Clin Oral Invest (2007) 11:5–33
DOI 10.1007/s00784-006-0095-7

REVIEW

Recommendations for conducting controlled clinical studies of dental restorative materials

R. Hickel · J.-F. Roulet · S. Bayne · S. D. Heintze · I. A. Mjör · M. Peters · V. Rousson · R. Randall · G. Schmalz · M. Tyas · G. Vanherle

Received: 14 December 2006 / Accepted: 14 December 2006 / Published online: 30 January 2007

Abstract About 35 years ago, Ryge provided a practical approach to evaluation of clinical performance of restorative materials. This systematic approach was soon universally accepted. While that methodology has served us well, a large number of scientific methodologies and more detailed questions have arisen that require more rigor. Current restorative materials have vastly improved clinical performance and any changes over time are not easily detected by the limited sensitivity of the Ryge criteria in short term clinical investigations. However, the clinical evaluation of restorations not only involves the restorative material *per se* but also different operative techniques. For instance, a composite resin may show good longevity data when applied in conventional cavities but not in modified operative approaches. Insensitivity, combined with the continually evolving and non-standard investigator modifications of the categories, scales, and reporting methods, has created a body of literature that is extremely difficult to meaningfully interpret. In many cases, the insensitivity of the original Ryge methods is misinterpreted as good clinical performance. While there are many good features of the original system, it is now time to move to a more contemporary one. The current review approaches this challenge in two ways: (1) a proposal for a modern clinical testing protocol for controlled clinical trials, and (2) an in-depth discussion of relevant clinical evaluation parameters, providing 84 references that are primarily related to issues or problems for clinical research trials. Together, these two parts offer a standard for the clinical testing of restorative materials/procedures and provide significant guidance for research teams in the design and conduct of contemporary clinical trials. Part 1 of the review considers the recruitment of subjects, restorations per subject, clinical events, validity versus bias, legal and regulatory aspects, rationales for clinical trial designs, guidelines for design, randomization, number of subjects, characteristics of participants, clinical assessment, standards and calibration, categories for assessment, criteria for evaluation, and supplemental documentation. Part 2 of the review considers categories of assessment for esthetic evaluation, functional assessment, biological

R. Hickel (✉)
Department of Operative Dentistry and Periodontology,
University of Munich,
Goethestraβe 70,
80336 Munich, Germany
e-mail: hickel@dent.med.uni-muenchen.de

J.-F. Roulet · S. D. Heintze
Ivoclar-Vivadent AG,
Schaan, Liechtenstein

S. Bayne · M. Peters
Department of Cariology, Restorative Sciences, and Endodontics,
School of Dentistry,
Ann Arbor, MI, USA

I. A. Mjör
Department of Operative Dentistry, University of Florida,
Gainesville, FL, USA

V. Rousson
Department of Biostatistics, ISPM, University of Zurich,
Zurich, Switzerland

R. Randall
3M ESPE Dental,
St. Paul, MN, USA

G. Schmalz
Department of Operative Dentistry and Periodontology,
University of Regensburg,
Regensburg, Germany

M. Tyas
School of Dental Science, University of Melbourne,
Melbourne, Australia

G. Vanherle
Department of Operative Dentistry and Dental Materials,
Catholic University Leuven,
Leuven, Belgium

responses to restorative materials, and statistical analysis of results. The overall review represents a considerable effort to include a range of clinical research interests over the past years. As part of the recognition of the importance of these suggestions, the review is being published simultaneously in identical form in both the "Journal of Adhesive Dentistry" and the "Clinical Oral Investigations." Additionally an extended abstract will be published in the "International Dental Journal" giving a link to the web full version. This should help to introduce these considerations more quickly to the scientific community.

Keywords Filling · Technique · Trial · Adhesive · Composite · Ceramic

Introduction

In order to carry out a prospective clinical investigation of dental materials and/or techniques, most researchers use the Ryge Criteria for assessment of study restorations. Ryge developed this measurement scale more than 35 years ago ([10]; reprinted 2005) as a standardized method to clinically evaluate restorations [3]. The work was done during his tenure at the United States Public Health Service and his measurement standard is also known as the USPHS criteria. The criteria were drawn up at a time when the longevity of direct restorative materials, other than amalgam, was limited; deterioration and inadequacies of these materials were more pronounced, and defects appeared earlier than with present-day materials. Researchers often adapt the criteria in an effort to make them more discriminating for modern restorative materials, with the consequence that there are many so called modified Ryge criteria in use. Virtually every modification is different with the result that comparison between studies has been eroded. Despite the criteria changes, the majority of restorations in many studies continue to receive an Alpha score at the six, 12 and 18 month evaluations. In order to detect early deterioration and differences between restorations it would be helpful to have a more discriminative scale. There is therefore a definite need for new or improved criteria for the clinical analysis of direct and indirect restorations.

This paper will primarily focus on evaluation criteria for prospective clinical studies on all types of restoration, and in addition some general ideas on research methodology for conducting studies will be considered. The following main topics will be addressed: Study design and criteria for publication, standards and calibration, and criteria for evaluation.

The majority of the elements that make up the study design and clinical evaluation criteria are also applicable to fixed partial dentures such as crowns and bridges. This paper may therefore also serve as a basis for the development of a standardized protocol for clinical trials in fixed prosthodontics.

Part I. Study design for clinical trials of direct and indirect restorations including onlays and partial crowns

Shortcomings in clinical study design

In the last three decades numerous reports on clinical studies in restorative dentistry have been published. Many studies have exhibited some deficits in design which have negatively impacted the outcome value or possibilities of interpretation for statements or guidelines. Also many publications have not reported in sufficient detail thus impairing later analyses of data and comparison with other studies. There are many publications on clinical studies available but in most cases it is virtually impossible to analyze and compare the results; meta-analyses for example are not feasible. This is largely due to inadequate design and description of the clinical study in publications, insufficient reporting on restoration placements or their evaluation, inadmissible pooling of different groups (e.g. posterior and anterior restorations) etc. Furthermore, incomplete reporting of results or inappropriate statistical analysis can impair the outcome and value of a study.

All clinical studies involve a great deal of work, high budgetary costs (grants) and occupy scientists' and patients' time. It is necessary to plan them carefully and to optimize the evaluations to obtain the most valid outcome; otherwise the studies will be a waste of time, energy, and money. Simple studies may cost less but they still require much work, including recruitment of patients and IRB or ethics committee approval. If the outcome is of no or very limited value because of insufficient design or inadequate tools for evaluation, the amount of work and money is eroded, and it could be argued that such a study is unethical in that the results are of such limited value. Therefore it is better to have fewer but excellent, and preferably multi centre studies with high reliability, validity and reproducibility.

Evidence-based medicine requires that the aim and structure of all future clinical studies be such that the results can be included in a meta-analysis [52]. A meta-analysis is a way of combining quantitative evidence from different studies, and by statistical analysis, extracting a more objective and quantitative summary of the evidence than could be obtained from a traditional review [2]. Published studies need to include information in sufficient detail to be included in a meta-analysis, for example, give an adequate description of how randomization was carried out, not just a statement that it was done. The reader needs

to be able to appraise the quality of the study design from the published report. The description of the restoration placements and documentation of their performance should be in enough detail to permit reproducibility [72]. This lack of detail can result from a journal editor's requirement to limit page numbers per manuscript as much as from an error of omission on the part of the author.

Ongoing studies should continue with their existing criteria but consideration should be given to ensuring that on publication optimal information is made available to the reader and is of maximal value for further use of the results e.g. for consideration in reviews, guidelines etc.

Shortcomings of clinical studies can be divided into two main areas:

1. Deficiencies in study design, insufficient reporting on operational procedures and/or inadequate statistical analysis and results presentation.
2. Use of insufficient, invalid and non evidence-based evaluation criteria.

Improvements concerning the first point can be facilitated by applying existing standards currently adopted in the medical field concerning the planning and conduct of clinical trials. The application of sound evaluation criteria may be difficult in some studies and not readily manageable; validation of some criteria still requires research. The interpretation of the outcomes of clinical studies that are carried out in the meantime should be limited until more evidence is available. It therefore seems reasonable to utilize only those criteria that have a sound scientific basis. As these evaluation criteria are directly related to the evaluator's decision as to whether the restoration is scored as a failure or not the criteria determine the survival rate and annual failure rate of the restorative materials and/or restorative technique.

The rationale to judge restorations as acceptable or as failures should be communicated to general practitioners as guidelines for improvement in dental health care by reducing erroneous treatment and over treatment.

Study design

Many publications omit to adequately describe the details of how a study was run (e.g. inclusion/exclusion criteria not listed, control group not defined, or without randomization etc.). In many papers there is not enough information about subject or case selection. Attitudes and habits of patients such as bruxism, as well as oral hygiene and caries risk, may greatly influence the results and in many clinical studies these factors are not, or not adequately, considered or described in sufficient detail. The same is true for pre-existing damage to the tooth or location and size of the cavity etc.

Recruitment of subjects

The recruitment of subjects is a decisive stage as it may influence the outcome of the study. Often the convenience of using dental students and dental school faculty and staff in order to reduce patient drop out during a longitudinal study is tempting. However, the performance of a restorative material in a dental student population may be different from that in a community of socially deprived subjects. The difference in dental awareness towards oral hygiene and caries preventive measures in these groups would account for this. Recruitment should be carried out in such a way that the study group consists of more or less a cross-section of the population where the study is carried out; otherwise the conclusion is inevitably limited to the selected study population.

Number of restorations per subject

Each subject that receives the interventional treatment should be regarded as a statistical unit. The performance of multiple restorations placed in one subject cannot be considered independent from one another as the influence of the individual subject may play a crucial role. The statistical power of a clinical study cannot be enhanced by simply putting more restorations in a limited number of patients and it is reasonable to limit the number of restorations to one per group/material per subject.

Including one test and one control group in a study of paired restorations, where possible, is considered to be the preferred option so that both the test and control materials are placed in the same type of tooth with comparable cavity size, preferably on contralateral sides of the jaw, and preferably at the same appointment. This procedure however can drastically limit the available patient pool. A different study design and larger sample size may offer an alternative in situations where the requirement of paired restorations will slow down patient accrual and undesirably prolong the duration of the recruitment period. Each part of the clinical evaluation should be explicitly described during the design phase of the study. This approach, combined with appropriate published data from clinical studies on restorative materials, provides the basis for the power calculations, and thus the sample size.

Operational procedures

The procedures required during cavity preparation and placement of restorations, such as bevelling, liner and adhesives, and incremental placement techniques have always been considered very important however in some studies this information has not been given in detail. For example the mode of curing may play an important role in

strength, wear, and/or gap formation of resin-based restorations. Therefore the abbreviated but frequently seen wording in Materials and Methods "that the material was light cured" is not sufficient. Details about the curing device, such as intensity, time and mode of light curing should be included to increase the potential for analysis of failures and for comparison of different studies. Marginal staining and gap formation may not be exclusively linked to the restorative material but may also be influenced by cavity preparation, the conditioning and bonding procedure and/or placement technique etc. It would be unsatisfactory to only link marginal staining to the specific restorative material. Self-etching adhesives for example can show more marginal staining of restorations compared to etch and rinse systems, and there may also be differences between products in the same class of adhesives. A detailed description of all materials and techniques facilitates later analyses and comparisons.

Clinical events

Fracture and loss of retention Fracture of restorative material or tooth tissue, and loss of retention are so-called "hard criteria" of evaluation as diagnosis is clear cut and often apparent to the patient, and usually results in re-restoration of the cavity. In some instances repair is also an option. As these clinical events are easily recognized they should not pose a problem during the evaluator calibration process.

Caries Associated with Restorations (CAR)-synonymous with recurrent caries or secondary caries The detection of caries associated with the restoration margin (CAR), being synonymous with recurrent caries or secondary caries, and the interpretation of clinical signs that may indicate CAR, is not straightforward [17]. Up to the present time diagnosis of recurrent or secondary caries has been more or less based on experience and 'preventive' treatment approaches; even 'intuition' has been suggested as a foundation for the diagnosis [7].

The aetiology of secondary caries is similar to that of primary caries. The same types of cariogenic bacteria are involved (e.g. S. mutans, lactobacilli), both at the beginning of caries development in enamel and in deeper dentinal regions [14, 30, 46]. Secondary caries is primary caries adjacent to a restoration that starts on the tooth surface closest to the restoration [30, 35], either in enamel or dentine. It therefore follows that the same criteria for diagnosing primary caries should be applied to diagnosis of secondary caries. Systemic fluoride sources, such as fluoridated drinking water, do not appear to affect the occurrence of secondary caries [25], the likely reason being that many clinically diagnosed secondary caries lesions are not caries *per se* but localized restoration defects [46]. Secondary caries develops mainly on the proximal gingival floor of Class II restorations and to a lesser extent at occlusal margins of Class II/I restorations [44, 45, 49], being independent of the restorative material.

Marginal discoloration has often been considered a sign of secondary caries. Since objective criteria or a suitable procedure for the diagnosis of secondary caries have been lacking it can be assumed that secondary caries is often diagnosed when no caries is present, and more rarely vice versa. A study of extracted restored teeth has shown that in 25% of cases a replacement restoration was recommended although no secondary caries was present, and in 9% of cases existing secondary caries had not been diagnosed [78]. These results are in agreement with another study showing that diagnosis of secondary caries corresponded with histological findings in only 37% of cases [43]. The supposition that an erroneous diagnosis followed by unnecessary replacement of the restoration (overtreatment) was justified is not acceptable [11]. Radiographs may help detect carious lesions in the cervico-gingival area [12, 27] however, these lesions can only be identified if the central x-ray beam is aimed directly at the defect, provided the lesion is not masked by other structures [83].

In many studies no differentiation has been made between stained restoration margins and secondary caries, both being diagnosed as secondary caries due to insufficient experience, improper criteria or a lack of calibration and recalibration. Secondary caries scores will often have been over-estimated. Mjör and Toffenetti [50] demonstrated notable differences in secondary caries between practice based cross-sectional and controlled longitudinal studies which cannot be explained otherwise. These problems will markedly influence the rate of scoring of acceptable and unacceptable restorations. Without sound examiner calibration based on current scientific information it is not possible to have a high reproducibility rate including high intra- and inter-examiner reliability. Notwithstanding that secondary caries was likely to have been misdiagnosed and therefore overestimated in a number of clinical trials, especially those with cross-sectional designs or those evaluating reasons for replacement of restorations [28, 29, 41]; the few longitudinal trials with an observation period in excess of ten years suggest that secondary caries is an infrequent occurrence with composite resin restorations if placed under ideal conditions. A number of studies on composite restorations have reported that between only 4% and 8% of restorations developed secondary caries over a ten year period [15, 28, 29, 40, 41, 62].

To improve the criteria for diagnosis of caries, including secondary caries, an international committee has defined a new international caries detection and assessment system [30]. It is recommended that these criteria be taken into

consideration when diagnosing and rating restoration margins for caries (CAR).

