

Review Article

Dental restorations for oral rehabilitation – testing of laboratory properties versus clinical performance for clinical decision making*

S. C. BAYNE *Department of Cariology, Restorative Sciences, and Endodontics, School of Dentistry, University of Michigan, Ann Arbor, MI, USA*

SUMMARY At the outset, the categories of physical, chemical, mechanical, biological and clinical properties of biomaterials are reviewed in terms of their definitions and relevant examples. Clinical performance for restorative materials is considered in terms of five crucial categories of factors (operator, design, materials, site and patient). Clinical performance assessment in actual clinical trials is described in terms of United States Public Health Service (USPHS) and modified USPHS categories of ratings collected from direct observations. Clinical failure analysis is characterized using reverse s-shaped curves to summarize longevity (failure or success) and clinical longevity for 50% failures (CL_{50}) is defined. Actual practice effectiveness is demonstrated as being approximately

one-half of clinical trial efficacy. Types of restorative dental material clinical trials are contrasted (longitudinal versus cross-sectional, short-term versus long-term, university-based versus practice-based research networks). Poor correlations between laboratory test values and clinical performance are explained. The need for risk assessment is emphasized. Evidence-based dentistry is defined in terms of available published information and precautions. At this point, the evidence base for clinical performance of biomaterials is scant.

KEYWORDS: dental materials, biomaterials, restorative dentistry, clinical performance, longevity, clinical research

Accepted for publication 22 August 2007

Introduction

In clinical practice, a dentist encounters a certain amount of frustration when he or she is trying to negotiate the plethora of treatment choices and complicated options for various dental materials. A constant series of questions ensues each time. *Which material should I use? Which one is it the best? How long will it last?* The response is always categorized into a series of factual comments that often provide little or no information to truly answer these questions. *Use the material with the least complicated techniques. Use this material because it is the strongest. There are no clinical*

research data about this recommendation but the company is a reliable one in terms of product quality. The focus of the following discourse is to decipher what all of this information really means and decide if any real answers can be provided. While someday it might be possible to approach this topic from the point of the desirable clinical properties and their connection to specific laboratory properties, the limitations of both make this approach unworkable at the moment. The following are the three main thrusts: (i) explaining biomaterials properties relative to clinical decision making, (ii) examining the relationship between laboratory testing and predictions of clinical performance and (iii) interpreting the real value of any biomaterials information as evidence for evidence-based decision making.

*Based on a lecture given at the JOR Summer School 2006 sponsored by Blackwell Munksgaard.

Biomaterials research

Biomaterials properties

There are many schemes used to classify all the properties associated with any biomaterial. However, the one that seems to collect properties in the most meaningful way (1, 2) utilizes the categories of *physical, chemical, mechanical, biological and clinical properties* of materials. In materials science and dental materials science, most often we are concerned with properties of solids and occasionally liquids. Consider the technical definitions for each category and examples of actual properties [see also Mjør 2007 (3)].

Physical properties are those that involve motion of electrons, protons, or atoms within the solid, but which do not involve any major changes in bonding patterns or types. One example would be the electrical conductivity of a material. Electrons may move as a part of the current through the material, but the material remains as a metallic solid with the same atomic positions throughout. The category of physical properties can be subdivided into thermal, electrical, optical and mass physical properties. Examples would include thermal conductivity, thermal diffusivity, thermal expansion, reflectivity, radiopacity, density and others.

Chemical properties are ones that involve changes in bonding patterns and/or hydration states of the atoms or molecules on the surfaces or within the interior of the material. Examples would include water adsorption (onto the surface), water absorption (into the interior), chemical corrosion, electrochemical corrosion, biodegradation and/or new chemical reactions.

