Efficiency of split-mouth designs

P. P. Hujoel[1,2] and W. J. Loesche[2]
[1]Department of Biologic and Material Sciences, School of Dentistry, The University of Michigan, Ann Arbor, MI 48109; [2]Department of Biostatistics, School of Public Health, University of Michigan, 109 South Observatory, Ann Arbor, MI 48109, USA

Abstract. The purpose of this paper is (1) to investigate the similarity of the amount, distribution, and, severity of periodontal disease of the within-patient experimental units, (2) to estimate the relative efficiencies of split-mouth designs when compared to whole-mouth designs, and (3) to discuss how stratification on initial pocket depth can result in large differences in the power of the test-statistics in the different disease categories. Periodontal disease characteristics are not always homogeneously distributed over the within-patient experimental units and this heterogeneity can reduce the efficiency of split-mouth designs. In particular, if analyses are stratified on initial pocket depth, sites with an initial probing depth deeper than 6 mm may be small in number and asymmetrically distributed when compared to sites with an initial probing depth less than 6 mm. This may result in large differences of the power of the test statistics among the different disease categories and should lead to a careful interpretation of the statistical significance tests. When disease characteristics are symmetrically distributed over the within-patient experimental units and a sufficient number of sites is present per experimental unit, the split-mouth design can provide moderate to large gains in relative efficiency. In the absence of a symmetric disease distribution, whole-mouth clinical trials may be preferable.

At least 11 different types of split-mouth design have been described in the periodontal literature. The design is characterized by the subdivision of the mouth into quadrants, posterior sextants, contra- or ipsi-lateral sextants or quadrants, or, (a) symmetric combinations of these. Subsequently, the 2, 3, or, 4 within-patient experimental units are (randomly) assigned to a range of 2 to 6 different periodontal treatment modalities. The expectation for these designs is that they provide a powerful tool for the comparison of periodontal treatments by increasing the efficiency of the statistical tests. However, minimal information is available as to how efficient split-mouth designs are, and, as to which design provides the most homogeneous experimental units and a maximum amount of information.

It has been suggested that the distribution of the quantity, location and severity of periodontal disease may be so un-equal (dissimilar) within a mouth that it forms a contra-indication for split-mouth designs (Imrey 1986). Bone loss in the maxilla tends to be more severe than in the mandible, except for the anterior region where the situation is reversed (Carranza 1984, Marshall-Day et al. 1949, Beagrie et al. 1962). Also, there may be a slightly higher tendency to gingivitis on the right half side of the arch than on the left half side (Suomi et al. 1968). Although randomization should balance out these asymmetries as sample sizes get larger, the noise or the experimental error of split-mouth designs may be needlessly augmented by the choice of heterogeneous experimental units.

A better perception of the efficiencies of the different types of split-mouth designs can lead to insights in the statistical results of clinical trials and help in the planning of future clinical trials. The purpose of this study was (1) to investigate the similarity and suitability of 7 different types of split-mouth designs and (2) to investigate the relative efficiencies of split-mouth designs as compared to whole mouth-designs, and (3) to evaluate the precision of the split-mouth designs in the different disease categories. The consequences of the statistical design for the interpretation of treatment comparisons in clinical trials are discussed. Validity issues of split-mouth designs, such as problems with carry-over effects, are not addressed in this paper.

Material and Methods
Subjects

A set of 69 patients with advanced adult periodontitis seen at the graduate periodontal clinic of the University of Michigan who participated in double blind randomized clinical trial were used to investigate the similarity and the suitability of the within-patient experimental units at baseline. 7 different types of split-mouth design were investigated: (1) ipsi-lateral quadrants (ILQ), (2) ipsi-lateral sextants (ILS), (3) contra-lateral quadrants (CLQ) (4) contra-lateral sextants (CLS) (5) quadrants (Q), (6) posterior sextants (PS), and (7) (all) sextants (S).

A subset of 38 patients who received uniform treatment (scaling and root planing) over all within-patient experimental units and were re-evaluated 6 weeks post-therapy were employed to determine the relative efficiencies of the split-mouth designs. This type of trial where all within-patient experimental units receive a similar treatment is com-

monly referred to as a uniformity trial (Cochran 1937). The material and methods will be presented in two parts: (a) investigation of similarity and suitability on the baseline data of 69 patients, and (b) investigation of the relative efficiencies of different types of split-mouth clinical trials on the sample of 38 patients.