Marginal integrity Many clinical research workers focus on in vitro marginal integrity of direct and indirect restorations; for instance pre-clinical laboratory testing to evaluate dye penetration at the tooth restoration interface by sectioning the stained restored teeth [16], or by two-dimensional examination of the margin using scanning electron microscopy (SEM) [69]. Researchers claim that the greater the dye penetration or the more discontinuous the margin, the higher the risk of secondary caries. No evidence-based acceptance levels have been established to date. A dental institute has stated that if an evaluation of a restorative material in vitro in Class II cavities yields a percentage of continuous margins in enamel of at least 90%, and at least 80% in dentine, this can be considered a good performance [34]. These threshold values are arbitrary and not supported by clinical evidence. A systematic analysis that compared the in vitro results of Class V restorations considering the above-mentioned threshold values with those of clinical studies on Class V restorations involving the same adhesive system revealed that in nine out of the 11 studies selected according to specific criteria (prospective clinical study of at least two years involving at least two adhesive systems in a split-mouth design) the clinical outcome did not match the prognosis based on the laboratory tests [23]. No correlation was seen between the percentage of continuous margin of fillings placed in extracted premolars and the percentage of teeth that showed no retention loss in clinical studies, also between discoloured margins, acceptable margins or absence of secondary caries.

Further evidence that the microscopic view in the laboratory does not have clinical relevance comes from clinical studies where restorations were evaluated both clinically and by SEM of replicas. In a clinical study of indirect ceramic inlays using a replica technique with SEM, only about 70% of the enamel margins were rated as continuous after six months, irrespective of the luting resin used; furthermore overhangs were present in about 15-20% of the whole margin, which may have masked marginal openings; in the clinical evaluation, however, the marginal integrity of the majority of the restorations was rated as excellent [56]. A similar finding has been made for direct resin restorations based on annual SEM evaluation in conjunction with clinical examination over a period of ten years [15]. In a study on Class II composite restorations in premolars scheduled for orthodontic extraction a high percentage of overfilled (43%) and underfilled (25%) proximal margins were detected on replicas by means of SEM [58]; only 27% of the restorations were considered satisfactory. This finding was independent of the type of operative technique used (incremental technique with transparent matrices/horizontal layers with metal matrix, the type of adhesive system, and operator experience (student/dentist)). In a similar study marginal gaps were present in up to 20% of the internal interfaces of Class I restorations shortly after placement [57]. The presence of overfilled material impedes the accurate evaluation of the cervical margin of Class II restorations. In spite of a lack of scientific evidence for clinical relevance however numerous laboratory tests of this kind are still carried out.

Other factors to consider are gap width and continuity of the margin as well as gap depth. Most studies that use quantitative marginal analysis differentiate between the presence or absence of a continuous margin rather than between differences in gap width. It has been reported that neither the amount of marginal opening nor microleakage is related to post-operative sensitivity [57, 59]. Restorations with both marginal deterioration and cavomarginal discoloration at three years however failed 8.7 times more frequently at five years than restorations with sound margins at three years [21]. These authors concluded that clinical investigation of present-day posterior composite materials should seek to determine if marginal deterioration and cavomarginal discoloration is an important predictor of the failure of posterior composites, especially when marginal deterioration and cavomarginal discoloration occur simultaneously.

As far as the clinical diagnosis of secondary caries is concerned, a clinical-histological study has shown that in areas that are easily accessible to oral fluids and oral hygiene, such as occlusal surfaces, gaps of 400 micrometers or greater are necessary for secondary caries to develop [37]. For the critical, difficult-to-clean, proximal cervical areas of Class II restorations, the gap width related to the development of secondary caries is probably smaller. Independent of the size of the gap, this highlights the importance of the essential presence of cariogenic plaque for caries to develop in potential stagnation areas. Currently, there are no clinical data available with respect to caries development linking the dimensions of clinically relevant marginal discrepancies with the presence of cariogenic plaque. Özer [61] however demonstrated a relationship between even minimal marginal overhangs on restorations and the development of secondary caries. This finding corroborates the observation that secondary caries is most often diagnosed at the proximal cervical margin, which is a difficult area to evaluate clinically [44] and also a difficult area for the patient to keep plaque free. Secondary caries seldom occurs at the occlusal surface which tends to be self-cleansing [46]. Patient-related factors such as caries activity and oral hygiene level are confounders which are not often assessed in clinical evaluations of restorative procedures [38, 39]. Composite resin materials have been considered to promote an increase in caries-related mutans

streptococci in the interdental plaque [24, 81]. This finding may partly explain why adhesively luted ceramic inlays do not necessarily harbour less mutans streptococci in this area than conventional luting cements. The cervical margin of Class II ceramic inlays may be just as prone to secondary caries as the cervical margins of Class II direct composite restorations.

There is evidence suggesting that the integrity of the marginal seal of restorations correlates with the occurrence of marginal discoloration. This has been proven for ceramic inlays. A longitudinal clinical trial investigating ceramic inlays has shown that the frequency of marginal discrepancies on replicas analysed by SEM coincided with the frequency of clinical signs of discoloration [22]. A discoloured margin is not indicative of secondary caries [46, 50]; however marginal deterioration and cavosurface discoloration may be predictors of future failure [21].

As the evidence for the influence of marginal integrity and gap width on secondary caries is inconclusive and lacking in correlation to future failure rate, recording and grading gap width is important to include in future clinical studies to develop a scientific basis for such a potential relationship.

Wear and surface roughness A primary aim of many highly standardized longitudinal clinical studies is to demonstrate changes or deficiencies in new materials or differences in materials or techniques compared to conventional procedures, at an early stage. For several criteria such as wear, roughness and gap formation, this evidence can only be obtained using sophisticated methods. Wear or surface roughness should not be measured by outdated means. Sometimes a precise value is reported which cannot be achieved or reproduced by manual clinical evaluation. For example, in the 1970's, wear of composite restorations was excessive and detected by probing the step at the restoration margin either directly with explorers in the mouth or by comparing casts with models that showed defined steps at the margin (e.g. Vivadent scale, M-L scale). As long as the materials showed distinct wear in a short period of time, and no other tools for wear measurement were available, the extent of material loss at the restoration margin was considered indicative of wear. This clinical observation was the accepted standard. Contemporary restorative materials have been improved to the extent that wear cannot be scientifically quantified using a subjective procedure [82] as this would systematically underestimate wear rates even when comparing replicas with standardized models [63, 79].

Aesthetics The aesthetic acceptability of a restoration may be compromised in many different ways including loss of anatomic form, roughness and surface staining, staining of the material or the margins, and/or colour shift. The evaluation is somewhat subjective and therefore prone to a great deal of variability, especially if the deterioration or alteration is of minor extent. The aesthetic appearance of a restoration is to a large degree dependent on how well it blends in with the surrounding tooth structure.

Failure rate

Some publications focus on results for specific individual criteria, e.g., anatomic form, fracture, gaps, and secondary caries, and it can be unclear how many restorations failed in total. Sometimes details of failures for two or three different criteria are given but it is not apparent whether or not these involve different restorations, resulting in uncertainty as to how many restorations failed in total. This is one reason why we not only need improved criteria for evaluation, standardization, and calibration but also for improved design and statistical analysis of studies.

Validity versus bias

Validity

A study is **internally valid** if the conclusions represent the truth for the individuals studied, in other words the results were unlikely to be due to the effects of chance, bias, or confounding factors because the study design, execution, and analysis followed accepted good practices [42]. These features are characteristic of the randomized controlled clinical trial. If the study conclusions represent the truth for all populations to whom the results can be applied, the study is also **externally valid,** as in many practice-based cross-sectional studies; these however are not as scientifically rigorous as the randomized controlled clinical trial. It would therefore be desirable for practice-based research to also follow a randomized controlled clinical trial design.

Bias and confounding factors

A systematic error can occur during the design and execution of a study due to errors in sampling, selection, or allocation methods. The study comparisons may be between groups that differ with respect to the outcome of interest for reasons other than those under evaluation. Almost all studies do have some inherent bias, but the degree varies from study to study. The more information available on study design and execution the better one may evaluate the possibility of bias and therefore correct it. Systematic errors can be reduced to a large extent by standardization of study design, randomization of subjects, calibration of measurement equipment and evaluation criteria.

Systematic errors can also occur during interpretation of the results. Such bias can be caused by lack of clinical experience, tradition, credentials, prejudice and human nature. The human tendency is to accept information that supports pre-conceived opinions and to reject or trivialize that which does not support these opinions or that which is poorly understood [60].

To avoid such errors, the evaluation should always be performed independently from personal or situation bias and be based on sound decision making. For example the operator and evaluator should be different individuals to avoid biased assessment of the restorations even after long periods between recalls; decisions should be taken according to a set of pre-determined criteria, the decision to replace a restoration should be made by an independent evaluator and not the operator or private practitioner whose practice management situation may influence the outcome (open schedule for appointments or full schedule for months ahead). If the design of the study does not address possible bias, the resulting data may be skewed.

Confounding variables and effect of modifying factors may cause distortion of the effect of one influencing factor by the presence of another. Factors that are not primarily related to the study itself but influence the outcomes are called confounding factors. For instance, in the evaluation of a medicament to reduce the incidence of periodontitis an important confounder would be "smoking" as smoking is a risk factor for periodontal disease and therefore the response of smokers to pharmaceutical interventions may be different compared to non-smokers. Likewise the occurrence of restoration failure in a group of bruxing patients may differ from a group of non-bruxers. Factors such as upper or lower jaw may not be regarded as important confounders for the success of dental restorations; however the operator may be a powerful confounding factor. Restorations placed by trained and skilled personnel may have a very different outcome to those placed by untrained and unskilled operators. The manual dexterity of clinicians can vary. Invariably a confounding risk factor is not distributed randomly within the study or between the test and control groups. Confounding can be controlled by defined inclusion and exclusion criteria, and randomization of both treatments of teeth/subjects, and operators, by matching the confounding variable and/or by including it in the statistical analysis.

Legal and regulatory aspects

All restorative and prosthodontic materials which do not have a biological effect are defined as medical devices both in the USA and the European Community and European Economic Council. Many other countries have also adopted this definition [13] however research workers in countries other than the EU or North America should verify whether their specific regulatory bodies apply the same directives to medical devices or have implemented other national regulations.

Clinical studies must be approved by an ethics committee (Institutional Review Board or IRB in the USA) that evaluates whether the proposed study design complies with international standards with regard to product safety, necessity of the study, and the protection and rights of the subjects and investigators. In addition an efficient and comprehensive adverse event and adverse device effect, as well as a complaints management system should be implemented. A recent requirement in the EU is to register clinical studies with the regulatory body, also the results, whether positive or negative, must be made available to the public. Regulatory and funding bodies in other countries may follow similar guidelines. This requirement will ensure that unfavourable clinical results from the performance of medical devices and pharmaceutical products are not put aside but are published within a certain time period after the study has been completed.

Rationale for conducting clinical studies on dental materials and/or operational procedures

The objectives of most dental restorative clinical studies are two-fold:

1. To evaluate whether a new or modified dental material and/or procedural technique is suitable for its defined indication for use intra orally according to accepted requirements such as
   - restoration/maintenance of function
   - improvement/restoration of aesthetics
   - protection of biological structures such as the pulp and periodontal tissues
   - that the dental material does not cause unacceptable side-effects.
2. To further evaluate whether a new or modified dental material or clinical procedure is suitable for use by the majority of professionals who will carry out the service, mainly dentists or specifically trained health workers (e.g. for materials such as fissure sealant and Atraumatic Restorative Treatment (ART)).

In consideration of the first point it is recommended that longitudinal studies, designed as randomized controlled clinical trials, be carried out at universities or specialized clinical institutions. To estimate an operator effect, a control material and/or control technique should be included. Further, at least two independent test centres that follow

the same protocol should be involved to determine whether the outcome is reproducible. The purpose of such a study is also referred to as an evaluation of the "efficacy" of a treatment under 'ideal' conditions, or 'feasibility' testing.

To address the second point, prospective clinical trials involving a representative cross-section of general practitioners, preferably in different countries should be conducted to test the technique-sensitivity of a material and/or operational technique. This kind of evaluation has been referred to as practice-based research [53]. The procedure can be thought of as the evaluation of the 'effectiveness' of a treatment under normal realistic conditions. Further objectives could be to test the efficacy or effectiveness of a certain material when applied to specific populations such as subjects with high caries activity.

The need for a specific clinical study on a dental material and/or clinical technique has to be determined prior to the development of the study design. Written justification of the necessity for the study is the responsibility of the investigator or trial sponsor. This involves establishing, from published literature and any internal data available, whether the material or technique warrants clinical testing, or whether there is sufficient similarity to existing products or techniques for which adequate and valid data already exist.

Before starting a clinical study the physico-chemical properties and biocompatibility of the test material must be analysed and be acceptable. For example for composite materials certain physical properties such as flexural strength, hydrolytic expansion over time are likely to have a clinical correlate, and sufficient laboratory test data should be available prior to the clinical trial. These data should be carefully examined and assessed to help circumvent a high clinical failure rate due to inadequate material properties. Standards published by the International Organization for Standardization (ISO), in particular ISO 4049 on polymer-based restorative materials [32] and the technical specification on testing of adhesion to tooth structure [33] offer useful guidelines. The standards are up-dated on a regular basis and describe test methodologies as well as threshold values, e.g. for flexural strength, water solubility, water absorption, curing depth, sensitivity to ambient light, colour stability, etc. Other parameters such as shrinkage, shrinkage stress, Vickers hardness, rate of conversion, and modulus of elasticity, may also be of interest, however, there is less evidence in the literature that these parameters have a well-defined clinical correlate.

An important prerequisite for clinical testing is the availability of adequate data to show acceptable biocompatibility of the product. As with clinical trials the type and scope of biocompatibility testing depends on how closely related the new material is to an existing commercially available product with regard to new molecules etc. Further information on test methods for biocompatibility may be found in ISO 7405 [31].

Guidelines for study design and publication (for authors, reviewers and editors)

The first step in planning a clinical study is to determine the aim(s) of the study. The test parameters must be clearly defined, often as a primary end point or hypothesis, and some secondary end points or more minor parameters. A decision has to be made on what is to be analysed, e.g. the potential of a new restoration material in the clinical context, or the possibilities of a new treatment procedure, or both. It is important to determine the test variable at the beginning of the study. Once the variables are identified the control group can be defined; the control group will allow a comparison of the results. It is possible to compare several arms of a study but if too many factors are compared between test and control groups, or different treatment arms, the interpretation of the results becomes difficult and limited. It is usually not recommended to evaluate many variables within a single study.

Randomization can be done at several levels and the advice of a statistics expert is highly advisable. If possible each subject should receive similar treatment from a limited number of experienced, calibrated operators. If feasible a split mouth design or paired tooth design should be selected as the preferred method [1]. There should not be more than one restoration per group per patient. A sufficient number of restorations should be placed taking into account potential subject dropout during the trial period. For this reason extra patients should be recruited into the study to ensure an adequate number of restorations. For each patient there must be a relevant treatment need, and all treatment offered must be ethically justifiable.

