $\sqrt{T_t}$ - law. Rather, it is likely that for T_t larger than a certain relative low threshold the variation of a median does not depend on T_t any more.

as well as double determinations, and for all evaluation procedures for fixed I .

When comparing scores in order to distinguish a *procedure* the factor $\sqrt{T_t}$ is irrelevant because it is independent of the respective plot. Therefore, the distinction of procedure 2 remains valid even for an actual establishment of assigned values, though it has been derived originally from medians.

However, the factor $\sqrt{T_t}$ should be taken into consideration when individual plots are studied in order to distinguish a *design*. For $I < I_{\max}$ they differ at most by factor $\sqrt{2} \approx 1.4$ for creatinine, glucose, urea, creatine kinase and γ -glutamyltransferase, and by $\sqrt{3} \approx 1.7$ at most for alanine and aspartate aminotransferase. When planning the simulation we had to select from two contradicting strategies:

Either to perform an equal number of randomizations for each I , or to have constant proportions both as far as actually permitted by combinatorics. The first strategy would result in different proportions relative to $\binom{I_{\max}}{I}$ and thus produce another kind of variation

between designs of different I concerning representativity of samples of randomizations. The latter one however would result in unequal numbers T_t and the mentioned consequences. We preferred our strategy of constant proportions because its influence on the results can be assessed more easily: Designs D which are larger than the designs D^* of sufficient sample size given in table 6 approximately yield the same result as D^* , though they are often based on a smaller number of randomizations than D^* itself. This holds especially for the designs D with $I = I_{\max}$ which are based on one randomization only. Therefore, if these designs D would be based on the same number as D^* it could be expected that their results would also be approximately concordant with those of D^* .

In using medians we have seen from our study that up to 6 laboratories are required. It can be concluded that 6 laboratories will in general be sufficient for an actual establishment of assigned values.

The further characteristics of an optimized design cannot be derived from figures 1 to 14. Rather, we need some additional information. Since we restrict ourselves now to the fixed number of 6 laboratories the dependence between the individual randomization steps caused by repeated randomization out of the given I_{\max} laboratories is identical. Therefore, the complete distributions of all individual randomization steps may be compared. This will lead to the optimized design (7): It covers 6 laboratories each performing 5 series.

The design 6|5 is the limiting one in the criteria for judging when we distinguish a procedure. If we had started from a design D_0 other than 6|5 for judging

we would not have attained a clear decision between procedure 2 and 3 except D_0 would have been larger than 6|5. In this case we would have also chosen procedure 2.

Certain laboratories are selected out of a set of laboratories for an actual establishment of assigned values. We simulated this by random selection of laboratories from our pool of I_{\max} laboratories. Furthermore, each laboratory performs analyses in a natural order and the result of each analysis may be influenced by the previous ones. That is why we did not randomize the series but used them according to their given order. Finally, we used the given analytical data only and we did not perform any interpolation in order to simulate a continuous distribution. Thus it was unnecessary to assume a distribution of the data for the sake of the performance of the simulation. The disadvantage however is that single extreme values which have not been eliminated by means of the controls may bias the results more easily.

Originally, we had performed one set of randomizations only for each constituent. In order to check whether the proportion out of all randomizations existing in total was sufficiently large we performed the second set. The results of both sets corresponded very well with each other so that the randomization error due to sampling was sufficiently small and the first set was large enough⁹). We have based our results on both sets for the sake of completeness only.

It is an elementary statistical fact that the precision of estimations increases if the number of underlying values is enlarged – provided that all values are independent and identically distributed. Therefore it is obvious to take the assigned value which is based on all namely $2 \cdot 15 \cdot I_{\max} \geq 270$ values as a reference quantity. Actually, these assumptions need not be fulfilled and certain values may randomly bias the reference assigned value or the results obtained from smaller designs. Especially, the results for $K = 1$ need not be concordant with those for $K = 2$, see figure 5. That is – when judging the figures – why we stress more the robustness than the concordance with the red net, and why the results of one design may be neglected. On the other hand, the criterion of concordance is essential too, because it reflects the accuracy of results obtained from smaller designs relative to the largest design enclosed in the study.