Mechanical properties are those which involve the ways in which a material responds to load. All mechanical properties are normalized (stress, strain, modulus) and they depend on the direction of loading (compressive, tensile, shear), are time dependent (static versus dynamic, strain-rate sensitivity), may involve cyclic loading (fatigue) and involve considerations of deformation that may not be uniform in all directions (Poisson's ratio). For any particular application, it is crucial to discern the direction of loading (e.g. tensile) so as to know the strength of interest (e.g. tensile strength), strain-rate sensitivity and expected life-time in fatigue. Even if someone is diligent enough to commit to long-term cyclic testing to understand fatigue, it is hard to represent the full range of environment variations that might occur in the mouth.

In frustration, many investigators have created simulators that pretend to represent different combinations of environmental variables and produce responses typical of perhaps, 3 years of clinical performance. Generally, examining the time-dependent changes in mechanical properties is the focus of these efforts.

Biological properties of biomaterials are those that represent interfacial interactions of a material with the hard and soft tissues and which may produce local or systemic responses in the patient. Generally these are ascertained through a series of cell culture tests, tissue culture tests, small animal model tests, or human usage tests in mammals. Examples of biological properties include toxicity, sensitivity and mutagenicity tests. Little information has been accumulated directly for most biomaterials. Rather, materials have been screened historically as clinically acceptable or not, without fully understanding the details of biological activity. Materials may be potentially toxic but may not produce any major side-effects in use. Almost all reactions depend on a combination of dose and time in determining the observed response.

Clinical properties or clinical performance is defined in terms of safety and effectiveness. Effectiveness is subdivided into a series of clinical assessments based on the clinical acceptability (acceptable and unchanged; changed but still acceptable; unacceptable). The most famous of the rating scales for clinical performance was developed by Dr Gunnar Ryge (4, 5), who rated each assessment as alfa, bravo, or charlie (or A, B, C, respectively) for each of several clinical categories of interest [caries resistance, maintenance of anatomical form (resistance to wear), colour match, surface texture, marginal integrity, marginal staining]. This process was called the United States Public Health Service (USPHS) method because Dr Ryge worked for the USPHS at the time when it was first introduced. This list of categories has been extended considerably (modified USPHS) and now also includes occlusal contacts, marginal contacts, resistance to fracture, retention and many others. Collectively, the assessments are used to judge the overall restoration acceptability or failure, as a function of the length of time in service. These will be discussed in detail in a moment.

Clinical performance or clinical outcome factors

Remember the original questions posed by dentists. *Which material should I use? Which is the best? Which will*

last longer? Now, consider the factors that might be involved in the overall performance of any restoration as a function of time. There are number of ways to collect wide range of variables, but a practical way has been introduced (6, 7). There are five categories of variables or factors that describe many influences on clinical outcome. Taken in a logical order, they include operator factors, design factors, material factors, intra-oral location factors and patient factors. Now, let us consider the details associated with each one.

Operator factors deal with potential differences in skill (not judgement) of different dentists. What is the technical ability of an individual to perform a particular procedure? That ability may be influenced by manual dexterity affected by one's natural psychomotor skills and/or the effects of fatigue or ageing (loss of dexterity, changes in manual control, visual impairments). Clearly, there are differences among individuals.

Design factors involve judgements made by the operator in determining the appropriate cavity preparation for the material type being used. Are cavosurface margins created as butt joints or beveled? Are the proximal margins flared and/or beveled? Are dentin seats included at various positions along the pulpal floor? Is the overall design gross retentive or adhesive? All of these impact the transfer of stresses at the restoration-tooth boundaries and margins.

Material factors include all the laboratory properties that were previously discussed. What is the tensile strength? What is the modulus? What is the coefficient of thermal expansion? Are the values sufficient to protect against restoration failure?

Intra-oral location factors consider the intra-oral variations in saliva, stress, temperature, or other effects in relation to the restorative material located in (i) anterior, premolar, or molar tooth, (ii) maxillary versus mandibular arch, or (iii) primary versus permanent tooth.