The clinical procedure has to be fully standardized before the start of the investigation and should have the approval of an ethics committee or IRB. The clinical protocol should be written up in detail and include all important elements such as cavity preparation (e.g. bevelling of margins, location or nature of margins, (enamel or dentine), cavity size, lining procedures, bonding procedure, the insertion and finishing technique, and complete information regarding the variables of resin polymerization). If the protocol requires it, post-finishing steps such as an additional layer of bonding or topical fluoride application must also be described. Too strict a regime however should be avoided as this is an unrealistic expectation for clinical studies in restorative dentistry.

For *indirect restorations,* in addition to the requirements outlined above, the type of the material and fabrication method, as well as the luting material and technique

employed, including isolation and conditioning should be specified. In addition, prior to luting the restoration, the quality of the various components (made by dentists, dental technicians, copy milling or CAD/CAM devices) should be assessed/measured and documented.

If impressions are taken for the construction of replicas to allow indirect observations and/or measurements, the procedure and materials used should be described. Once the clinical procedure has been established it should not be changed during the study. If an alteration to the protocol is required then this may necessitate further review by the ethics committee or IRB.

For all clinical studies a patient's information sheet and informed consent form must be prepared in the local language. A comprehension level equivalent to 6th grade/ primary school is recommended.

The following points should be considered during the design of a clinical study:

– Control group

A study without a control group is of limited value. By definition a control group is required in all studies that adopt a randomized control clinical (RCT) trial approach. If possible, and reasonable, a split mouth design is preferable. A multi-centre study is desirable to evaluate the influence of different test centres and/or operators/ examiners. Note that the following text emphasizes the RCT and that the results from these trials may have limited application in general dental practice because they are mostly conducted at universities with selected subjects and under ideal conditions.

The guidelines and recommendations of CONSORT (Consolidated Standards for Reporting of Trials; www.consort-statement.org) with regard to the conduct of RCT should be adopted as far as possible [51, 52]. The randomization procedure should be fully described in the protocol along with the inclusion and exclusion criteria for the recruitment of subjects. Confounding variables that may influence the clinical outcome have to be defined, evaluated and taken into account during the recruitment and randomization of the subjects.

– Randomization

Randomization is needed to ensure that the treatment and control groups are comparable according to all known and unknown confounding variables; it then follows that a difference between the groups with respect to the outcome of the study may be attributed to the treatment effect. On the other hand, if the groups differ according to some important variable, for example if one group comprises patients who are much younger than the other group, any difference with respect to the outcome could be due to age differences or to a treatment effect. The random assignment of a subject, or tooth to be restored, to the treatment or control group should not, in principle, present any difficulty. The assignment can be done using sealed envelopes which indicate either "control" or "treatment". In practice, however, the procedure is not always correctly applied because more than one envelope may be opened. Thus, it may be safer to use a computer to produce the random assignment. Since the computer algorithm is reproducible, it is then easier to check (or even to prove) that the study was truly randomized. The algorithm may also ensure that the two groups contain the same number of subjects at the end as well as at some earlier stages of the trial (a so-called "block-randomization"). It should be recognized however that the two groups will not be identical.

By definition, if the randomization procedure was perfect, the groups would differ from one another by only five percent of the confounding variables, and this would be the case whatever the sample size. With a large sample size, statistically significant differences between groups will often be small so the groups tend to be comparable. For small samples, it is sometimes recommended to define a few "strata" i.e. subpopulations, based on important confounders (for example a subgroup of caries-active and caries-inactive subjects) and to perform a block-randomization within each subgroup. For multi-centre studies, each centre may typically be a subgroup. If some confounding variables are known to have a major influence on the outcome of a study, it is better to include them as covariates in the statistical analysis. In addition to eliminating bias due to a possible group difference based on these variables, this has the further advantage of increasing the power of the study.

Number of subjects

Prior to starting the study a sample size calculation and statistical power analysis should be carried out. The specific sample size needed to show a treatment effect should be calculated statistically. This is based on the effect size expected to be observed between the test groups which can often be estimated from published data as well as the researcher's own experience. Differentiation should be made between anterior and posterior restorations, and cavity class or size according to the indication and/or type of material. A minimal sample size for planned cells or clusters that meet the same criteria (e.g. two-surface restorations on upper molars with same material combination and same operational procedure) should be calculated. Pooling of restorations of different indications e.g. posterior with anterior or cervical restorations is not recommended.

The number of subjects enrolled in the study is usually determined such that there is high probability of obtaining a statistically significant difference between the control and

treatment groups with respect to the main outcome of the study (typically 0.8 or 0.9). This probability is called "power" and is calculated under the hypotheses of the study (based on the expected difference between the groups using e.g. information from previous studies). Thus, given some hypotheses, one can calculate the sample size which is required to get a power of at least 0.8 or 0.9. If the calculated sample size is too high (i.e. is more than the number of subjects which is possible to enrol), one may conversely calculate for a given (available) sample size how large the difference between the groups should be in order to get 80 or 90 percent chance to detect it as statistically significant. If this difference is in turn too high (i.e. unrealistic), then it is probably not worth performing the study.

Calculation of sample size depends on the statistical test which is used. Explicit (although often approximate) formulae are available for some classical tests [8]. Useful software for calculating power and sample size incorporating various problems and situations is nQuery Advisor (current version 6.0) by Statistical Solutions, Crosse's Green, Ireland (http://www.statsol.ie).

Operator and procedure characteristics

- The number of operators must be specified including numbers of restorations placed per operator. Each operator should carry out sufficient restorations to statistically exclude or document an operator effect. The operators should be blinded to the test and control material where possible.
- The types of teeth involved (premolar, molar, anterior and primary or permanent teeth, upper and lower jaw) should be described.
- It is important to give a description of the size of the cavity, not simply Black's classification or "MOD"; also the proportions of the restoration margins located in enamel and dentine.
- For studies on direct and indirect restorative materials, cavity preparation features such as bevelling and lining, and the bonding protocol must be described in detail.
- Additionally for indirect restorations the process and quality of the fabricated restoration should be assessed and documented prior to luting, for example by photographs and notes on restoration fit. The luting protocol should also be described in detail.

Table 1 includes a checklist of the core requirements for the design of a study on restorative materials.

Clinical assessments

Clinical assessments should be carried out at pre-stated regular time intervals. A three year follow-up is advised for direct restorative materials while an observation period of five years is recommended for indirect restorations such as veneers or inlays. Longer observation periods are very useful especially when a new treatment method is being tested. For an observation period of three years, four recall sessions could be planned. The baseline evaluation should be carried out about one week (or at the latest one month) after the insertion of the restorations, and never immediately after placement. The remaining three recalls can either take place after 6, 18 and 36 months or preferably can be yearly recalls, namely after 12, 24 and 36 months. An annual survival rate is more easily calculated from the second option For a five year period six assessments are preferable (baseline and every year); alternatively five reviews, e.g. baseline, 1, 2 , 3 and 5 years can be carried out.

Before the start of an assessment consideration should be given to requesting that each patient brush his or her teeth for three minutes, with the materials to do so provided free of charge. The plaque index can then be recorded followed by restoration polishing.

The clinical assessments should be carried out by experienced, calibrated examiners who were not involved in the placement procedures. Ideally they should be blinded to the treatment conditions. At least two investigators should record the results consecutively at the same appointment. If disagreement occurs it should be resolved by discussion at once. Clinical assessments are optimized using loupes therefore their use is highly recommended.

Clinical evaluations should include pulp vitality testing, photographs of the restorations and in some cases evaluation of replicas for indirect observation and/or measurements. Evaluation of replica models should of course be done blind. For ethical reasons radiographs should be taken only if clinically necessary. It is expected that a number of countries will place increased restrictions on the use of x-rays in the near future, including use in clinical studies.

The number and reasons for patient drop out should be reported at each evaluation period. Any restorations excluded from the evaluation at a review appointment must also be documented. The reasons for patient drop out should be recorded, including tooth extraction due to reasons other than the study, changes in the opposing dentition in posterior studies, re-treatment by another dentist not involved in the study.

At each recall visit the subject should be re-evaluated with regards to the inclusion and exclusion criteria. In particular bruxism, oral hygiene and changes to the medical history should be noted. Archival storage of all the records for each study subject is mandatory and these data should remain accessible for further assessment. At each review session all records must be carefully checked for correctness including the photographs and impressions/replicas.

Table 1 Core requirements for conducting and reporting clinical studies on dental restorative materials

1. Design
   –Study design according to CONSORT recommendations, as far as applicable for the study
   –Define hypothesis and primary and secondary aims/endpoints
   –Use a randomized controlled clinical trials approach (RCT)
   → control group and randomization (as far as applicable for the study)
   –Whenever possible a multi-centre design is preferable.
   –Plan and check the statistical study design before starting the clinical part, including: calculation of sample size, and power analysis.
   → number of patients and restorations, preferably only one restoration per group per patient, estimate drop out rate of patients
   –Define baseline and recall intervals
2. Description of study population
   –Method of recruitment of patients
   –Inclusion and exclusion criteria
   –Patient characteristics including caries history and present activity, oral hygiene, parafunctional habits, and diet.
   –Type of teeth included, e.g. first and/or second premolars / molars, and numbers of each tooth type
   –Specify and quantify cavity size in some detail, not simply Class II or MOD
3. Description of interventions for experimental and control series
   –Describe operative procedures in detail: cavity preparation including bevelling, acid etching, conditioning, bonding, placement, curing, and polishing.
   –Provide information on all additional materials being used, including luting materials
4. Additional information
   –Acceptance/approval by an ethics committee
   –Results of laboratory testing sufficient and available including biocompatibility, physical parameters, radiopacity, material stability, and shelf life
   –Sources of financial support
   –Institution, where the tests were carried out
   –Operator qualifications e.g., student, dentist, specialty training or years of professional experience
5. Description of conduct of the study
   –Number and reasons for drop out given independently for each group
   –Adverse effects necessitating other interventions
6. Outcome evaluation
   –Specify periods for recall evaluations
   –Evaluator qualification. Evaluator(s) should not be the same individual(s) as the operator(s) and ideally should involve two independent calibrated individuals
   –Baseline assessment should be carried out at a separate appointment to the placement, and should include restoration polishing
   –Recommended that assessments be done after tooth cleaning.
   –Clinical evaluation should be done with magnification, e.g., loupes, and photo-documented
   –Evaluation criteria (see Table 2); overall failure rate, reasons for failure
   –Re-evaluation of inclusion/exclusion criteria
   –Photo documentation pre-operatively, at baseline and all recalls
   –State results in absolute numbers and percentage if feasible/justified

Standards, calibration and recalibration

General aspects

The following criteria will focus primarily on evaluation of the clinical performance of restorative materials, not on the dentist's ability to perform the restorative work. The aim is to distinguish clinically important differences between (new) materials and/or techniques and control groups with high reliability at an early stage of the study.

Many research workers seem to forget an important aspect of the original idea behind the Ryge criteria, namely calibration of the evaluators. This exercise should be revived. In order to obtain as much validity as possible only calibrated examiners should be used to assess study restorations. To assist with this a web based standard has been set up (http://www.dent.umich.edu/cer/) which will make it possible to train clinicians in clinical evaluation of restorations. It will include standard sets of photographs as reference instruments to illustrate the different criteria and groups of restorations and materials. Calibration courses during annual workshops will be encouraged at IADR, FDI and other research meetings. Participants will be able to be recalibrated at intervals, for example via the web, receiving instant feedback on accuracy and reproducibility.

Some criteria such as surface roughness and wear cannot be accurately measured without the use of sophisticated tools. If such advanced techniques/devices are not available at a study centre it is advisable to co-operate with another centre where such measuring apparatus is accessible rather

Table 2 Allocation of criteria to clinical observations

Clinical investigation ……..
ID patient / restoration ……..
Date (dd /mm/yy): Baseline…… 1. Recall …….. 2. Recall…… 3. Recall…… 4. Recall……
Photographs (n° and date): …… …
Replica (n and date): …… 5. Recall……...

a) Aesthetic properties	1. Surface luster	2. Surface staining	3. Colour stability and translucency	4. Anatomic form
1. Clinically excellent / very good	1.1 Luster comparable to enamel.	2.1 No surface staining.	3.1 Good colour match No difference in shade and translucency.	4.1 Form is ideal.
2. Clinically good (after polishing very good)	1.2 Slightly dull, not noticeable from speaking distance.	2.2 Minor staining, easily removable.	3.2 Minor deviations.	4.2 Form is only slightly affected.
3. Clinically sufficient / satisfactory (minor shortcomings, no unacceptable effects but not adjustable w/o damage to the tooth).	1.3 Dull surface but acceptable if covered with film of saliva.	2.3 Moderate surface staining, also present on other teeth, not aesthetically unacceptable.	3.3 Clear deviation but acceptable. Does not affect aesthetics: 3.3.1 more opaque 3.3.2 more translucent 3.3.3 darker 3.3.4 brighter	4.3 Form differs but is not aesthetically displeasing.
4. Clinically unsatisfactory (but repairable)	1.4 Rough surface, cannot be masked by saliva film, simple polishing is not sufficient. Further intervention necessary.	2.4 Surface staining present on the restoration and is unacceptable; major intervention necessary for improvement	3.4 (Localised) clinically unsatisfactory but can be corrected by repair 3.4.1 too opaque 3.4.2 too translucent 3.4.3 too dark 3.4.4 too bright	4.4. Form is affected and unacceptable aesthetically. Intervention (correction) necessary.
5. Clinically poor (replacement necessary)	1.5 Quite rough, unacceptable plaque retentive surface.	2.5 Severe staining and/or subsurface staining (generalized or localized); not accessible for intervention).	3.5 Unacceptable. Replacement necessary.	4.5 Form is completely unsatisfactory and/or lost. Repair not feasible / reasonable, replacement needed.
Overall aesthetic score	Acceptable aesthetically (n and %):		Not acceptable (n, % and reasons):	