Similarity and suitability of experimental units

The similarity of the experimental units at baseline was investigated with respect to (a) the amount, (b) the distribution, and (c) the severity of periodontal disease.

(a) The amount of periodontal disease within an experimental unit was represented by mean probing depth and attachment level measurements. All measurements were taken by a single examiner who had been periodically calibrated for consistency of measurements. Similarity of these measurements for the experimental units were investigated with a weighted two-factor analysis of variance: one random factor being the patient and a fixed factor being the experimental units. The variance component of patients (s_1^2) and of sites within patient (s_2^2) were determined with a variance components analysis and used to weigh the mean experimental unit values by the inverse of $s_1^2 +$

s_2^2/n, where n is the number of sites (Healy 1972, Osborn 1987).

(b) The distribution of periodontal disease within an experimental unit was represented by the variance of probing depth and attachment level measurements. The variance of the attachment levels tend to be close to zero with a horizontal attachment loss pattern and increase with an irregular pattern of attachment loss. Similarly, the variance of probing depth measurements tends to be close to zero when probing depth is uniformly distributed over the within-patient experimental unit, and tends to increase as probing depth measurements are unevenly distributed over the within-patient experimental units. Since both variables exhibited substantial skewness, the data were re-expressed with a natural log transformation. The similarity of these indexes for the experimental units were compared using a two-factor analysis of variance. A multivariate ANOVA was used to investigate the similarity of the joint distribution of the variability of probing depth and attachment level measurements.

(c) The severity of periodontal disease within an experimental unit was represented by the presence or absence of the two disease categories: pockets 4 to 6 mm deep, and pockets deeper than 6 mm. The binomial probability model was used in order to test the hypothesis

that the proportion of patients with asymmetric disease severity is significantly different from zero.

Criteria for determining the suitability of experimental units is difficult since different clinicians use different criteria to determine the suitability of a particular experimental unit for a particular treatment. Only one criteria, the extent of disease, will be presented by the number of patients where in at least one of the experimental units the maximum probing depth did not exceed 3, 4 or 5 mm.

All analyses were performed using the Generalized Linear Model (GLM) and the Variance Components (Varcomp) procedures in SAS (SAS Institute Inc.)

Uniformity trial

The data were stratified according to initial probing depth into three disease categories: category I: probing depth < 4 mm, Category II: probing depth 4–6 mm, category III: probing depth > 6 mm. Site-specific differences in probing depth and attachment level measurements were obtained and mean values calculated per type of experimental unit. These data were analyzed using a two-factor analysis of variance with a sums-of squares due to grouping and a sums of squares due to error. This analysis was not weighted by the number of sites

Table 1. Investigation of the homogeneity of the experimental units

Experimental design	sites[a]	Similarity in quantity and distribution					
		mean probing depth	mean probing attachment level	v (p)[b]	$v(a)$[c]	$v(p,a)$[d]
		P-values †	*P*-values	*P*-values	*P*-values	*P*-values
Ipsi-lateral quadrants (2[+])	35–80	0.0781* (0.10)	0.8869 (0.01)	0.1138	0.1891	0.2694
Ipsi-lateral sextants (2)	10–50	0.0255* (0.18)	0.2507 (0.11)	0.0359*	0.0061*	0.0215*
Contra-lateral quadrants (2)	40–80	0.8011 (0.01)	0.3264 (0.07)	0.8314	0.0143*	0.0043*
Contra-lateral sextants (2)	10–40	0.7609 (0.02)	0.4485 (0.07)	0.9893	0.2557	0.4025
quadrants (4)	5–40	0.0001* (0.37)	0.2345 (0.22)	0.0259*	0.0001*	0.0001*
posterior sextants (4)	5–25	0.0997* (0.27)	0.0001* (0.78)	0.0483*	0.0007*	0.0076*
Sextants (6)	5–30	0.0001* (0.91)	0.0001* (0.73)	0.0001*	0.0002*	0.0001*

+Number of within-patient experimental units. [a]Range of the number of sites per experimental units; [b]variability of probing depth measurements, [c]variability of attachment levels, [d]joint distribution of the variability of probing depth and attachment level measurements; † maximum mean differences between within-patient experimental units (mm).

Stars indicate significant differences among the within-patient experimental units. Since the purpose of these statistical tests is to investigate the homogeneity of the experimental units, 0.1 was selected as the significance level. Note that the CLS-design is the most homogeneous with respect to the presented parameters.