Patient factors involve environmental effects associated with the patient's behaviour or genetic predispositions. What is the dental IQ of the patient? What is the patient's fluoride exposure history? What is patient's likelihood or risk towards caries? What is the relative saliva production by the patient?

For any single clinical assessment, the list of important factors may be different. Yet, the list is ordered in a remarkably consistent way. The most important factor in terms of affecting risk for clinical failure is operator! Generally, operator risk is considered to be in the

neighbourhood of 50% or more of the overall risk. The least important factor contributing to risk is generally the restorative material. Reflecting on the original questions from dentist, suddenly the reality is that material is not nearly as important as the skill of the operator. A skilled operator can make a poor material work relatively well. An unskilled operator cannot make even the best material work well. This point will be revisited in a moment.

Risk assessment can be performed for any type of factor. Once the impact of the operator can be controlled or understood, the underlying effects in the other categories can be revealed. For example, from a series of well-controlled clinical trials of posterior composite restorations placed by a highly uniform and calibrated group of academic dentists at the University of North Carolina, it was determined that the relative risk for occlusal wear rate of medium-sized bonded composites could be assigned, on the basis of intra-oral location as approximately 100% for first molars, 60% for second molars, 40% for second premolars and 30% for first premolars. Thus, for a skilled operator, one could conclude that posterior composite restorations in first premolars are not at much risk at all and should be selected whenever possible.

Clinical performance assessment in clinical trials

Clinical performance assessment is most frequently performed using the USPHS or modified USPHS categories (5). These are all based on direct observation ('direct analyses') of the conditions in the mouth of the patient. However, it is possible to conduct 'indirect analyses' by taking an impression of the cast of gypsum or epoxy and then conduct laboratory analyses (such as wear measurements) (8) or magnified inspections (such as scanning electron microscopy on the surfaces or margins) (9, 10).

Consider direct analyses in more detail. Remember that these are judgements that were originally selected on the basis of relevance in determining clinical acceptability and were able to be performed with no more than an explorer and mouth mirror. While many operators now use magnification routinely as part of chair-side procedures, that aid was not known when the USPHS system was originally designed. Actual selection of USPHS categories may also depend on the restoration location. Retention may be an issue for adhesive class V preparations but irrelevant for gross

retention of class II restorations which are placed as amalgam replacements.

Caries resistance is interpreted as resistance to secondary caries or recurrent caries in the neighbourhood of the restoration of interest. Many individuals now conclude that there may be no such event as secondary caries (11). They interpret each incidence of caries event as new caries. With that in mind, if one adopts simply the point of view of caries near the restoration, then this measures incidence in the test population of patients. Most clinical trials select patients who are not at high caries risk (12) and thus reports are low for most restorative investigations. Typically, the recurrent caries level in clinical trials is less than 3% (13) [see Baelum *et al.* 2007 for a review (14)].

Tooth-coloured restorations are rated for their ability to match the colour of the tooth into which they are placed. A number of interesting factors contribute to the actual rating. In clinical trials of new materials, there is a high likelihood that the manufacturer will only have a couple of shades (A2, A3) available for the investigation rather than a full palate of 10–14 shades. Therefore, a darker tooth may not have a good colour match at the baseline. It is possible that 10% of the restorations in a clinical trial actually start with a bravo (or B) rating. Over long periods of time both the restoration and tooth can change colour. With most composites manufactured after 1990, little colour degradation is anticipated in the restorative material because of the care taken by manufacturers to add UV stabilizers in the production of the resin. However, the tooth may change colour intrinsically over the time. Dentin tends to become darker and more yellow during middle age (35–50 years). In long-term clinical trials, small changes in restoration towards darkening may be compensated by darkening of the tooth, resulting in a continued good match. Even more likely is the possibility that middle-aged patients' teeth will darken and restorations will appear lighter than originally selected.