b) Functional properties	5. Fractures and retention	6. Marginal adaptation	7. Wear	8. Contact point/food impact	9. Radiographic examination (when applicable)	10. Patient's view
1. Clinically excellent / very good	5.1 Restoration retained, no fractures / cracks	6.1 Harmonious outline, no gaps, no discoloration.	7.1 Physiological wear equivalent to enamel (80-120% of corresponding enamel).	8.1 Normal contact point (floss or 25 µm metal blade of can be inserted but not 50 µm blade).	9.1 No pathology, harmonious transition between restoration and tooth	10.1 Entirely satisfied
2. Clinically good (after polishing very good)	5.2 Small hairline crack.	6.2.1 Marginal gap (50 µm). 6.2.2 Small marginal fracture removable by polishing.	7.2 Normal wear with only slight difference to enamel (50-80% or 120-150 % of corresponding enamel).	8.2. Slightly too strong but no disadvantage	9.2.1 Acceptable cement excess present 9.2.2 Positive/negative step present at margin <150 µm	10.2 Satisfied
3. Clinically sufficient / satisfactory (minor shortcomings, no unacceptable effects but not adjustable w/o damage to the tooth).	5.3 Two or more or larger hairline cracks and/or chipping (not affecting the marginal integrity or proximal contact).	6.3.1 Gap < 150 µm not removable 6.3.2. Several small enamel or dentin fractures	7.3. Differing wear rate to enamel but within the biological variation (< 50 % or 150-300 % of corresponding enamel)	8.3. Slightly too weak, no indication of damage to tooth, gingivae or periodontal structures (50 µm metal blade can pass easily but not 100 µm)	9. 3. 1 Marginal gap < 200 µm. 9. 3. 2 Negative steps visible < 250 µm. no adverse effects noticed. 9.3.3 Poor radiopacity of filling material.	10.3 Minor criticism of aesthetics 10.3.1 Aesthetic shortcomings, 10.3.2 Some lack of chewing comfort, 10.3.3 Ttime consuming procedure and/or similar; No adverse clinical effects.
4. Clinically unsatisfactory (but repairable)	5.4 Chipping fractures which damage marginal quality or proximal contacts; bulk fractures with or without partial loss (less than half of the restoration).	6.4.1 Gap > 250 µm or dentine/base exposed. 6.4.2. chip fracture damaging margins 6.4.3 Notable enamel or dentine wall fracture	7.4 Wear considerably exceeds normal enamel wear; or occlusal contact points are lost (restoration > 300 % of enamel wear or antagonist > 300 %).	8.4 Too weak (100 µm metal blade can pass) and possible damage (food impaction). Repair possible.	9.4.1 Marginal gap >250 µm. 9.4.2 Cement excess accessible but not removable. 9.4.3 Negative steps >250µm and repairable	10.4 Desire for improvement (reshaping of anatomic form or refurbishing etc.)
5. Clinically poor (replacement necessary)	5.5 (Partial or complete) loss of restoration.	6.5 Filling is loose but in situ.	7.5 Wear is excessive (restoration or antagonist > 500 % of corresponding enamel).	8.5 Too weak and/or clear damage (food impaction) and/or pain/gingivitis. Requires replacement	9.5.1 Secondary caries, large gaps 9.5.2 Apical pathology 9.5.3 Fracture/loss of restoration or tooth	10.5 Completely dissatisfied and / or adverse effects incl. pain.
Overall aesthetic score	Acceptable function (n and %):			Not acceptable (n, % and reasons):		

Table 2 (continued)

c) Biological properties	11. Postoperative (hyper-)sensitivity and tooth vitality	12. Recurrence of caries, erosion, abfraction	13. Tooth integrity (enamel cracks)	14. Periodontal response (always compared to a reference tooth)	15. Adjacent mucosa	16 Oral and general health
1. Clinically very good	11.1 No hypersensitivity, normal vitality.	12.1 No secondary or primary caries	13.1 Complete integrity.	14.1. No plaque, no inflammation, no pockets	15.1 Healthy mucosa adjacent to restoration	16.1 No oral or general symptoms.
2. Clinically good (after correction very good)	11.2 Low hypersensitivity for a limited period of time, normal vitality.	12.2 Very small and localized 1. Demineralization 2. Erosion or 3. abfraction. No operative treatment required	13.2.1 Small marginal enamel split (<150 µm). 13.2.2 Hairline crack in enamel (<150 µm not probable).	14.2. Little plaque, no inflammation (gingivitis), no pocket development	15.2 Healthy after minor removal of mechanical irritations (sharp edges etc.)	16.2 Minor transient symptoms of short duration (of known or unknown origin) local or generalized.
3.Clinically sufficient / satisfactory (minor shortcomings with no adverse effects but not adjustable without damage to the tooth)	11.3.1 Premature / slightly more intense 11.3.2 Delayed/weak sensitivity; no subjective complaints, no treatment needed.	12.3 Larger areas of 1. Demineralisation, 2. Erosion or 3. Abrasion/abfraction but only preventive measures necessary (dentine not exposed)	13.3.1 Enamel split < 250 µm 13.3.2 Crack <250 µm; no adverse effects.	14.3.1 Plaque accumulation at acceptable level 14.3.2 Gingival bleeding acceptable. 14.3.3 Pocket formation acceptable.	15.3 Alteration of mucosa but no suspicion of causal relationship with filling material.	16.3. Transient symptoms, local and/or general.
4. Clinically unsatisfactory (repair for prophylactic reasons)	11.4.1 Premature/ very intense 11.4.2 Extremely delayed/weak with subjective complaints 11.4.3 Negative sensitivity Intervention necessary but not replacement.	12. 4.1 Caries with cavitation 12.4.2 Erosion in dentine 12.4.3 Abrasion/ abfraction in dentine Localized and accessible and can be repaired	13.4.1 Major enamel split (gap > 250 µm or dentine or base exposed). 13.4.2 Crack >250 µm (probe penetrates).	14.4.1 Plaque accumulation not acceptable. 14.4.2 Gingival bleeding not acceptable. 14.4.3 Pocket depth increase > 1 mm	15.4 Suspected mild allergic, lichenoid or toxicological reaction.	16.4 Persisting local or general symptoms of oral contact stomatitis or lichen planus or allergic reactions (or remitting). Intervention necessary but no replacement.
5. Clinically poor (replacement necessary)	11.5 Very intense, acute pulpitis or non vital. Endodontic treatment is necessary and restoration has to be replaced.	12.5 Deep secondary caries or exposed dentine that is not accessible for repair of restoration.	13.5. Cusp or tooth fracture.	14.5 Severe / acute gingivitis or periodontitis	15.5 Suspected severe allergic, lichenoid or toxicological reaction.	16.5. Acute / severe local and/or general symptoms.
Overall biological score	Acceptable biologically (n and %):			Not acceptable (n, % and reasons):		
TOTAL SCORE	Acceptable biologically (n and %):			Not acceptable, reasons:		

than adopt inappropriate criteria with low discrimination or reproducibility, resulting in data which profess an accuracy which is not achieved.

It should be however be kept in mind that the more detailed a clinical assessment becomes, the less reproducible are the observations recorded [48]. It seems that once the amount of detail sought reaches a certain limit it cannot then be universally applied. If a study requires recording of minute detail, calibration becomes difficult with the concomitant risk of recording differences in clinical judgement between evaluators rather than between experimental and control groups.

Criteria for evaluation of all restoration types

The criteria that constitute restoration failure have to be clearly and independently defined for each study, and should be formulated in terms of outcomes related to the aim of that study. The description of these criteria should primarily be based on established clinical measures reporting on failures of restorations. This approach is independent of whether the failure was caused by the material, the operator or the patient.

Early failures after 0-6 months:

- severe post-operative hypersensitivity
- loss of restoration (e.g. Class V)
- allergic side effects and or possible toxic reactions

Failures in a medium time frame (6–18/24 months):

- cracked tooth syndrome or tooth fracture
- marginal discoloration
- discoloration/staining of material
- chipping of material and or bulk fractures
- loss of tooth vitality

Long-term failures after >18/24 months:

- bulk fractures
- tooth fractures
- CAR (secondary caries)
- excessive wear of the material or opposing tooth
- periodontal side effects

The definition of clinical criteria should be based on these clinical events. Before the clinical criteria are presented and discussed in detail, some general and specific aspects of evaluation (grading) will be discussed.

Aspects of grading assessment

With the improvement in quality of present-day restorative materials there is a need for a more sensitive assessment (scoring) method with enhanced discriminative power compared with the original Ryge criteria. There is also a need for inclusion of a separate subgroup of repaired restorations. Some criteria such as wear, roughness and colour have to be measured by indirect methods using apparatus which allows precision and reproducibility; other criteria are evaluated with clinical aids such as probes, blades and articulating paper. For the allocation of criteria to clinical observations and findings see Table 2.

It is recommended that the examiner(s) use a two step approach for assigning scores for each parameter [70, 84]. The first step is to assess the restoration and to determine the level of clinical acceptability for each parameter in each of the categories.

It must be recognized that there is not always a clear definition characterizing the transition zone between an acceptable and an unacceptable result. Clinically a rule of thumb is that the result becomes unacceptable whenever re-treatment is necessary or highly advisable, in spite of the negative association with a re-intervention (see below). Even with this rule in mind some decisions or judgements remain difficult. Only experienced examiners with sufficient training and after proper calibration at ≥85% level will guarantee a reproducible result.

If a parameter is judged to be acceptable, as a second step a further distinction can be made between an excellent, good and clinically satisfactory result.

Score 1 means that the quality of the restoration is excellent/fulfils all quality criteria, and the tooth and/or surrounding tissues are adequately protected. Score 2 should be selected when the quality of the restoration is still highly acceptable, though one or more criteria deviate from the ideal. The restoration could be modified by polishing and upgraded to an 'excellent' rating but this is not normally necessary. There is no risk of damage to the tooth and/or the surrounding tissue; scores 1 and 2 would correspond to Ryge's Alpha rating; score 3 is equivalent to Bravo. Score 3 means that the quality of the restoration is sufficiently acceptable but with minor shortcomings. Because of their location/extent, however these cannot be eliminated without damage to the tooth, though no adverse effects are anticipated. Scores 4 and 5 correspond to Ryge's Charlie and Delta scoring which means that a restoration scored 4 is unacceptable but repairable whereas a restoration scored 5 has to be replaced. A five-step grading modification to the USPHS/Ryge criteria has already been described in a textbook by Charbeneau (in the section dealing with training of new faculty) [6].

When reporting over short observation periods, e.g. 6 or 12 month results, there are usually no or only a few unacceptable restorations, and the differentiation of excellent and good results becomes more important. Changes in the percentages of the first three scores can disclose a trend in the behaviour of these restorations and indicate the weak points in a particular material.

If a parameter is judged to be clinically unacceptable then the exact cause of failure has to be recorded and it has to be decided whether the restoration can be repaired or requires replacement. Not all problems lead to replacement of a restoration. Defects with easy clinical access can be repaired, e.g. sealing of gaps, adding new material to fractured restorations, re-contouring, part removal and veneering of stained restorations etc.

The decisive difference between score 4 and 5 is not the need for an immediate or a later replacement of the restoration, it is whether the restoration can be corrected/repaired or whether it has to be replaced immediately or later. Frequently score 5 will show worse clinical results than score 4 but that is not a must. Score 4 and consequently the possibility for repair depends more on the location of the problem and whether it is accessible for repair or not.

Although repair of restorations is nowadays especially recommended as a tool in minimally-invasive dentistry [46, 47] it is not often carried out by general practitioners or consistently taught at universities [5]. Published data on longitudinal studies to date are only available for amalgam restorations. The studies show equivalent survival of repaired, compared to replaced, restorations after five years, with however lower survival rates for repaired restorations after ten years of clinical service [77]. Data on repaired composite restorations are rather limited. Results after one and two years of service showed a similar performance of repaired versus replaced restorations [19, 20, 54].

Some examples of conditions suitable for repair:

- Marginal opening, or staining, or secondary caries without deep undermining caries, if accessible
- Exploratory preparation in the case of suspected secondary caries, e.g. replacement of only an approximal box of an MOD restoration if cervical caries is present.
- Moderate staining of the material, loss of surface gloss/increased surface roughness
- Chipping/partial fracture or marginal fracture of restorative material (incremental addition of material)
- Marginal breakdown of enamel or non-functional cusp fracture (incremental addition of material)
- Access preparation for endodontic treatment
- Amalgam restorations with accessible defects can be repaired using adhesive techniques such as bonded amalgam or composite.

– Ceramic inlays or partial crowns with fractures and/or chipping may be repaired by intraoral sandblasting/ silication, silanization and composite bonding.

A repair is a minimally-invasive approach whereas non-invasive interventions such as polishing or the application of sealants or adhesives with no restorative material are defined as refurbishments. Based on this definition a restoration that requires a repair should be considered a failure. Repaired restorations can however be further monitored and evaluated as a subgroup. Research is currently ongoing to obtain more detailed data on survival of repaired restorations [18, 19, 73–75].

It may sometimes be difficult for a patient to make a timely appointment for repair or re-treatment In some of these cases an unacceptable result may survive a number of years due to a favourable oral environment precluding further breakdown or dental discomfort; differentiation between active and inactive caries lesions is essential in these cases. The survival time however should not affect a decision to record an unacceptable result. Every unacceptable result will impede proper function of that particular tooth and most of the time will result in further dysfunction, breakdown, dental discomfort or pain. Such clinical conditions move an experienced clinician to restore normal function and/or prevent further damage therefore re-intervention is preferred, keeping in mind the principles of minimal intervention whenever possible. The need for major re-intervention or re-treatment is invariably interpreted as an unacceptable result.

To allow further analysis of the main study findings, we propose a classification of restoration assessment into three groups: Aesthetic, functional, and biological categories. Each group has sub-categories and the final result is determined by the sub-category scores; the final score in each group being dictated by the highest/most severe score among all the sub-scores. If one property/category is totally unacceptable, the whole aesthetic or functional or biological result becomes unacceptable, and the final, overall score is also unacceptable. Therefore, when summarizing the three scores (aesthetic, functional and biological) in one overall rating, the highest score prevails and gives the final score.

A restoration with a recurrent caries lesion at the margin would be unacceptable in the biological category and also in the final overall score, even though there may be acceptable ratings for the aesthetic and functional categories. The same holds true for a total colour shift visible from a distance of 60-100 cm etc. Grouping the restoration features in this way enables the annual failure rate to indicate the performance of the material or treatment mode, and the failure rates within each category/group will help to indicate where the main problems lie. Although the calculation of an annual failure rate may suggest a linear failure rate this will not always be the case.

The following criteria should be reported separately for the different classes of restorations. It is *not* necessary to adopt all the parameters in every study. Depending on the study objective some categories can be omitted or are not applicable, in particular aesthetic properties with amalgam or gold restorations, or for an experimental composite that is only available in two or three shades, or an ART material. For direct restorative materials occlusal anatomic form is not critical for the material's performance. This feature primarily indicates the operator's skill level and can be omitted as a main criterion in many studies. It is important however for evaluation of proximal contact areas, and where anatomical form is a primary goal of the investigation it should be included. Other factors such as tooth mobility, pocket depth, gingivitis etc. can be omitted in most cases as they do not have a high impact on the longevity of the restoration itself but rather reflect the functional oral environment. Periodontal assessments are of greater interest where the restoration is in direct contact with the periodontal tissues, as is the case for veneers in the anterior region, direct cervical restorations, and posterior proximal restorations with a margin located apical to the CEJ. If the plaque affinity of the material is of special interest, such as in bioactive or antibacterial materials, then these sub-categories should of course be investigated.

Specific aspects of restoration grading categories

Several features of a restoration determine aesthetic outcome such as colour stability and translucency, surface staining, surface gloss and anatomic form. These sub-categories are separately analysed and recorded, all scores then being taken into consideration to provide a final overall rating for the aesthetic properties, with the highest/ most severe score prevailing.