Table 2. Number of patients and experimental units where disease category II is distributed asymmetrically over the within-mouth experimental units

Experimental unit	Similarity in disease severity category II		
	% patients with asymmetric disease	% experimental units missing	range of # sites per experimental unit
Ipsi-lateral quadrants (2)[a]	0 (0/69)	0 (0/138)	5–41
Ipsi-lateral sextants (2)	0 (0/69)	0 (0/138)	5–31
Contra-lateral quadrants (2)	0 (0/69)	0 (0/138)	5–43
Contra-lateral sextants (2)	0 (0/69)	0 (0/138)	5–31
quadrants (4)	0 (0/69)	0 (0/276)	5–26
posterior sextants (4)	9* (6/69)	2 (6/276)	0–16
sextants (6)	14 (10/69)	3 (11/414)	0–19

[a] The number of within-patient experimental units. Stars indicate a proportion of patients significantly different from 0. The last column gives the range of the number of sites in disease category II.

to permit comparisons with earlier research. (Weighted analysis provided similar trends in the relative efficiency estimates for the different designs). All analyses were performed for the probing depth and attachment level measurements in the different disease categories and the overall mean values. 2 different error variances of whole-mouth designs were determined. One whole-mouth error variance was estimated excluding the anterior teeth and used for estimating the relative efficiency of the ILS, CLS, and PS design. The other whole-mouth error variance was estimated using the whole dentition and was used to estimate the relative efficiency of the CLQ- ILQ-, S-, and Q design. Relative efficiencies of the different designs were determined as the inverse ratio of the variance of the split-mouth divided by the inverse ratio of the variance of the whole mouth as an experimental unit. The precision of the split-mouth design in the different dis-

ease categories was estimated as the inverse of the error variance.

Results
Investigation of similarity

An evaluation of the 7 different types of split-mouth design revealed that significant differences between the experimental units existed with respect to the amount, distribution, and severity of periodontal disease. Significance levels are shown in Table 1. The amount of periodontal disease as quantified by probing depth and attachment level measurements differed significantly between the within-patient experimental units for the ILQ-, ILS-, Q-, and S-design. Pocket depth and attachment loss at baseline was significantly higher in the maxilla than in the mandible with the upper left quadrant being the worst.

The distribution of periodontal disease as measured by the joint variability of probing depth and attachment level

measurements was significantly different for most experimental units, except for the ILQ- and CLS-design. Pocket depth measurements and attachment level measurements were more variable in the maxilla than in the mandible with the upper left quadrant being the most variable.

An important aspect of the investigation of asymmetries was the determination of the percentage of patients where at least one of the experimental units did not have pockets of category II or III. An asymmetry of the severity of periodontal disease can have a large impact on the efficiency of the design due to the loss of error degrees of freedom. Pockets 4 to 6 mm deep (Category II) were symmetrically distributed for most types of split-mouth designs with exception of the PS- and S-design (Table 2). Pockets deeper than 6 mm (Category III) were distributed asymmetrically for a significant proportion of the patients for all types of split-mouth design. This is primarily due to an absence of teeth with this degree of morbidity in the within-patient experimental units. When the mouth is subdivided into more than two experimental, such as when the Q- or PS-design are used, over 50% of the patients have an asymmetric distribution of disease category III. In fact, when the S-design were used 83% of the patients had at least one sextant where the probing depth did not exceed 6 mm (Table 3).

Investigation of the suitability of the within-experimental units revealed that in 6 out of 69 patients, there was at least one posterior sextant with probing depths no deeper than 3 mm. 14 out of 69 patients had at least one posterior sextant with probing depths no deeper than 4 mm. Specifics are summarized in Table 4.

Table 3. Number of patients and experimental units where disease category III is distributed asymmetrically over the within-mouth experimental units

Experimental unit	Similarity in decrease severity, category III		
	% Patients with asymmetric disease	% experimental units missing	rang of # sites per unit experimental
Ipsi-lateral quadrants (2*)	17* (11/65)	8 (11/130)	0–33
Ipsi-lateral sextants (2)	19* (12/62)	10 (12/124)	0–26
contra-lateral quadrants (2)	15* (10/65)	8 (10/130)	0–36
contra-lateral sextants (2)	16* (10/62)	8 (10/124)	0–19
quadrants (4)	55* (36/65)	25 (64/260)	0–19
posterior sextants (4)	68* (42/62)	29 (71/298)	0–14
sextants (6)	83* (54/65)	34 (145/390)	0–14

*The number of within-patient experimental units. Stars indicate a proportion of patients significantly different from 0. The last column gives the range of the number of sites in disease category III.