Marginal integrity refers to the mechanical durability of the margins. Have the margins fractured or ditched? The appearance of good or poor integrity depends on the position of the margin. Along relatively flat portions of the occlusal table, a ditched margin represents a dramatic change in surface topology. On cuspal inclines, ditches may not be nearly obvious. While an explorer can be helpful in detecting the disrupted interface, detection depends on the orientation of the explorer and the margin path. For composite restora-

tions, it is quite common to encounter remnants of flash at the margins. Thin geometry of flash makes it probable that it will quickly fracture away. This may lead to an impression of a rough or fractured margin, when in fact no important marginal changes have taken place.

Ideally, surface texture of restorations should be smooth. To some extent, this depends on operator techniques and choices for polishing and burnishing procedures. Most restorations start with highly polished surfaces with no detectable surface texture. In earlier times, when amalgams or composites utilized larger-size particles as part of their formulation, it was possible to encounter initial surface roughness, but that no longer is a reality if proper procedures and techniques are employed.

Anatomical form refers to the ability to resist wear. If one only observed the surface of a restoration without reference to the margins, it would be almost impossible to detect changes in contours. The easiest reference for intra-oral wear, particularly along the occlusal table, is to compare the relative height of the restoration to the remaining tooth structure. If the preparation was constructed with butt joints, then this should be a sharp demarcation in most cases and allows detection visually or with an explorer. In almost all cases, the restoration is expected to wear at a markedly greater rate than the enamel at the preparation margin, providing a quasi-reference for the process. However, over many years or where highly wear resistant materials are being used and the process of wear occurs slowly, some wear occurs at the enamel margin making it a poorer reference and may actually hide true changes (15).

Fracture resistance refers to the bulk of the restoration and not simply to the margins. Material from the restoration may or may not be lost. This event is more commonly observed in situations such as large class V restorations where in the cervical third may be fractured and lost, a large class II undergoing cusp fracture with material being lost, or a class IV involving loss of the entire incisal restoration.

Occlusal contact is determined from occlusal markings with carbon paper and reference to the original contacts shown in reference slides at baseline. For small posterior restorations, there typically can be two or three occlusal contacts that may be on the tooth or restoration (16). If they are on the tooth, then changes may not impact the restoration at all. Also, the

likelihood of change under these circumstances would be small. This parameter has more meaning for a larger restoration with initial contacts on the restoration. Other contacts help to distribute stress and prevent any important changes for the sole restoration contact. What is much more important is the determination that this tooth is actually in function and not sheltered by the fact that it has no initial contacts at all.

The quality of a proximal contact is detected using dental floss in most cases. Floss is snapped through the contact and rated as a tight contact (alfa), light contact (bravo), or no contact (charlie). The broadness of the contact is not noted but may influence any long-term changes in the feel of the contact.

Any post-operative sensitivity rating can become a red herring category depending on how the information is used. Most studies pre-select patients without sensitivity and then measure the appearance of any post-operative sensitivity. However, some studies intend to study the effect of materials in reducing pre-operative sensitivity and so this outcome is extremely important. Unfortunately, readers of the results misconstrue the results that low levels of post-operative sensitivity imply that materials are reducing sensitivity when, in fact there may be no relationship between two things at all. When post-operative sensitivity does arise, the normal strategy is to delay treatment from weeks to months to see if things resolve by themselves. In that case, the rating would change from beta-to-alfa in transitioning to the disappearance of post-operative sensitivity. Only in the case when a restoration required replacement would the rating become a charlie.

Loss of retention usually can be observed only in the case of a preparation designed with adhesive-only retention such as for saucer-shaped class V lesions. Part or all of the retention fails and the restoration is discovered as missing during a recall appointment.

One of the challenges for converting all of this information into a clinical judgement is determining the relative importance of each category and creating a method to combine the specific values. For example, one might ask if there are two or three clinical events (e.g. caries resistance, wear resistance, retention) that are more important than others. If one only examined the failure occurring in three categories, would they count all equally or would there be a hierarchy of relative value? When do you replace a restoration? If there is significant wear and the rating is a charlie (clinically unacceptable), does that mean that replace-

ment should occur in a timely manner? It could be that despite the failure, there is little or no risk to the patient and the replacement could be indefinitely deferred. No one has ever determined a method to manage these questions.