Restoration retention is an important functional property. For retained restorations a careful evaluation of the margins is carried out to detect defects such as gaps, under- or overfilling. Fractures at the margin and so-called bulk fractures within the body of the restoration are recorded. Chip fractures involving the restoration are identified, both small and severe chip fractures involve material loss and leave an irregular oblique fracture plane. Lastly marginal sealing capability and marginal discoloration are scored.

Functional form is analysed according to whether there is physiological wear comparable to enamel, or abnormal wear. If wear is to be evaluated quantitatively in a clinical study of posterior restorations, this is best carried out indirectly using replica models and a measurement tech-

nique with high reproducibility in the micrometer range. Occlusal contact area (OCA) and contact free area (CFA) wear are measured and compared with enamel facet wear. Differential wear is the term used to describe OCA composite resin wear minus enamel wear at the same facet. Ideally the wear rate should approximate enamel wear taking into account biological variation. If the total occlusal surface has been restored, comparison should be with corresponding enamel of a selected unrestored reference tooth. For clinical evaluation of wear of ceramic materials with cuspal coverage, consideration should be given to measurement of enamel wear of the opposing tooth.

The tightness of proximal contact areas of posterior restorations should be estimated in a reproducible manner. For a basic evaluation of the contact points use of waxed dental floss with a yes/no decision is sufficient, however, the brand of dental floss should remain the same for all re-evaluations. For a more precise differentiation, metal blades with increasing thicknesses (25, 50 and 100 μm) inserted into the interdental space have been recommended [67].

The final score for restoration functionality is determined from different scores with the highest or most severe score prevailing as the final rating.

In the *biological category* ratings include postoperative sensitivity or hypersensitivity, pulp vitality, and the recurrence of previous pathology such as caries, erosion, abrasion/attrition or abfraction at the margins. A careful analysis of the hard tissues is also required, including evaluation of tooth integrity and recording any enamel or dentine cracks at the restoration margins. Periodontal tissue health should be analysed if the restoration is adjacent to the gingivae. The Plaque Index (PI) [76], Sulcus Bleeding Index (SBI) [55] and probing depth are also recorded. As the restorative material is often located in the proximo-gingival area, where this is a region of primary interest, the Papilla Bleeding Index (PBI) [71] or proximal plaque index may be preferred. Plaque accumulation detected with a probe or after staining, or bleeding on probing, should always be compared with a matching tooth with no proximal restoration, on the contra-lateral side of the mouth or the opposing jaw. If there is no comparable tooth, no evaluation should be performed.

Radiographic examination should be restricted to patient treatment need, not for the purposes of a study. The results are recorded on the clinical report form as for the other categories. A radiographic evaluation should include the restoration and mineralized tissues, including the periapical area.

Finally a careful examination of the oral tissues and some questions concerning possible consequences of the dental treatment on the subject's general health will complete this part of the assessment. Again all these scores will contribute to the final biological score for each restoration.

Supplemental documentation

A *photograph* of the restoration, including the surrounding structures and part of the adjacent teeth, should be taken. In the anterior region it is sometimes helpful to record the dentition from canine to canine to show the anterior contact areas. In the posterior region a conventional buccal view of the restoration should be followed by a photograph of the occlusal surface showing the contact areas marked with articulating paper.

Replica models may be taken for each restoration in the posterior region to enable indirect observation and measurement. In the anterior region replicas may be useful to analyse certain changes on the surface and at the margins of the restoration. An impression should be taken for each restored tooth under investigation using a light-body polyvinylsiloxane material in a conventional partial tray. A recent comparative analysis of accuracy of clinical wear measurement using replica models revealed no difference between individually fitted and conventional trays [65]. If further replicas are needed a second impression should be taken. Immediately after removal of the set impression a thorough analysis of the impression should be done using a stereo microscope or magnifying loupes. Air bubbles or other imperfections in critical areas will make the impression useless; a new impression is the only solution. After measurements are complete, the calculated wear rates, including the OCA, CFA and differential wear, are added to the wear category for the restoration and graded as acceptable or unacceptable. The replica impressions may be stored for several years without detrimental effect if a polyvinylsiloxane material is used; the replica models should be stored for up to ten years after completion of the study for documentary purposes and later follow-up comparisons.

Part II. Clinical criteria for evaluation of direct and indirect restorations including onlays and partial crowns

The overall rating for a particular restoration is determined after completion of the assessments of the final scores for aesthetic, functional and biological properties. The most severe score will prevail. A description of the criteria and grading is presented in Table 2. Whenever a restoration receives a score of 4 or 5 independent of the specific criteria below, it must be recorded as a failure.

Aesthetic properties

It is recommended that surface staining and gloss or lustre are evaluated before judging colour stability and translucency. Changes in colour/staining should be documented by

clinical photographs and/or electronic shade measuring devices.

Surface gloss/lustre and roughness

Clinical assessment of surface roughness is difficult, especially with regards to reproducibility. Precise data on surface roughness should be evaluated on replica models using measuring devices such as profilometry and optical sensors. A qualitative possibility is to define roughness in relation to neighbouring enamel as similar to or rougher than enamel. Changes in roughness can also be seen as changes in lustre, being as shiny as or less shiny than enamel; ideally surface roughness should approximate that of enamel. When assessing these criteria it is recommended that the operator light be switched off and the evaluation carried out at a distance of 60–100 cm.

1.1 Surface gloss/lustre is comparable to that of the surrounding tooth tissues (mainly enamel).
1.2 Surface is slightly dull but not noticeable from a speaking distance of 60–100 cm.
1.3 Surface is dull but still acceptable if the surface of the restoration is covered with a film of saliva
1.4 Surface is rough and not masked by salivary film. Major re-finishing or veneering is necessary and possible.
1.5 Surface is unacceptably rough which makes it anaesthetic and/or it retains noticeable biofilm (plaque). Improvement by finishing or veneering is not feasible.

Surface and marginal staining

a) Surface staining

The restoration surface is evaluated with regard to staining by comparison with the surrounding hard tissues. When assessing discoloration a distinction should be made between staining and poor colour match. Accumulated stain on the surface of the restoration and analysis of this (including eating/drinking habits from the patient's history) should be differentiated from staining within the material itself. Slight staining is defined as being visible on clinical inspection using mirror and illumination, while severe staining is visible at a speaking distance of 60–100 cm.

Surface staining should have been removed during the tooth brushing and polishing session in the dental office. If not, one must examine whether the staining is present only on the surface of the restoration or also on other teeth. A heavy smoker will usually show surface staining of the teeth but if that staining is only seen on the restorations, these surfaces may be retaining staining pigments which is not acceptable. The same holds true for all extrinsic pigmentation such as coffee, tea, wine and the use of chlorhexidine mouth rinses. A distance of 60–100 cm is recommended for proper evaluation with the operator light switched off.

b) Marginal staining

The sealing capacity of restorations has often been assessed by colour change along part or the whole of the restoration margin. The colour change results from seepage or leakage of oral fluids between the restoration and tooth structure. Minor discoloration is only visible during dental inspection with a mirror and operating light, while severe discoloration is visible at a speaking distance of 60–100 cm [80, 84]. The same quantification as described for marginal deterioration (see point 6) should be used.

2.1 No marginal or surface staining.
2.2 Minor marginal staining (under dry conditions) and/or mild surface staining is present but is evenly spread over all the teeth. It does not affect the aesthetic properties because it is generalised and acceptable
2.3 Moderate marginal or surface staining not noticeable from a speaking distance
2.4 Surface staining is present on the restoration but not the tooth and is clearly recognisable from a speaking distance. Or severe localized marginal staining is present and not removable by polishing. The aesthetic properties of the dentition are affected. Restoration requires major correction and layering of new material.
2.5 Surface staining is totally unacceptable/unsightly and the restoration needs to be replaced. Or generalized and profound marginal discoloration is present.

There is usually no need to evaluate the intensity of marginal discoloration in more detail than outlined above. The same holds true for the extent of discoloured margins which can be expressed as a percentage of the discoloured margin in relation to the entire restoration margin. In some studies marginal staining may be of special interest and would then be classified or quantified in greater detail.

Colour match/stability and translucency
(not applicable for metallic inlays, or restorations not visible during normal function)

Restorations or parts of restorations that are easily visible at a speaking distance or during laughing, including Class IV, Class III restorations involving the labial surface, Class V in anterior teeth and premolars, and perhaps large MO or

MOD premolar restorations, should be distinguished from restorations that are not easily visible such as Class II and Class V in molars, and Class III if restricted to the lingual/palatal surface.

At baseline a restoration in a visible area of the mouth, mostly anterior restorations, should have a colour match comparable to that of the surrounding tooth tissue, which usually mostly comprises enamel. The size of the restoration is important as colour and translucency usually differ slightly in the incisal, middle and cervical thirds of the tooth. The operator light should be switched off and a distance of 60–100 cm is recommended for proper evaluation of colour match. If further details are required, sub-scores such as too opaque, too translucent, too dark, or too bright can be recorded (see Table 2).

A method to standardize digital photographs taken over time, with respect to colour, has been described [4]. By including a grey card in the picture as a neutral reference, colour casts can be eliminated and image brightness can be fine tuned using a standard image-editing programme before the relevant colour values are metered by the same software. A standard set of photographs showing colour deviations should be provided for comparison.

3.1 Colour and translucency of the restoration have a clinically excellent match with the surrounding enamel and adjacent teeth. There is no difference in shade, brightness or translucency between restoration and tooth.
3.2 Colour match is clinically acceptable but minor deviations in shade between tooth and restoration are apparent.
3.3 Colour match is satisfactory; there is a clear deviation in colour match that does not affect aesthetics.
3.4 Colour and/or translucency are clinically unsatisfactory. There is a (localized) discoloration or opaqueness in the restoration making it immediately recognisable from a speaking distance and affecting the appearance of the dentition. Partial removal and repair (veneering) is possible.
3.5 Colour match and/or translucency are clinically unsatisfactory. The restoration displays an unacceptable alteration in colour and/or translucency. Restoration needs replacement.

Anatomic form

The anatomic form of the restoration is evaluated for its effect on the overall aesthetic appearance, the aesthetic outcome being partly determined by an acceptable form. Wear will cause this to alter but as long as the change is not noticeable from a speaking distance the result is categorized as acceptable. In the case of chip fractures involving the form of the restoration, the aesthetic result, would not acceptable. The operator light should be switched off and a distance of 60–100 cm is recommended for proper evaluation.

4.1 Anatomic form is ideal.
4.2 Anatomic form deviates slightly from the remainder of the tooth.
4.3 Anatomic form differs from the homologous tooth but does not affect appearance; other irregularities in the dentition allow this to be aesthetically acceptable.
4.4 Anatomic form is altered, the aesthetic result is unacceptable. Correction is necessary
4.5 Anatomic form is unsatisfactory. Replacement of the restoration is necessary.

Functional properties

Fracture of restorative material and restoration retention

Restoration fracture and retention parameters are straightforward to assess. It is recommended to chart and characterize the localization of any cracks or chipping. A chip fracture is a small fracture with loss of material at the surface of the restoration. Chip fractures at baseline are not acceptable and should be reported, these restorations should not enter the study.

5.1 Restoration is present with no fractures, cracks or chipping
5.2 Small hairline cracks are visible
5.3 Several hairline cracks are present and/or limited chipping of material without damage to marginal quality or proximal contacts.
5.4 Fractures affect marginal quality and/or proximal contacts; bulk fractures with probable gap >250 μm with or without partial loss of less than half the restoration
5.5 Loss of restoration or bulk fracture with probable gap >250 μm with or without partial loss of the restoration.

Marginal adaptation

A distinction can be made between discoloration of the margin, as outlined under aesthetic properties, fractures of the enamel or dentine wall, as discussed under biological properties, and the presence of gaps or chipping of the material. Marginal morphology is of special interest for all

restorations and particularly for indirect restorations (luting gaps). Again, in general, magnifying aids such as loupes are recommended for evaluation.

a) Morphology

– Overfilled/Underfilled:

These criteria may provide information about the quality of the restoration placed by the clinician as well as the material itself in regards to its adaptability to the cavity margins. The evaluation should not exclusively focus on the restorative material; patient and/or dentist related reasons for failure should also be considered. The primary aim is to document marginal quality without specifically examining whether this is related to the material or the operator; further analysis can be done later if this is of interest.

The proximal area of the restoration should be examined with a blunt explorer with a 50 μm tip and dental floss. Overfilling or underfilling must especially be assessed during the baseline evaluation. Either the defect (underfilling or overfilling) is acceptable clinically and causes no long-term problems or the defect influences the normal function of the tooth. In the latter case the restoration has to be replaced. If this is noticed at baseline the restoration should not enter the study however it should be reported.

– Marginal deterioration including marginal degradation, irregularities, and gaps:

a) Marginal degradation and irregularities

Margins should be assessed quantitatively as a proportion (e.g. 25%) of the total length of the margin and should be differentiated as being located on the occlusal or proximal part of the restoration. Only similar sized restorations should be grouped for comparison. A method to easily quantify marginal deterioration is SQUACE (*S*emi *QUA*ntitative *C*linical *E*valuation). On an evaluation sheet (Fig. 1) sketches are made of the occlusal as well as the mesio-and disto-proximal parts of Class I and II restorations. The evaluator outlines the extent of the observed event on the sketch using different coloured pens according to defined criteria. The lines are then related to the size of the sketch and scored according to defined categories. Additionally, defects such as wear, fractures, tooth sensitivity, surface smoothness etc. may also be included in the evaluation sheet.

A comparative study evaluating 30 Class I/II composite restorations found a satisfactory correlation between the clinical criteria submargination and marginal discoloration as quantified by SQUACE, colour slides, and SEM evaluation of replica models, [26]. This method has been applied in several clinical trials of posterior composite restorations and has revealed more marginal discoloration in the proximal areas of Class II restorations than occlusally [64, 66, 68]. Greater marginal discoloration in proximal areas of restorations placed with a self-etching adhesive system were found compared to the same composite bonded with a conventional etch and rinse adhesive system (Heintze et al. unpublished data). If USPHS criteria had been used, defining the whole restoration as a unit, these differences would not have been detected.

b) Marginal gaps

The dichotomy in differentiation between continuous margins (defined as <2 μm gap) and gaps >2 μm is not a predictor for secondary caries or failure. Development of secondary caries has only been correlated to gaps >250 μm [61] and >400 μm for amalgam restorations [37]. A further study evaluating composite restorations in situ concluded that only a frank carious lesion is a reliable indicator of infected dentin beneath the restoration or at the margin [36]. Marginal deterioration and cavomarginal discoloration however may be prognostic for future failure [21]. To obtain better quality data for clinical prediction therefore, restoration gap width should be classified as a parameter for secondary caries development and restoration failure. A set of blunt explorers, straight and double angled for proximal sites, with different blunt tips of 50, 150 and 250 μm are recommended to be used to assess the gap size between tooth and restoration.

6.1 No clinically detectable gap. Margins represent a harmonious continuation of the outline at the tooth/restoration transition.