Investigation of relative efficiency

Site-specific differences of probing depths and attachment level measurements in patient who were scaled and root planed were compared before and after treatment using each of the different split-mouth designs. The relative efficiencies of the various split-mouth designs when compared to whole-mouth designs ranged from 0.6 to 9.3 depending on the disease category studied (Table 5). For this set of patients, divisions of the mouth into 2 experimental units are more efficient than divisions of more than 2. When the mouth is

Table 4. Number of patients where at least one of the experimental units has pockets depths no deeper than 3, 4 or 5 mm

	Suitability		
	maximum probing depth per experimental unit		
Experimental design	3 mm	4 mm	5 mm
Ipsi-lateral quadrants (2)	0/69	1/69	5/69
Ipsi-lateral sextants (2)	0/69	2/69	7/69
Contra-lateral quadrants (2)	0/69	1/69	5/69
Contra-lateral sextants (2)	0/69	2/69	7/69
quadrants (4)	0/69	7/69	22/69
posterior sextants (4)	6/69	14/69	28/69
sextants (6)	10/69	27/69	41/69

Table 5. Results of the uniformity trials; relative efficiencies of different types of designs

	Relative efficiency							
	mean		<4 mm		4–6 mm		>6 mm	
Experimental design	p [a]	a [b]	p	a	p	a	p	a
Ipsi-lateral quadrants (2) [c]	7.3	3.6	4.4	5.1	3.2	1.9	0.8	2.0
Ipsi-lateral sextants (2)	2.5	1.3	1.8	0.9	1.9	1.3	0.8	1.2
Contra-lateral quadrants (2)	9.1	3.0	2.8	5.3	4.8	1.6	1.7	2.0
Contra-lateral sextants (2)	3.5	1.7	1.9	0.8	1.9	1.0	1.0	1.0
quadrants (4)	3.7	1.6	1.4	1.7	2.0	0.9	0.8	1.0
posterior sextants (4)	1.5	0.8	0.5	0.3	0.9	0.6	0.7	0.6
sextants (6)	1.8	0.9	0.5	0.9	1.0	0.5	1.0	1.1

[a] Probing depth measurements. [b] Probing attachment level measurements. [c] Number of experimental units.

The relative efficiency number indicates how many more replications one will need with a whole-mouth design as compared to a split-mouth design to get the same precision on any estimated treatment contrast. Note the inefficiency of designs where the mouth is divided into more than 2 experimental units.

divided into 2 experimental units, the CLQ and CLS design are generally more efficient than their ipsi-lateral counterparts, and, the CLQ design is more efficient than CLS-design. A surprising observation was the inefficiency of the PS-design. The precision differed largely among the different disease categories. Measurements in disease category III can be up to 60 times less precise than measurements in disease category I.

Discussion

The split-mouth design has been the principal research tool in periodontal clinical trials to compare different treatment modalities. The actual design of the study has varied widely, ranging from simple designs where 2 treatments are assigned (randomly) to ipsi-lateral quadrants, to incomplete block designs where 6 different treatment modalities are randomly assigned to quadrants. Surprisingly, little information is available as to how efficient these different designs are, and, to which extent ques-

tions raised about the similarity in the quantity, distribution, and, severity of the periodontal disease may affect these efficiencies.

Data from this set of patients, who were consecutive admissions to the graduate periodontal clinic, suggest that periodontal disease is not always a symmetrical disease. From the different types of asymmetries investigated, the most significant problem is the absence of lesions of disease category III in an appreciable fraction of the within-patient experimental units. A review of the literature indicates that this observation is not limited to this group of patients. One study, which required symmetric disease distribution, screened 1500 patients to find 12 patients with symmetric disease, indicating that only 1 in 125 patients exhibited symmetric disease (Smith 1982). This scarcity of symmetrically diseased patients may be one of the reasons why a symmetry of the periodontal lesions has not been required for entry into some studies. The absence of this entry criteria has several consequences: (1) there may be a loss of

efficiency of split-mouth designs, to the extent that a whole mouth design may be more efficient; (2) elaborate statistical methods may be required to estimate treatment differences for more complicated split-mouth designs with missing observations, since none of the well-known statistical packages will provide a correct analysis of variance (Milliken & Johnson 1984); (3) split-mouth designs can result in experimental units with minimal disease, which may be unsuitable to evaluate periodontal therapies.