Clinical failure analysis

Whether one considers all direct evaluation categories in combination or individually, it is possible to generate a summary of survival versus time as a curve (17). It is more typical to plot success versus time, but call it a failure analysis. An example of the curve is shown in Figure. 1. The curve is reverse-s shaped. Typically, this is termed a longevity curve.

Survival decreases over time from 100% to 0%, theoretically. Actual curves rarely go to zero. A few restorations seem to survive indefinitely. The curve can be reported in terms of the time to failure for half of the restorations in the pool and that is called the clinical longevity for 50% or CL₅₀. Typical values for the CL₅₀ fall into the range of 5–25 years. Points along the curve may be reported in terms of both time and survival rates such as the 5-year survival is 92%.

A failure curve truly represents a wide range of actual results related to different factors. It is just the average performance for a pool of restorations. An important factor influencing the curve is the clinical judgement involved in deciding when to replace a restoration. In controlled clinical trials (CCTs), replacement generally occurs only at the time of a failure. In clinical practice, failure is anticipated and restorations are replaced long

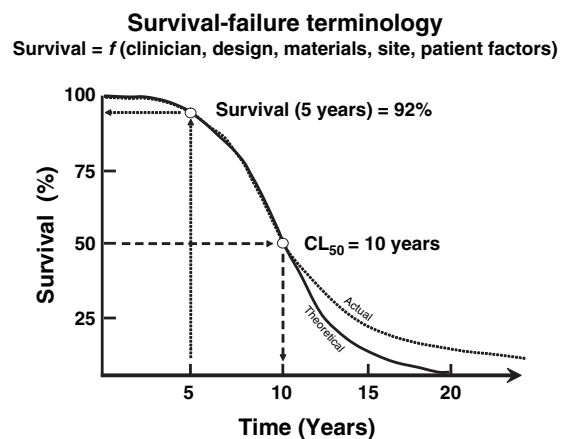


Fig. 1. Survival curve for population of restorations displaying a typical reverse s-shape.

before the actual failure in response to small changes or intuition about potential failure. Therefore, a more appropriate presentation might be something like Fig. 2.

It is also important to recognize that failure in any clinical trial is generally a combination of causes or events. Imagine the situation (Fig. 3) in which early failures are primarily driven by technical problems, midterm events could be caused by dental caries and long-term events might be related more to bulk fracture.

Failure curves for controlled clinical trials (CCTs) do not have the same CL_{50} as the ones observed for materials placed by general practitioners in their normal practice. As mentioned before, the major factor

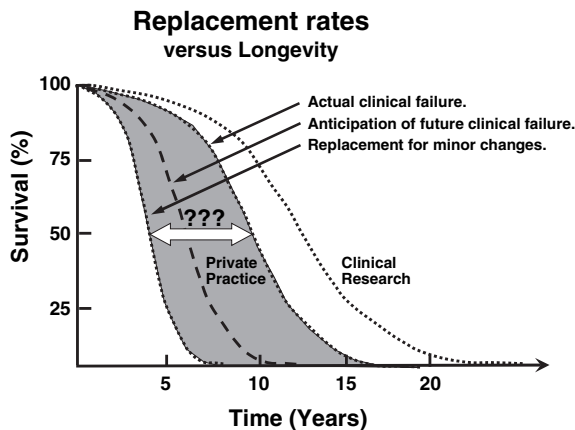


Fig. 2. Failure curves for different replacement reasons, describe a collection or zone of failure curves observed in private practice. The typical clinical longevity at 50% (CL_{50}) is lower in practice than in clinical trials.