6.2 Marginal integrity deviates from the ideal, but could be upgraded to ideal by polishing. Small marginal chip fracture of the restoration can be eliminated by polishing and/or a localized gap is just perceptible with a dental probe >50 μm and <150 μm.

6.3 Leakage/discoloration is present but limited to the border area of the margins. Generalized marginal gap >150 μm but <250 μm, is easily perceptible on probing but cannot be modified without minor damage to the tooth or surrounding tissue, and is not considered to result in long-term negative consequences for the tooth or surrounding tissue if left untreated. Presence of several small marginal fractures that are unlikely to cause long-term effects.

6.4 Localized gap larger than 250 μm, may result in exposure of dentine or base. Repair is necessary for prophylactic reasons.

6.5 Generalized gap larger than 250 μm or the restoration is loose but in situ, replacement is necessary to prevent further damage or there are large fractures at the margins and loss of material is too extensive to be repaired.

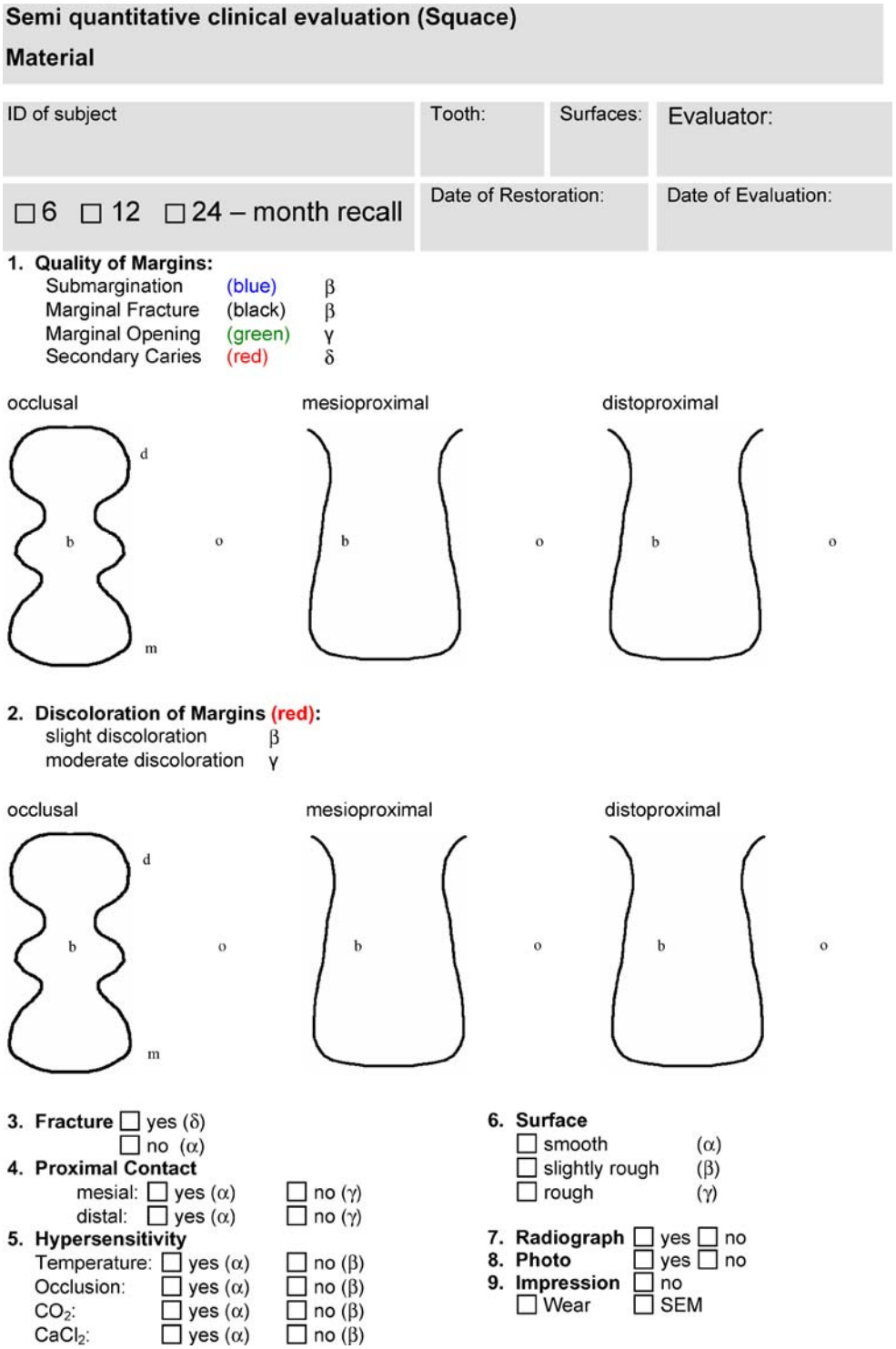

Fig. 1 Evaluation sheet for the semi-quantitative clinical evaluation of restorations. (The greek characters α, β, γ, δ relate to the USPHS criteria alpha, beta (bravo), gamma (charlie), delta)

Wear

a) Occlusal wear

Chemical degradation of a material in vivo can sometimes be difficult to differentiate from wear which is a physical phenomenon. Often there is a combination of the two resulting in loss of restorative material. If an overall loss of material from the occlusal surface is noted, as was seen with silicate cements and early resin composites for example, chemical as well as mechanical degradation can be considered to be the primary reason for material loss.

Wear cannot be measured quantitatively with clinical tools [82], indirect techniques are required. Semi-quantitative methods which link the occurrence and magnitude of

step formation at the occlusal interface of the restoration/enamel to wear by comparison of plaster replicas with standardized models showing defined steps, systematically underestimates wear, as has been shown in comparative studies [63]. Reliability and reproducibility of wear measurements of modern materials such as ceramics and resin composites cannot be recorded with old manual scale systems. It is not recommended to calculate mean values of wear with a precision of 1 or 0.1 micrometers if the original measurements were based on 50- or 100-μm steps. For example, if at the one year recall there are two restorations with a score of 100 μm wear; and the remaining 28 of 30 restorations score 0 μm; then it should be reported in this way. It is not scientifically sound to calculate and publish a mean wear value of 6.7 μm as this suggests an exactness which is not realistic.

Differentiation should be made between occlusal contact area and contact free area wear. Marginal ditching or deterioration is a different phenomenon to occlusal wear and should be evaluated separately. Wear is not homogeneous across the whole restoration surface so it is not representative to only measure a few points. Sophisticated 3-D-scanning of the whole occlusal surface of the restoration is recommended for quantitative measurement, however 3-D scanning is only feasible when excellent replica models are available.

– ***Qualitative wear measurement:***

Qualitative information is provided by photo-documentation of occlusal contact areas of the restoration and hard tissues registered with thin occlusal foil or by documentation of wear facets from models. For clinical semi-quantitative evaluation of wear, baseline photos can be compared with the clinical condition at each recall.

– ***Quantitative wear measurement:***

Wear must be measured indirectly if detailed information is required. Wear can be defined as 1) actual absolute wear on the occlusal contact areas (OCA) and contact free areas (CFA) and 2) as differential wear relative to enamel. Differential wear is material wear minus enamel wear at OCA as measured by 3D-Scanning of the total (occlusal) tooth surface. A differentiation between OCA wear marked with occlusal foil and documented on intra-oral images, and CFA wear is important. Volume loss of the whole occlusal surface should be reported relative to the whole measured surface to give an indication of generalized material degradation and wear. Vertical loss should be recorded at each OCA and averaged for all OCA. Differential wear calculations (compared to enamel) based on single surface points or on single profiles are not considered to be sufficiently representative. Only restorations with similar sizes and tooth types should be compared and within these groups the average wear rate is calculated for the material and referenced enamel at the OCA. If a difference of more than 300% (3x) is seen either in the restoration or the antagonistic enamel compared to the reference enamel, the wear is rated as unacceptable. Note that this statement is not evidence based and must be seen as a recommendation. The decision for replacement of such a restoration should also consider clinical assessment.

7.1 No difference in wear rate to enamel wear measured qualitatively or quantitatively with 3D equipment, or difference is in the range of 80–120% of the reference enamel wear

7.2 Minor difference in wear to that of enamel; or wear rate of restoration and antagonistic enamel is not less than 50% and not greater than 150% of the reference enamel.

7.3 Wear rate differs from enamel wear but is still within biological variation; or wear rate on restoration and antagonistic enamel is less than 50% or >150–300% of the reference enamel wear rate.

7.4 Wear rate significantly exceeds normal wear of enamel; occlusal contact points have been lost. Wear rate of restoration or enamel of opposing tooth >300% of reference enamel wear as measured quantitatively.

7.5 Wear rate is excessive and distinctly different from normal wear of enamel; wear rate on restoration or antagonistic enamel >500% of reference enamel when measured quantitatively.

b) Proximal wear

Proximal wear cannot be measured intraorally as the mesial drift of teeth often compensates for this, however in vivo wear and material degradation may cause loss of the proximal contact. If an open contact is present this is most often due to an operational problem during the placement of the restoration. If proximal wear is of specific interest, it can be measured using 3D equipment from casts which include several teeth in a quadrant, measurements being done at baseline and each recall.

Proximal contact point and food impaction

The proximal contact points can be checked by passing waxed dental floss through the interdental space. An appropriate degree of contact is necessary to prevent food impaction. A proximal contact point has physiological strength when dental floss or a 25 μm metal blade can pass through it and is evaluated for a certain degree of resistance or “snap” effect. Metal matrix strips of different thicknesses of 25 μm, 50 μm and 100 μm allow for a more precise determination than dental floss.

The firmness of the proximal contact can be initially established using waxed dental floss in order to avoid possible discoloration of the proximal surface as a result of a forced insertion of a metal matrix or blade. If the proximal contact is weak, thin metal blades of increasing thicknesses should be used to quantify the loss of contact. Lack of adequate contact leading to food impaction and discomfort during chewing is unacceptable. Surfaces with non-existing proximal contacts, e.g. teeth with a diastema, are excluded from this evaluation and scored as a missing value. The same is recommended for patients with advanced periodontitis and teeth which are graded as mobile or "loose". Teeth with non existing proximal contacts, e.g., absent neighbouring tooth, should be avoided in clinical studies of Class II restorations.

Food impaction related to open contacts and/or an inappropriate shape of the proximal restoration should be recorded.

8.1 Proximal contact is physiological, i.e., dental floss can only be inserted into the interdental space under pressure. A metal blade of 25 μm thickness can be inserted but not a 50 μm blade.
8.2 Proximal contact is slightly too strong but acceptable. Floss or 25 μm metal blade can only be passed through the contact with force/pressure.
8.3 Proximal contact is weak, a 50 μm metal blade can pass through the contact area but not a 100 μm blade, or floss passes very easily with only a slight snap effect. There is no indication for removing/repairing the restoration and there is no damage to the tooth, gingivae or other periodontal structures. There is no cervical caries, inflammation of the gingival papilla through food impaction, or pocket formation.
8.4 Proximal contact is weak and a 100 μm metal blade can easily pass through. In addition there are signs of damage to the tooth, gingivae or other periodontal structures (i.e. cervical caries, inflammation of the papilla through food impaction, pocket formation). Repair is necessary.
8.5 Proximal contact is weak allowing damage due to food impaction and demonstrating pain/gingivitis requires immediate intervention. Repair is not feasible and replacement is necessary.

Radiographic examination

Although radiographs of restorations may give valuable information with regard to gaps, secondary caries, overhangs, steps/underfilling and the level of alveolar bone; for ethical reasons routine radiographic examination for research purposes should only be performed if clinically indicated as a treatment need. Ideally the restorative material under test should have an adequate level of radiopacity. A threshold radiopacity of at least 100% aluminium is defined as sufficient in ISO Standard 4049 [32] however radiopacity of at least 200% Al is recommended for Class II restorations to clearly distinguish between hard tissues and restorative material [12]. If an older material shows no/ insufficient radiopacity this can be recorded as a missing value. For the detection of secondary caries on radiographs a semi-radiopaque material whose radiopacity slightly exceeds that of enamel seems most suitable [12].

9.1 Radiographic examination reveals no pathological findings. There is a harmonious transition between restoration and the tooth with no identifiable excess or insufficiency of restorative material or cement.
9.2 Radiographic examination reveals a small visible but acceptable excess (9.2.1) or/and a positive/negative step at the restoration margin is present <150 μm (9.2.2)
9.3 Marginal gaps <250 μm, and/or negative steps <250 μm (9.3.1) are identifiable with no clinically negative effects; removal is not possible due to their location or due to inadequate radiopacity of the restorative material (9.3.2).
9.4 Unacceptable, non adjustable marginal gaps >250 μm and/or marked interradicular excess material (9.4.1). Major intervention or repair is necessary to avoid damage to the tooth and adjacent tissues.
9.5 Radiographic examination reveals verifiable large gaps >500 μm and/or with the suspicion of secondary caries (9.5.1) or apical pathological changes (9.5.2) or severe tooth fracture or inlay fracture or loss of restoration (9.5.3). Replacement of restoration is necessary.

Patient satisfaction with restoration

Patient satisfaction with a restoration is by necessity a subjective response and is usually scored by means of a Visual Analogue Scale (VAS). Patient liking with regard to aesthetics, chewing comfort, pain/hypersensitivity, ease of ability to clean the restoration with toothbrush/dental floss, gingival bleeding or other problems such as detection of the restoration with the tongue, are items which are summarized under this topic.

10.1 Patient is entirely satisfied, would accept the same material again, is happy to recommend a restora-

tion using this material to others. The patient cannot detect the restoration with his/her tongue.

10.2 Patient is satisfied and would agree to the same material again. Patient can detect the restoration with his/her tongue but does not judge it to be disagreeable.

10.3 Patient criticizes aesthetic shortcomings, and/or lack of comfort when chewing. Repair or replacement of the restoration is not considered clinically necessary. Patient can detect the restoration with his/her tongue and judges the situation to be slightly annoying; restoration can be improved by simple refurbishing such as grinding or polishing.

10.4 Patient requests an improvement in the restoration, such as reshaping anatomic form or removal of discoloration. Patient's tongue is irritated or locally inflamed and the patient judges the situation to be annoying; simple refurbishing such as grinding or polishing cannot solve the problem

10.5 Patient is completely dissatisfied. There are objective reasons to support this position, the restoration has to be replaced immediately to prevent further adverse effects and/or pain. Patient does not want to have the same material or type of restoration again.

Biological properties

Pulp: postoperative sensitivity and tooth vitality

The terms postoperative sensitivity and postoperative hypersensitivity are used interchangeably in many publications. Postoperative sensitivity may perhaps be misunderstood in some cases, e.g. in the context of no sensitivity due to postoperative loss of vitality, therefore, the term 'hypersensitivity' is preferred.