The extent of disease necessary to justify periodontal therapy is a matter of debate. Where some clinicians find pockets of 4 mm suitable candidates for surgical correction, other clinicians may disagree (anonymous 1982). Other suitability criteria may be important in clinical periodontal research: different patterns of bone destruction may require different surgical management styles (Schluger 1979), different forms of periodontal disease may require different clinical managements (Page 1982), and, different bacterial infections may require different antimicrobial treatments (Loesche 1979). Clinicians and researchers have expressed their concern of these suitability issues and have suggested that results of studies may be falsely skewed against treatment modalities which are assigned to unsuitable experimental units (anonymous 1983). Periodontal disease should not be regarded as one uniform disease entity, neither within nor across patients. These concerns of the suitability and the similarity of the experimental units and of the suitability of the assigned treatment modalities emphasize the need for blinded randomization processes. In fact, absence of randomization or non-blinded randomizations may produce biases which could determine the outcome to a larger extent than the treatments themselves (Chalmers et al. 1983).

The interpretation of the statistical significance levels (*P*-values) of split-mouth clinical trials which have stratified their results according to initial pocket depth should be done with caution. Apart from problems with the choice of the primary statistical test (Hujoel & Moulton 1988), 2 other issues are of importance in the interpretation of *P*-values; (1) the number of sites measured per disease category; (2) the % of missing within-patient experimental units per disease category.
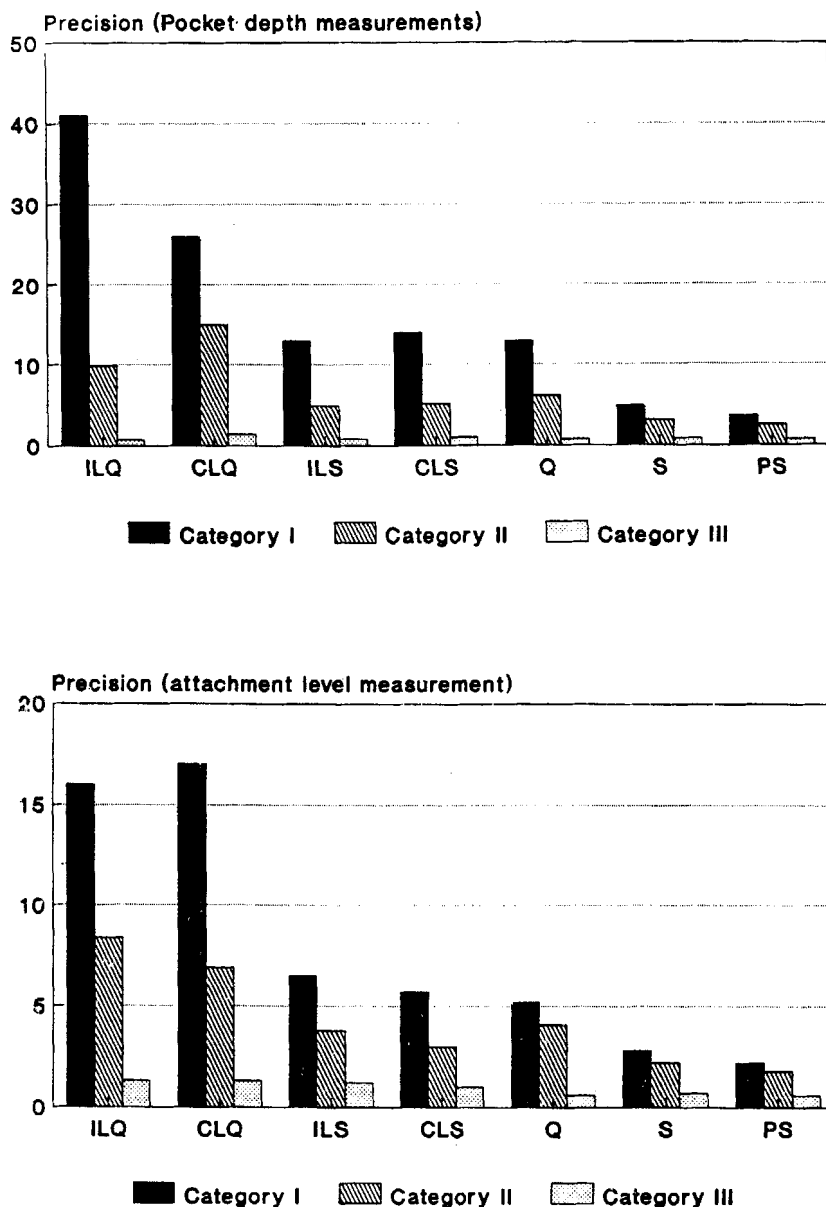
The larger the number of sites