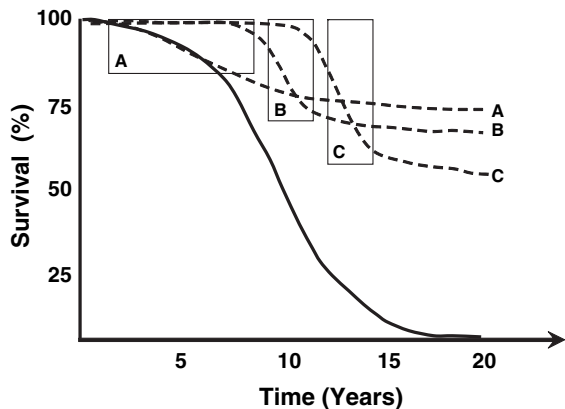


Fig. 3. Theoretical combination of three different failure events (a, b and c) making contributions at different times.

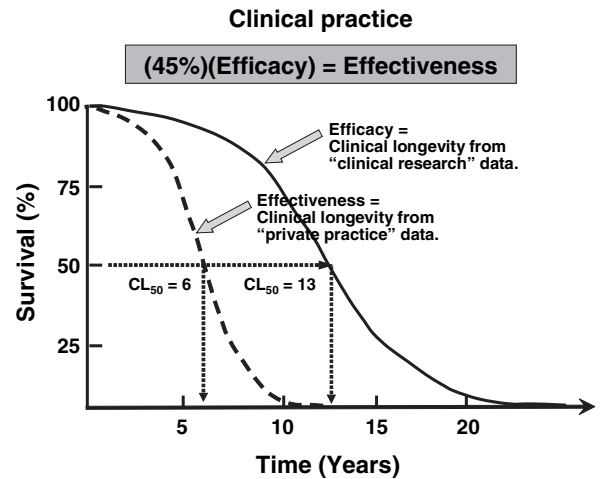


Fig. 4. Clinical longevity at 50% (CL_{50}) for clinical practice versus clinical trials.

influencing the outcomes is the operator. This is reflected in Fig. 4.

Generally, longevity observed for a CCT and private practice are different by a factor of two. Remember that the goals of a clinical trial are to determine the safety and efficacy. The outcome (or performance) for a CCT is called ‘efficacy’, while success or a private practice situation is normally designated as ‘effectiveness’. The ratio of the effectiveness to efficacy is generally around 0.45 as stated above. It is hard to know the impact of uncalibrated evaluators, but it is expected to be a major one (12).

Types of clinical trials

Clinical trials can be categorized in terms of their overall design as: (a1) retrospective = examining existing databases; (a2) cross-sectional = observing data for a fixed period of time without following the progress of any single restoration; (a3) prospective or longitudinal = designing a study and monitoring restoration performance over the time or their length of study as: (b1) short term = 1–5 years or long term = 5–20 years. Clearly, long-term longitudinal clinical trials are the most expensive and least likely to occur. There are only a few examples. Work carried out by Wilder *et al.* (18) is a famous one.

Retrospective clinical research is confounded by the absence of adequate information about the range of factors in play and the actual details of many aspects of the trial. More recent trials are much more likely to

diligently present all the information about a patient pool (distribution of restorations by intra-oral location, premolars versus molars; distribution of patients by age, 20–40 versus 40–60 versus >60 years ; patient gender; relative restoration size, <one-half intercuspal distance versus large restorations; distribution of cavity preparation types, class I versus class II; effects of operators). It is now known that while there are only small differences between class I and class II wear risk, there is much greater risk for restorations wider than one-half the intercuspal distance. Thus, any pooling of retrospective data needs to consider these differences in trying to meaningfully merge data from different trials into a single group for analysis.