Postoperative hypersensitivity is recorded at the time of restoration placement, at baseline and at all recall visits, and can include type of pain, discomfort and duration and/or on stimulus at clinical assessment. Intensity can be assessed with a VAS. Often differing pulp responses can be noted between placement and baseline. Vitality can be tested with application of cold (dry ice) and should always be compared with the reaction of adjacent vital teeth. Transient pain elicited on stimulation is acceptable, persistent pain renders the restoration unacceptable and requires intervention to alleviate the problem. When normal sensitivity/vitality is recorded at restoration placement, and is subsequently lost or alteration occurs at baseline or at subsequent recall, the restoration should be rated as unacceptable. Great care must be taken in diagnosing pulp death and the restorative history of the tooth, such as former pulp capping should be taken into account. In clinical studies postoperative hypersensitivity is usually a secondary endpoint. Studies looking at postoperative hypersensitivity as a primary endpoint should follow a specific protocol for that purpose.

11.1 No postoperative hypersensitivity. Normal pulp vitality response.

11.2 Postoperative hypersensitivity of short duration (less than one week) and no longer present at the baseline assessment. Pulp vitality response normal at the baseline assessment (one week after placement).

11.3 Intense postoperative hypersensitivity of greater duration than one week but less than six-months. At baseline assessment, response to cold stimulus is premature/strong (11.3.1.) or delayed/weak (11.3.2.) but normal function is still present based on subjective patient comments and the clinical condition is unremarkable. Occlusal adjustment may be required.

11.4 Persistent postoperative hypersensitivity. Response to cold stimulus is markedly premature/strong (11.4.1.) and major intervention is necessary; or there is extremely delayed/weak or negative sensitivity (11.4.2.). Sensitivity level is significantly different from the situation prior to treatment. The clinical condition and patient complaint make imminent treatment a priority, an additional waiting period, and/or occlusal adjustment, and/or use of desensitizing products, and/or alteration of eating habits will be of no help. If pulp treatment is planned repair of the restoration should be considered. Tooth is removed from study and documented.

11.5 Negative sensitivity recorded at recall visit despite positive pulp response at baseline, or severe pain is noted. Removal of restoration with or without immediate root canal treatment is required or the tooth must be extracted. Tooth is removed from study and documented.

Recurrence of initial pathology

Recurrence of previous pathology and/or new pathology at the restoration margins such as caries, erosion, or abfraction that cannot be alleviated by a minor intervention should be scored as unacceptable. Active recurrent caries or erosion that cannot be treated by remineralization and has to be treated operatively is given an unacceptable score. Evaluation is done by visual assessment; if quantitative or semi-

quantitative measuring tools are used they should be described and validated.

Initial CAR (secondary caries) not requiring repair/replacement is recorded when there is visible demineralization without cavitation in tooth tissue adjacent to the restoration, this includes opacity and/or brown discoloration of arrested caries, which cannot be polished away. Care must be taken to distinguish defects from stained margins. Cavitation in the adjacent tooth tissue indicates established secondary caries and consequently the need for operative intervention, such as repair or replacement. The recommendations of ICDAS [30] should be used when diagnosing CAR.

12.1 No recurrence of initial pathology and no other pathology present.
12.2 Small/localized area of demineralization but no operative treatment required.
12.3 Areas of demineralization, erosion or abrasion/abfraction are noted, preventive measures only are necessary, and dentine is not exposed
12.4 Recurrence of initial or other pathology, including cavitated caries, erosion or abrasion/abfraction in dentine, is more localized and accessible and can be restored / repaired by operative intervention.
12.5 Severe recurrence of initial pathology or other pathology, generalized or localized such as deep caries or exposed dentine that is not accessible for repair and requires immediate restoration replacement.

Tooth cracks and fractures

The restoration is visually examined for splits/cracks and fractures. Provided no clinical symptoms are present, signs of tooth enamel cracks are acceptable. If a major intervention such as repair or replacement is needed to remedy the consequences of these defects, the restoration is rated as unacceptable. Assessment is done visually; if quantitative or semi-quantitative measuring tools are used they should be described and validated.

13.1 Complete integrity of the restored tooth.
13.2 Minor marginal crack <150 μm wide (13.2.1) or a hairline crack (13.2.2) which cannot be probed. The patient has no clinical symptoms.
13.3 Marginal split (13.3.1) in the enamel <250 μm wide. Not possible to remove by polishing without compromising the shape of the tooth surface or damaging the tooth and is left untreated as not expected to cause further damage. Crack (13.3.2) <250 μm. Patient has no or minimal discomfort.
13.4 Major marginal split >250 μm that requires repair and/or dentine or base exposed (13.4.1). A 250 μm probe/explorer can be inserted into a crack (13.4.2) and the defect requires treatment.
13.5 A cusp or major tooth fracture requiring immediate replacement.

Effect of the restoration on the periodontium

As a first step any inflammation of the papilla should be treated conventionally. If the problems can be resolved the restoration remains acceptable, if the restoration needs major intervention, more than simple refurbishing, the restoration is scored as unacceptable.

Use of the Papillary Bleeding Index (PBI scale 0–4) [71] is recommended. The Index can be modified per restored tooth and compared to a matching unrestored tooth in the same patient. An increase in pocket depth >1.0 mm in comparison to a reference tooth should be considered as noteworthy and the restoration carefully examined for a possible causal relationship.

14.1 No plaque, no inflammation of the gingival papilla.
14.2 Minimal plaque is present, PBI equivalent to baseline
14.3 Difference up to one grade in severity of PBI compared to baseline and in comparison to control tooth.
14.4 Difference of more than one grade of PBI worsening in comparison to control tooth or increase in pocket depth >1 mm requiring major intervention
14.5 Severe/acute gingivitis or periodontitis if related to the restoration requiring immediate replacement of the restoration.

Localized reactions of soft tissue in direct contact with the restoration

A soft tissue contact allergic reaction may occur in relation to a restorative material. If the mucosa adjacent to a restoration has signs of an allergic reaction caused by the restorative material, including mild or severe redness and swelling and/or lichenoid reaction, and/or if there are indications of allergy based on the patient's history, the restoration should be removed and replaced by a material non-allergenic for the patient. This condition must be recorded as a failure. The study examiners require special training on these rarely occurring events, alternatively

experienced specialists may be included in the examiner group. Since the allergic reaction is inflammatory in nature, it is difficult to differentiate local contact allergy from local physical irritation.

15.1 Healthy soft tissue surrounding the restored tooth.
15.2 Healthy soft tissue associated with the restored tooth after minor removal of mechanical irritations such as sharp edges
15.3 Slight alteration of mucosa but clinically acceptable with suspicion of a causal relationship to the restorative material.
15.4 Suspected mild allergic, lichenoid or toxic reaction. Veneering of parts of the restoration in direct contact with the soft tissues may improve the situation.
15.5 Established allergic, lichenoid or toxic reaction; restoration has to be removed immediately.

Oral and somatic/psychiatric symptoms

True or suspected oral soft tissue pathological findings associated with the restored tooth, such as contact allergy, should be recorded as part of the previous section.

The patient's medical history should be evaluated, including psychiatric history, systemic disease and any medications used. The patient should be questioned as to any previous experiences of allergy. If a problem persists such as oral lichen planus or systemic reaction, and the restoration has to be replaced, it should receive an unacceptable score.

All serious unanticipated adverse events should be reported to the Ethics Committee/IRB, and to the study sponsor.

16.1 No oral or general symptoms of adverse effects.
16.2 Short-term and minor transient symptoms localized or generalized of known or unknown origin.
16.3 Minor oral and/or general symptoms of malaise e. g. lichen planus simplex, transient symptoms of inflammatory reactions.
16.4 Persistent oral or general symptoms, or recurrent symptoms, symptoms of oral contact stomatitis or lichen planus, or allergic reaction requiring veneering or replacement of the restoration in the near future using a different material.
16.5 Acute allergic reaction, acute/severe oral or general symptoms requiring medical consultation which will also consider toxic or psychiatric effects; restoration requires immediate replacement using a different restorative material and may be best done in consultation with a medical specialist.

Indirect restorations

The technical functionality of an indirect metal or non-metal restoration such as an inlay, onlay, partial crown or veneer should be evaluated prior to luting / cementation; this should be done both on the model *and* intraorally. If the restoration is deemed unacceptable it should be modified or remade. Criteria for assessment of indirect restorations consist of five categories comparable to clinical assessment of all restorations (Table 2), with three acceptable scores and two unacceptable. For fragile restorations such as ceramic inlays, checking the occlusal contact areas is not recommended prior to luting.

IR.1 Restoration has an ideal fit both on the model and the tooth. Colour and form exactly match and harmonise with the tooth
IR.2 Colour is not ideal but the difference is clinically acceptable. Minor marginal discrepancies may be present.
IR.3 Marginal adaptation or form is deficient but this does not affect normal function after cementation, the tooth is not compromised in any way. Or colour is not ideal but still acceptable.
IR.4 Proximal or occlusal contacts are insufficient but can be improved by the clinician or dental technician, or colour must be slightly adapted. (These corrections/modifications should be reported.)
IR.5 Marginal adaptation or anatomic form is deficient on the model and/or on the tooth and will cause dysfunction. Or the colour is clinically unacceptable. The restoration cannot be fitted and must be remade.

Only restorations scoring IR 1 to IR 3 should be inserted in the patient's mouth. Documentation is required if restorations need to be modified/corrected or remade.

At the recall assessments for indirect restorations the luting material should be evaluated, including wear and staining. Voids in the luting space should be recorded. It is clinically difficult, if not impossible however to differentiate between gaps at the interface between luting material and hard tissues, and between luting material and restoration. These assessments require indirect methods of evaluation such as SEM examination.

Miscellaneous

– Restorations with inserts

In the case of restorations with inserts, wear, marginal adaptation and insert retention loss should be recorded using the appropriate criteria as outlined for direct restorations.

– Teeth in the opposing arch

The teeth opposing a study restoration should be described with respect to the presence of enamel and type of restoration, if present, and should be documented at baseline. The placement or replacement of restorations in opposing teeth during the observation period should also be reported. This is especially important for fracture and wear analyses.

Some aspects of statistical analysis

Overall failure rate

The number and reasons for patient dropout should be reported, including the number of patients and restorations at the end of each observation period in order to calculate an accurate failure percentage. The overall failure rate should always be reported in a clinical study paper. The total number of restorations which have been and still require to be replaced and/or to be repaired should be stated, also the reasons for restoration replacement and repair.

All failures have to be reported, including failures from previous recalls. The number of failures from earlier recalls should be added to the total number of teeth at the last recall to calculate the correct failure percent.

Comparison of the treatment and control groups only with respect to the observed percentage of failures is usually considered suboptimal statistical practice. In some studies, the follow-up time differs between subjects due to dropout or because of late entry into the study. A subject with a late occurring failure may ultimately have a better outcome than a subject with no failure who was followed up for only a short period of time.

It is generally preferable and more powerful to statistically analyze the continuous outcome "time to failure", and hence to perform a "survival analysis", rather than the binary outcome "failure/no failure". For subjects with a failed restoration it would be preferable to record precisely when the failure occurred. It is imprecise to report that the restoration failed between the 12 and 24 month visit. In restorative dentistry however usually only those failures which are detected by the patient, such as pain, retention loss or fracture, can be recorded with any precision as to date of occurrence. Other failures may only be detected at the scheduled recall visit. Subjects with no observed failure due to loss to follow-up are said to be "censored". For these subjects, the time where a failure would have occurred is unknown and a lower bound may be applied, for example more than 36 months after receiving treatment.

Both censored and uncensored data are used in a survival analysis. The usual graphical tool to describe survival data is the well-known Kaplan-Meier plot. The estimated percentage of survival versus the percentage of subjects without failure is plotted for each unit of observation time, starting from 100 at time 0 and then remaining constant or sequentially decreasing. Separate Kaplan-Meier curves should be calculated for the treatment and control groups. From these curves, one may extract useful information such as the estimated percentage of subjects without failure after a given time interval, or the estimated time at which half the subjects will have a failed restoration, the estimated median survival time. The log-rank test is used to statistically compare the groups, in addition a multivariate analysis (to adjust for confounding variables, for example), and a Cox regression model, also referred to as proportional hazards model [9], can be calculated.

To give a fictitious example of how a Kaplan-Meier curve can easily be calculated for a group of patients, without using a computer or statistical programme, let us assume that a group of 25 subjects have each received a new material to restore one Class II cavity per subject, and the subjects were followed for up to 72 months after treatment, being recalled for evaluation every 12 months. Let us assume that a total of six restorations failed, one each at 15, 23, 23, 54, 66 and 68 months. Finally, let us assume that three subjects were lost to follow up, subject 1 not attending the first recall (hence this observation being censored at 0 months), subject 2 missing the third recall (hence censored at 24 months) and subject 3 missing the fifth recall (hence censored at 48 months). Since the remaining 16 subjects had no failure at the last recall, they were censored at 72 months. Such data can be reported as follows (where the symbol + denotes a censored observation):

0+ 15 23 23 24+ 48+ 54 66 68 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+ 72+

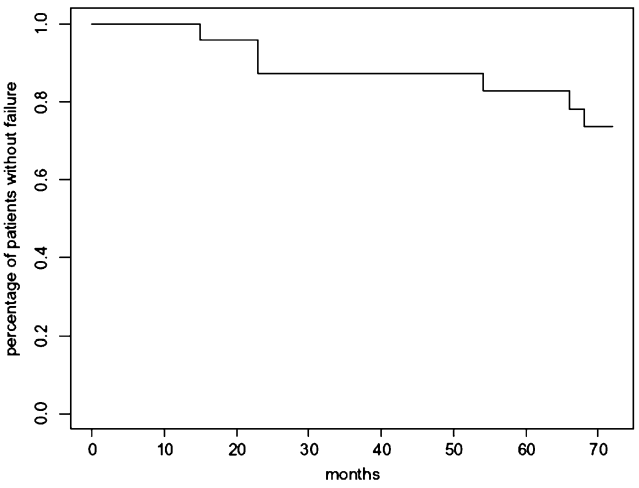

Fig. 2 Kaplan-Meier survival curve showing the percentage of patients without failure

From these data, one can calculate the quantities found in the following Table (Table 3):

Observed failure time	Number of failures	Patients followed up at that time	Observed percentage without failure	Kaplan-Meier estimate
15	1	24	23/24	$\frac{23}{24} \approx 0.95$
23	2	23	21/23	$\frac{23 \cdot 21}{24 \cdot 23} \approx 0.88$
54	1	19	18/19	$\frac{23 \cdot 21 \cdot 18}{24 \cdot 23 \cdot 19} \approx 0.83$
66	1	18	17/18	$\frac{23 \cdot 21 \cdot 18 \cdot 17}{24 \cdot 23 \cdot 19 \cdot 18} \approx 0.78$
68	1	17	16/17	$\frac{23 \cdot 21 \cdot 18 \cdot 17 \cdot 16}{24 \cdot 23 \cdot 19 \cdot 18 \cdot 17} \approx 0.74$

Coordinates of the corresponding Kaplan-Meier curve are found in the first and in the last column of this Table. Such a curve is plotted in Fig. 2.