Precision (Pocket depth measurements)

Precision (attachment level measurement)

Fig. 1. Precision of the different types of split-mouth design in the different disease categories. Note the precipitous drop in precision in disease category III.

sampled in a disease category, the more accurate the mean measurement will be. It commonly occurs that the number of sites sampled in the different categories decreases with increasing disease severity. For instance, of the sites measured in this study, 52% of the sites had an initial probing depth less than 4 mm, 38% had an initial probing depth between 4 and 6 mm, and, 10% of the sites had an initial probing depth greater than 6 mm. A decrease in number of sites is accompanied by an increase in the standard error, a decrease in the within-patient correlation coefficient of the experimental units (Fleiss 1988),

and, as a consequence a decrease in the power (efficiency) to detect treatment differences. As a result, if similar true treatment differences exist in the different disease categories, they will be most evident (significant) in the disease category with the largest number of sites. (This phenomenon is further accentuated due to the increase of measurement error in deeper pockets).

A second important design aspect of clinical trials whose result are stratified according to initial pocket depth is the percentage of patients with missing experimental units in the 3 different disease categories. Missing within-patient

experimental units in a large number of patients will result in a rapid decrease of the sensitivity of an experiment due to the loss of the error degrees of freedom. The % of missing within-patient experimental units tends to increase with increasing disease severity. If we take, for instance, quadrants as experimental units of analyses for this study, all patients had quadrants with pockets in the 1 to 6 mm range, but 55% of the patients had at least 1 quadrant without pockets greater than 6 mm. As a result, the split-mouth design will be efficient (powerful) in pockets less than 6 mm, and less efficient when pockets are greater than 6 mm. This indicates that, if true treatment differences are present, statistical tests are more likely to detect significant treatment differences in disease category I and II, than in disease category III.

These 2 aspects of the design of a study may be responsible for the finding of statistically significant results of clinically insignificant changes (e.g., 0.04 mm) in shallow pockets (where a large number of sites are sampled and minimal patients have missing within-mouth experimental units), and, statistically insignificant results for clinically significant changes (e.g., 1.25 mm) in deep pockets (where a small number of sites are sampled and a large % of patients have missing within-patient experimental units).

These problems often work in a synergistic fashion: the smaller the % of sites in disease category III, the more likely they will be distributed asymmetrically over the within-patient experimental units; the smaller the percentage of sites in category III, the larger the % of sites in category I and II, where they are most likely distributed symmetrically over the within-patient experimental units. This results in large differences in the power of the test statistics in the different disease categories. It is like using an light microscope for detecting differences between treatment modalities in shallow pockets, and using the naked eye for detecting differences between treatments in the deep pockets.

The loss of efficiency in disease category III can be minimized by proper selection criteria. In the split-mouth experimental design, the primary sampling unit is the patient and the secondary sampling unit is the split-mouth. Apart from patient-specific criteria, split-mouth specific selection criteria should be applied when selecting pa-

tients. For instance, if the purpose of the investigation is to evaluate the efficacy of periodontal treatments on sites with an initial probing depth deeper than 6 mm, every within-patient experimental unit should have at least one such site present. The effort of screening a large number of patients (if available) for obtaining patients with symmetric disease distribution in disease category III, must be weighted against the effort of performing surgeries and follow-up examinations on a larger sample of patients in whole-mouth designs. It should be realized however, that the use of split-mouth designs in patients with asymmetric disease distribution will probably be less efficient than a whole-mouth design, just as the use of a whole-mouth design will be less efficient than a split-mouth design when the disease is symmetrically distributed over the within-patient experimental units.

Results of this set of patients suggest that: (a) divisions of the mouth into 2 experimental units provide more efficient experimental designs than divisions of more than 2 experimental units, (b) the CLQ- and CLS-design is more efficient than their ipsi-lateral counterparts, (c) the CLQ-design is more efficient than the CLS-design, and (d) the CLS-design provides the most homogeneous experimental units. The selection of only 2 experimental units per patient does not preclude the investigation of more than 2 treatment modalities. Balanced incomplete block designs may provide an elegant means for eliminating the within-patient heterogeneity of experimental units to a larger extent than would be possible with randomized block designs.