For all three procedures the width of the respective uncertainty interval is expected to increase if a relatively small number n of values involved is enlarged. This theoretical aspect is immediate from the definition of how to truncate 5% of the values. But for large n the widths of the uncertainty intervals are expected to converge to a limiting width. The width of the uncertainty interval of the optimal assigned value, which is based on

at least 270 values, is considered as limiting width and therefore also used as a reference quantity. The fact that the median of widths of many designs, especially of the small ones, is below the red net is now explained as a systematic effect which must be expected. Moreover, small uncertainty intervals are not representative for the analyses in the customer's laboratory. Therefore, the comparison with the red net is essential in order to ensure that the width is large enough. In fact this is more important than the comparison of the assigned values themselves with the red net because of the systematic aspect. Yet, the width of the uncertainty interval of the reference assigned value or of a smaller design is influenced randomly to a much higher degree than the assigned values themselves, see figures 10, 11, 14. We cover these two opposite facts by giving less weight to the criterion of concordance than to robustness, just as above.

There are also theoretical arguments supporting the distinction of procedure 2, which was selected on the basis of our experimental data.

First. The assigned value of procedure 1 is defined as the mean of two extreme values which are naturally not robust. Therefore the assigned value is extremely sensitive as shown by the simulation. Furthermore, procedure 1 gives a symmetrical uncertainty interval which is justified only if the distribution of the underlying analytical data is symmetrical. This assumption generally fails (3, 4). Moreover, an erroneously assumed symmetry of the uncertainty interval may mislead the customer as discussed in l.c. (3).

Secondly. The assigned values of procedures 2 and 3 must approximately be concordant with each other because they are both medians of remaining values derived from the same set of analytical data being differently truncated only. This is why neither procedure can be distinguished using assigned values only. On the other hand, the different techniques concerning the elimination of extreme values either do (procedure 2) or do not (procedure 3) reflect the fact that the underlying distributions are skew. Therefore, procedure 2 must be superior.

Acknowledgments

This study was initiated by the Committee of Clinical Chemistry of the VDGH.

The author thanks *B. Müller* (Behringwerke AG) for many hints, and for assistance in writing this paper, and to *W. Bablok* and *M. Glocke* (Boehringer Mannheim GmbH) for their cooperation in solving statistical problems. The determinations of the cited constituents were performed by *G. von Seydlitz* (Asid Bonz und Sohn GmbH), *A. Berke* (Boehringer Ingelheim Diagnostika GmbH), *A. Benozzi*, *F. Bosslet*, *H. Brettschneider* (all Boehringer Mannheim GmbH), *W. von Thun* (Goedecke AG), *H.-G. Eisenwiener* (F. Hoffmann – La Roche & Co. AG), *B. Carl* (Dr. Bruno Lange GmbH), *F. Schindler* (E. Merck), *R. Spaethe* (Merz und Dade, Abteilung der American Hospital Supply Deutschland GmbH), and *D. Schlitzer* (Technicon GmbH). The consequences from the results were drawn jointly by all participants.

References

1. Passing, H., Glocke, M., Brettschneider, H. & Müller, B. (1980), *Lab. Med.* **4**, 154–159.
2. Passing, H., Müller, B. & Brettschneider, H. (1981) *J. Clin. Chem. Clin. Biochem.* **19**, 1137–1144.
3. Passing, H. (1981) *J. Clin. Chem. Clin. Biochem.* **19**, 1145–1151.
4. Hansert, E. & Stamm, D. (1980) *J. Clin. Chem. Clin. Biochem.* **18**, 461–490.
5. National Committee for Clinical Laboratory Standards (NCCLS), *Calibration reference materials and control materials in Clinical Chemistry* (1975).
6. Schumann, V. (1976) *Med. Labor.* **29**, 271–276.
7. Passing, H., Bablok, W. & Glocke, M. (1981) *J. Clin. Chem. Clin. Biochem.* **19**, 1167–1179.
8. Feller, W. (1968) *An introduction to probability theory and its applications*, Wiley, New York.

For reprint requests:
Geschäftsstelle des
Verband der Diagnostica- und
Diagnosticgeräte-Hersteller (VDGH)
Karlstr. 21
D-6000 Frankfurt 1