Conclusions

Under the auspices of evidence-based medicine and dentistry it is highly recommended that future clinical studies for testing new restorative materials or procedures be conducted as randomized controlled trials with a clear hypothesis and protocol description to allow subsequent inclusion in meta-analyses. A grouping of aesthetic, functional and biological categories has been proposed to simplify the clinical evaluation procedure, while at the same time allowing a more detailed analysis of failures. Depending on the aims and design of a specific study, the assessments used may differ in some aspects from those outlined in this paper, and not all parameters listed in Table 2 need to be included in every study. Every study however should include those categories which universally require evaluation, and should use criteria which can subsequently be compared. It is highly recommended that these criteria be adopted in future studies. On publication, a description of the overall study failure rate at specific observation periods should be given and, if feasible, a survival analysis (e.g. Kaplan-Meier).

Journal editors should make adequate space available for authors of RCT studies to be able to fully report on all the data.

This paper is a living document which will require regular updating. The present recommendations should be reviewed every three to four years for modifications based on available evidence. Input from all investigators is welcome.

Final remark

It is expected that a set of hand instruments specifically for clinical evaluation of dental restorations will soon be available. These will include explorers/probes with defined tip thicknesses to categorize marginal gaps; metal blades with defined thicknesses for evaluation of proximal contact areas; and high quality surface coated dental mirrors.

References

1. Antczak-Bouckoms AA, Tulloch JFC, Berkey CS (1990) Split mouth and cross over design in dental research. J Clin Periodontol 17:446–453
2. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC (1992) A comparison of results on meta-analyses of randomized controlled trials and recommendations of clinical experts: treatments for myocardial infarction. J Amer Med Assoc 268:240–248
3. Bayne SC, Schmalz G (2005) Reprinting the classic article on USPHS evaluation methods for measuring the clinical research performance of restorative materials. Clin Oral Investig 9(4):209–214
4. Bengel WM (2003) Digital photography and the assessment of therapeutic results after bleaching procedures. J Esthet Restor Dent 15(Suppl 1):S21–S32
5. Blum IR, Schriever A, Heidemann D, Mjör IA, Wilson NH (2003) The repair of direct composite restorations: an international survey of the teaching of operative techniques and materials. Eur J Dent Educ 7(1):41–48
6. Charbeneau GT (1981) Principles and practice of operative dentistry. Lea and Febiger, Philadelphia
7. Clark TD, Mjör IA (2001) Current teaching of cariology in North Amereican dental schools. Oper Dent 26:412–418
8. Collet D (1994) Modelling survival data in medical research. Chapman and Hall, London
9. Cox D (1972) Regression models and life tables. J R Stat Soc 34:187–220
10. Cvar J, Ryge G (1971) Criteria for the clinical evaluation of dental restorative materials. US DHEW Document, US Public Health Service 790244, Printing Office, San Francisco, pp 1–42 (and reprinted as Cvar J, Ryge G. Reprint of Criteria for the clinical evaluation of dental restorative materials. Clin Oral Invest 2005(9):215–252
11. Elderton RJ (1993) Overtreatment with restorative dentistry: When to intervene? Int Dent J 43:17–24
12. Espelid I, Tveit AB, Erickson RL, Keck SC, Glasspoole EA (1991) Radiopacity of restorations and detection of secondary caries. Dent Mater 7(2):114–117
13. EU (1993) Council directive concerning medical devices. Directive 93/42 EEC
14. Fitzgerald RJ, Adams BO, Davis ME (1994) A microbiological study of recurrent dentinal caries. Caries Res 28(6):409–415
15. Gaengler P, Hoyer I, Montag R, Gaebler P (2004) Micromorphological evaluation of posterior composite restorations - a 10-year report. J Oral Rehabil 31(10):991–1000

16. Going RE (1972) Microleakage around dental restorations: a summarizing review. J Am Dent Assoc 84(6):1349–1357
17. Goldberg AJ (1990) Deterioration of restorative materials and the risk for secondary caries. Adv Dent Res 4:14–18
18. Gordan VV, Shen C, Mjör IA (2004) Marginal gap repair with flowable resin-based composite. Gen Dent 53:390–394
19. Gordan VV, Shen C, Riley J, Mjör IA (2006a) Two-year clinical evaluation of repair versus replacement of composite restorations. J Esthet Restor Dent 18(3):144–153
20. Gordan VV, Riley J, Blaser PK, Mjör IA (2006b) 2-year clinical evaluation of alternative treatments to replacement of defective amalgam restorations. Oper Dent 31:418–425
21. Hayashi M, Wilson NH (2003) Failure risk of posterior composites with post-operative sensitivity. Oper Dent 28(6): 681–688
22. Hayashi M, Tsuchitani Y, Kawamura Y, Miura M, Takeshige F, Ebisu S (2000) Eight-year clinical evaluation of fired ceramic inlays. Oper Dent 25(6):473–481
23. Heintze SD (2007) The correlation between the evaluation of marginal quality and bond strength and their clinical relevance - a systematic review. J Adhes Dent;accepted for publication
24. Heintze SD, Twetman S (2002) Interdental mutans streptococci suppression in vivo: a comparison of different chlorhexidine regimens in relation to restorative material. Am J Dent 15(2): 103–108
25. Heintze SD, Bastos JRM, Roversi MJ (1997) The prevalence of dental caries and fluorosis in 6-to-44-year-olds in 2 Brazilian cities with and without water fluoridation. 6th World Congress on Preventive Dentistry, Cape Town, South Africa
26. Heintze SD, Müssner R, Reich E (2000) A new method for the clinical evaluation of direct posterior restorations. J Dent Res 79:186 (Abstract No 341)
27. Hewlett ER, Atchison KA, White SC, Flack V (1993) Radiographic secondary caries prevalence in teeth with clinically defective restorations. J Dent Res 72(12):1604–1608
28. Hickel R, Manhart J (2001) Longevity of restorations in posterior teeth and reasons for failure. J Adhes Dent 3(1):45–64
29. Hickel R, Manhart J, Garcia-Godoy F (2000) Clinical results and new developments of direct posterior restorations. Am J Dent 13 (Spec No):41D–54D
30. ICDAS (2005) Rationale and evidence for the International Caries Detection and Assessment System (ICDAS II). http://www.icdas.org Retrieved July 31, 2006
31. ISO (1997) Dentistry-Preclinical evaluation of biocompatibility of medical devices used in dentistry-Test methods for dental materials. International Standard:No. 7405
32. ISO (2000) Dentistry - Polymer-based filling, restorative and luting materials. International Standard No. 4049
33. ISO (2003) Dental materials-Testing of adhesion to tooth structure. Technical Specification:No. 11405
34. Kersten S, Lutz F, Besek M (1999) Zahnfarbene adhäsive Füllungen im Seitenzahnbereich Zürich: Eigenverlag PPK
35. Kidd EA (1990) Caries diagnosis within restored teeth. Adv Dent Res 4:10–13
36. Kidd EA, Beighton D (1996) Prediction of secondary caries around tooth-colored restorations: a clinical and microbiological study. J Dent Res 75(12):1942–1946
37. Kidd EA, Joyston-Bechal S, Beighton D (1995) Marginal ditching and staining as a predictor of secondary caries around amalgam restorations: a clinical and microbiological study. J Dent Res 74(5): 1206–1211
38. Köhler B, Rasmusson CG, Odman P (2000) A five-year clinical evaluation of Class II composite resin restorations. J Dent 28:111–116
39. Krasse B (1989) Prediction and prevention of recurrent caries based on microbiological assays. In: Anusavice KJ (ed) Quality evaluation of dental restorations. Quintessence, Chicago, pp. 199–208
40. Manhart J, Hickel R (2001) Longevity of restorations. In: Wilson HF, Roulet J-F, Fuzzi M (eds) Advances in operative dentistry-Challenges of the future. Vol 2. Quintessence, Chicago, pp. 237–304
41. Manhart J, Chen H, Hamm G, Hickel R (2004) Buonocore memorial lecture. Review of the clinical survival of direct and indirect restorations in posterior teeth of the permanent dentition. Oper Dent 29(5):481–508
42. Matthews DE, Farewell VT (1996) Using and understanding medical statistics. Karger, Basel
43. Merrett MC, Elderton RJ (1984) An in vitro study of restorative dental treatment decisions and dental caries. Br Dent J 157(4): 128–133
44. Mjör IA (1985) Frequency of secondary caries at various anatomical locations. Oper Dent 10(3):88–92
45. Mjör IA (1998) The location of clinically diagnosed secondary caries. Quintessence Int 29(5):313–317
46. Mjör IA (2005) Clinical diagnosis of recurrent caries. J Am Dent Assoc 136(10):1426–1433
47. Mjör IA, Gordan VV (2002) Failure, repair, refurbishing and longevity of restorations. Oper Dent 27(5):528–534
48. Mjör IA, Haugen E (1976) Clinical evaluation of amalgam restorations. Scand J Dent Res 84:333–337
49. Mjör IA, Qvist V (1997) Marginal failures of amalgam and composite restorations. J Dent 25:25–30
50. Mjör IA, Toffenetti F (2000) Secondary caries: a literature review with case reports. Quintessence Int 31(3):165–179
51. Moher D, Jones A, Lepage L (2001a) Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. JAMA 285(15):1992–1995
52. Moher D, Schulz KF, Altman D (2001b) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 285(15):1987–1991
53. Mjör IA, Gordan VV, Abu-Hanna A, Gilbert GH (2005) Research in general dental practice. Acta Odontol Scand 63(1):1–9
54. Moncado G, Martin J, Fernandez E, Vildosola P, Camano C, Mjör IA, Gordan VV (2006) Alternative treatments for resin based composite and amalgam restorations with marginal defects: 12-month clinical trial. Gen Dent 54:314–318
55. Mühlemann HR, Son S (1971) Gingival sulcus bleeding-a leading symptom in initial gingivitis. Helv Odontol Acta 15(2):107–113
56. Noack MJ, Van Meerbeek B, Roulet JF, Lambrechts P (1995) Marginal behavior of 10 different tooth-colored inlay systems in vivo. J Dent Res (Spec Iss) 74:165 (Abstract 1227)
57. Opdam NJ, Feilzer AJ, Roeters JJ, Smale I (1998a) Class I occlusal composite resin restorations: in vivo post-operative sensitivity, wall adaptation, and microleakage. Am J Dent 11(5): 229–234
58. Opdam NJ, Roeters FJ, Feilzer AJ, Smale I (1998b) A radiographic and scanning electron microscopic study of approximal margins of Class II resin composite restorations placed in vivo. J Dent 26(4):319–327
59. Opdam NJ, Roeters FJ, Feilzer AJ, Verdonschot EH (1998c) Marginal integrity and postoperative sensitivity in Class 2 resin composite restorations in vivo. J Dent 26(7):555–562
60. Owen R (1982) Reader bias. JAMA 247(18):2533–2534
61. Özer L (1997) The relationship between gap size, microbial accumulation and the structural features of natural caries in extracted teeth with class II amalgam restorations: Thesis, University of Copenhagen
62. Pallesen U, Qvist V (2003) Composite resin fillings and inlays. An 11-year evaluation. Clin Oral Investig 7(2):71–9. Epub 2003 May 10
63. Perry R, Kugel G, Kunzelmann KH, Flessa HP, Estafan D (2000) Composite restoration wear analysis: conventional methods vs. three-dimensional laser digitizer. J Am Dent Assoc 131(10):1472–1477

64. Peschke A, Heintze SD, Rheinberger V (2003) Clinical evaluation of large posterior composite restorations bonded with a new self-etching adhesive system: 6-month results. J Dent Res(Spec Iss B):195 (Abstract No.1468)
65. Peschke A, Heintze SD, Roulet JF (2005) Comparison of two impression methods for clinical wear measurement. J Dent Res (Spec Iss B) 84(Abstract No 350(Continental European and Scandinavian Divisions) (http://www.dentalresearch.org)
66. Peschke A, Heintze SD, Roulet JF (2007) Two-year clinical evaluation and wear analysis of posterior composite restorations. J Den Res 86 (Spec Issue A):Abstract No. 230
67. Peumans M, Van Meerbeek B, Asscherickx K, Simon S, Abe Y, Lambrechts P, Vanherle G (2001) Do condensable composites help to achieve better proximal contacts? Dent Mater 17:533–541
68. Rathke A, Heintze SD (2001) Prospective clinical trial on a posterior composite with a one-bottle adhesive. J Dent Res 80:1282 (Abstract No 114)
69. Roulet JF, Reich T, Blunck U, Noack M (1989) Quantitative margin analysis in the scanning electron microscope. Scanning Microsc 3(1):147–159
70. Sarrett DC (2005) Clinical challenges and the relevance of materials testing for posterior composite restorations. Dent Mater 21(1):9–20
71. Saxer UP, Mühlemann HR (1975) Motivation und Aufklärung. Schweiz Mschr Zahnheilk 85:905–909
72. Schmidlin P, Hoffer E, Lutz F (2002) Evaluation of published clinical studies for reproducibility, comparability and adhesrence to evidence-based methods. Am J Dent 15:26–30
73. Shen C, Mondragon E, Gordan VV, Mjör IA (2004) The effect of mechanical undercut on the strength of composite repair. J Amer Dent Assoc 135:1406–1412
74. Shen C, Speigel J, Mjör IA (2006) Repair strength of dental amalgams. Oper Dent 31:122–126
75. Shen C, Mondragon E, Mjör IA (2007) Effect of size of defect on the repair of amalgam. Quintessence Int (in press)
76. Silness L, Löe H (1964) Periodontal disease in pregnancy. II. Correlation between oral hygiene and periodontal condition. Acta Odontol Scand 22:121–135
77. Smales RJ, Hawthorne WS (2004) Long-term survival of repaired amalgams, recemented crowns and gold castings. Oper Dent 29(3):249–253
78. Söderholm K-JM, Antonson DE, Fischlschweiger W (1989) Correlation between marginal discrepancies at the amalgam/tooth interface and recurrent caries. In: Anusavice KJ (ed) Quality evaluation of dental restorations. Quintessence Publ., Chicago, pp. 95–108
79. Söderholm KJ, Roberts MJ, Antonson DE, Anusavice KJ, Mauderli AP, Sarrett DC, Warren JW (1992) Visual and profilometric wear measurements. Acta Odontol Scand 50(2):121–127
80. SSO (2000) Qualitätsleitlinien in der Zahnmedizin Bern: SSO
81. Svanberg M, Mjör IA, Ørstavik D (1990) Mutans streptococci in plaque from margins of amalgam, composite, and glass-ionomer restorations. J Dent Res 69(3):861–864
82. Taylor DF, Bayne SC, Sturdevant JR, Wilder AD (1989) Comparison of direct and indirect methods for analyzing wear of posterior composite restorations. Dent Mater 5(3):157–160
83. Tveit AB, Espelid I, Erickson RL, Glasspoole EA (1991) Vertical angulation of the X-ray beam and radiographic diagnosis of secondary caries. Community Dent Oral Epidemiol 19(6):333–335
84. Vanherle G, Hickel R (2007) In vivo testing of anterior filling materials. Eurocondensor Vol 9(1):16–20