Split-mouth clinical trials may under some circumstances provide suitable research designs for addressing more recent interests of the effects of anti-microbial and/or anti-inflammatory agents. There exists currently an interest as to how systemic and local periodontal treatment modalities interact. Multi-factorial experiments, where one or more systematic treatment factors are randomly assigned to patients, and certain types of surgical treatment modalities are randomly assigned to within-patient experimental units, may provide an efficient research design which permits inferences of the treatments with a greater range of validity (Neter 1985). It is concluded that split-mouth clinical trials may provide an efficient research tool for periodontal research,

provided that proper selection criteria are utilized. Validity issues such as carry-over effects should always be carefully evaluated. If patients with symmetric disease distribution are difficult to find, then a whole-mouth design may be more advantageous.

Zusammenfassung

Die Aussagekraft von Versuchsanlagen nach dem "Spalt-Mundprinzip" (split-mouth design)

Mit der hier vorliegenden Veröffentlichung wurde (1) beabsichtigt, die Vergleichbarkeit des Aussagewertes der Menge, der Verteilung und der Schwere der parodontalen Erkrankung in sog. Experimentaleinheiten am gleichen Patienten zu untersuchen, (2) die relative Aussagekraft von Spalt-Mund Versuchsanlagen (split-mouth designs) abzuschätzen und (3) zu besprechen, inwieweit eine, von initialen Taschentiefen ausgehende Resultatschichtung, die Trennschärfe der statistischen Bearbeitung deutlich beeinflussen kann. Die Parodontalkrankheit ist dadurch gekennzeichnet, dass sie in den experimentellen Einheiten eines Patienten nicht homogen verteilt vorkommt und diese Heterogenität kann die Aussagekraft von Versuchsanlagen nach dem Spalt-Mundprinzip reduzieren. Das gilt vor allem dann, wenn die Analysen von der initialen Taschentiefe ausgehen. Stellen mit einer initialen Sondierungstiefe von mehr als 6 mm, kommen selten vor und sie sind, im Vergleich zu den Stellen mit Taschentiefen von weniger als 6 mm, asymmetrischer verteilt. Das führt bei der statistischen Bearbeitung zu grossen Unterschieden der Trennungsschärfe zwischen verschiedenen Erscheinungsformen dieser Krankheit, was der Forderung nach vorsichtiger Deutung der Ergebnisse einer solchen Versuchsstatistik besonderen Nachdruck verleiht. Kommen dagegen die Krankheitssymptome in den experimentellen Einheiten des gleichen Patienten symmetrisch verteilt vor und ist die Menge der befallenen Stellen pro Experimentaleinheit ausreichend, kann eine Versuchsanlage nach dem Spalt-Mundprinzip (split-mouth design) mässige bis hohe Gewinne an relativer Aussagekraft erreichen. Es wird im allgemeinen empfohlen, nicht weniger als 2 Experimentaleinheiten am gleichen Patienten auszuwählen und kontralateral gelegene Kieferquadranten und -sextanten den ipsilateral gelegenen vorzuziehen. Liegt keine symmetrische Verteilung der Krankheit vor, empfielt es sich, klinische Versuche von den Verhältnissen im ganzen Mund ausgehen zu lassen.

Résumé

Efficacité des méthodes par bouches divisées

La but de la présente étude a été (1) d'analyser la similitude entre la quantité, la réparti-

tion et la sévérité de la maladie parodontale des unités expérimentales intra-individuelles, (2) d'estimer l'efficacité relative des méthodes par bouches divisées et (3) de discuter comment la stratification des résultats basée sur la profondeur de poche initiale peut s'accompagner de grandes différences dans la force des tests statistiques. Les caractéristiques de la maladie parodontale ne sont souvent pas réparties de manière homogène dans les unités expérimentales intra-individuelles, et cette hétérogénéité peut réduire l'efficacité des méthodes par bouches divisées. En particulier, si les analyses sont stratifées sur la profondeur de poche initiale, les sites avec une profondeur initiale supérieure à 6 mm sont rares et distribués asymétriquement lorsqu'ils sont comparés à ceux avec une profondeur initiale inférieure à 6 mm. Ceci résulte en grandes différences de force des tests statistiques parmi les différentes catégories de maladie, et doit donc s'accompagner d'une interprétation prudente des résultats statistiques. Lorsque les caractéristiques de la maladie sont réparties symétriquement parmi les unités expérimentales intra-individuelles et qu'un nombre suffisant de sites est présent par unité expérimentale, la méthode par bouches divisées peut s'accompagner de gains modérés à importants d'efficacité relative. En général, il est recommandé de ne pas utiliser plus de deux unités expérimentales intra-individuelles, les quadrants et sextants contralatéraux étant choisis plutôt que leurs équivalents unilatéraux. En l'absence de répartition symétrique de la maladie, des études cliniques portant sur l'ensemble de la cavité buccale sont préférables.

References

Anonymous. Emphasis (1982). Changing concepts in periodontal therapy, *Journal of the American Dental Association* **105**, 18–29.

Anonymous. Survey of current therapy (1983). Part II. *Periodontal Case reports* **5**, 14–22.

Beagrie, G. S. & James, G. A. (1962) The association of posterior tooth irregularity and periodontal disease. *British Dental Journal* **113**, 239–243.

Chalmers, T. C., Celano, P., Sacks, H. S. & Smith, H. (1983) Bias in Treatment Assignment in controlled clinical trials. *The New England Journal of Medicine* **309**, 1358–1361.

Cochran, W. G. (1937) A Catalogue of Uniformity trial data. Some uses of Uniformity trial data. *Journal of Agricultural Research*, Supp. **4**, 233–253.

Cochran, W. G. & Cox, G. M. (1957) *Experimental designs* New York: John Wiley & Sons.

Feinstein, A. R. (1985) *Clinical epidemiology. The architecture of clinical research*, 1st edition, p. 31. Philadelphia: Saunders.

Fleiss, J. L., Wallenstein, S., Chilton, N. W. & Goodson, M. (1988) A re-examination of

within-mouth correlations of attachment level and of change in attachment level. *Journal of Clinical Periodontology* **15**, 411–414.

Healy, M. J. R. (1974) Animal litters as experimental units. *Applied Statistics* **21**, 155–159.

Hujoel, P. P. & Moulton, L. H. (1988) Evaluation of test statistics in split-mouth clinical trials. *Journal of Periodontal Research* **23**, 378–380.

Imrey, P. B. (1986) Considerations in the statistical analysis of clinical trials in periodontitis. *Journal of Clinical Periodontology* **13**, 517–528.

Loesche, W. J. (1979) Chemotherapy of dental plaque infections. *Oral Science Review* **9** (63), 1976.

Marshall-Day, C. D. & Shourie, K. L. (1949) A roentgenographic study of periodontal disease in India. *Journal of the American Dental Association* **39**, 572–588.

Milliken & Johnson. (1984) *Analysis of messy data*, 1st edition, pp. 407. New York: Van Nostrand Reinhold Company.

Neter, J., Wasserman, W. & Kutner, M. H. (1985) *Applied linear statistical models*, 2nd edition p. 663–666. Homewood: R. D. Irwin.

Osborn, J. (1987) The choice of computational unit in the statistical analysis of unbalanced trials. *Journal of Clincial Periodontology* **14**, 519–523.

Page, R. C. & Schroeder, H. E. (1982) *Periodontitis in man and other animals. A comparative review*, 1st edition p. 222. Basel: Karger.

SAS Institute Inc. (1985) *SAS users guide: Statistics*, 5th edition. Cary, NC: SAS Institute Inc., p 817.

Schluger, S., Youdelis, R. A. & Page, R. C. (1979) *Periodontal disease* 1st edition pp. 322, ff. Lea & Febiger: Philadelphia.

Smith, D. H., Ammons, W. F. & Van Belle, G. A. (1982) A longitudinal study of periodontal status comparing flap curettage and osseous recontouring (I). 6 months results *Journal of Periodontology* **51**, 382–375.

Suomi, J. D. & Barbano, J. P. (1968) Patterns of gingivitis. *Journal of Periodontology* **39**, 71–74.

Varma, A. O. & Chilton, N. W. (1974) Crossover designs involving two treatments. *Journal of Periodontal Research* **9**, Suppl 14, 160–170.

Address:

Philippe Hujoel
Department of Dental Public Health Sciences
University of Washington, Sm-35
Seattle, Wa 98195
USA