Cross-sectional clinical research is relatively easy to accomplish because patients are examined once without need for recalls. A typical cross-sectional trial of class II restorations in patients coming to a clinic might try to measure the number of failures observed in posterior teeth that contained pre-existing amalgams, composites, or glass-ionomer restorations. As this is simply counting observations in an uncontrolled pool, there is no real knowledge of the five categories of factors (operator, design, material, site, patient) that are important for understanding failure. There is no expectation that the groups of restorations will be balanced in number of restorations of different materials, age of patients, or other factors. There is a high probability of concluding improper reasons for failures. The actual number of failures cannot be normalized. As a rule of thumb, it is observed that the clinical longevity reported in these types of studies is usually about half as great (e.g. CL_{50} of 12 years for molar composites in clinical practice) as the one observed in longitudinal clinical trials (e.g. CL_{50} of 20–25 years).

Longitudinal clinical trials provide the best opportunity for conducting well-controlled experiments that are capable of answering discrete questions. In this case, the five categories of factors influencing outcome can be controlled, as needed. The patient pool can be balanced by age, gender, or other important considerations. The primary complaints about information from these trials are that results do not represent the typical outcome in private dental practices where restorative situations might be much more challenging, patients are not ideal, dentists range widely in technical abilities and material usage might be far less uniform. These are certainly true. Yet, any conclusions

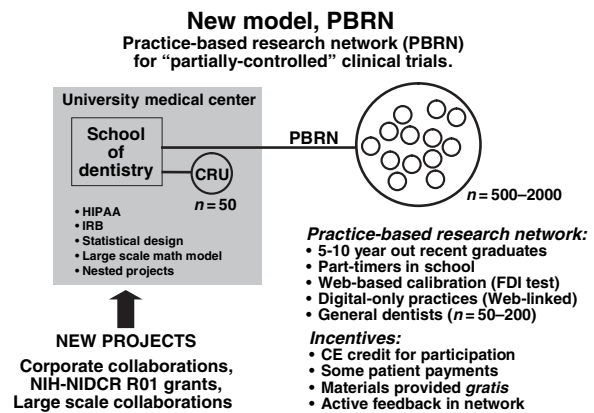


Fig. 5. Practice-based research networks (PBRNs) are a way of combining the information analysis based on the five categories of factors, affecting outcomes that universities normally consider with the greater variability and much larger pool of restorations possible in private practices. PBRNs generally succeed when the practices and practitioners are not too diverse and have reasons for participating in dedicated ways. Projects for PBRNs can be corporate ones or more esoteric ones.

about outcomes would be hard to interpret because the five categories of information under those circumstances would not be well known. To begin to deal with this conundrum, practice-based research networks (PBRNs) have been proposed to collect information in a more orderly manner from clinical practices (Fig. 5).

Both university CCTs and PBRN trials are needed. At the moment, most longitudinal clinical trials are simply short-term ones funded by individual dental companies to evaluate their own products, while looking at a limited pool of restorations. A small pool of restorations (typically $n = 50$) in a limited number of patients (typically $n = 20$ – 25) makes sophisticated statistical analysis almost impossible as the number of factors being considered is too small. The alternative is to use PBRNs to increase the patient pool significantly ($n = 1000$ – 2000), but give up on fully controlling the design for other patient or operator factors. PBRNs are also expensive and generally need research training for PBRN teams involved to control techniques of placement and evaluation of restorations. At the moment, three major PBRN sites at dental institutions (Washington–Oregon, NYU, Alabama–Florida) are being funded by the National Institute of Dental and Craniofacial Research (NIDCR) to explore the benefits from this type of research. Dr Ivar Mjor is one of the next presenters in this

workshop and will discuss the strengths and weaknesses of these operations [see Mjør 2007 (19)].

Prospective clinical trial designs

For any clinical research trial design, there must be a balance between statistical design needs, availability of patient types, levels of financial funding and details of research questions. In some cases, it is totally impractical to ask and answer a particular question. The size and related expense of the trial design can be prohibitive. This is unfortunate but part of the reality in conducting clinical research.

A perfect example is answering the question about the relative efficacy of fluoride releasing restorations. There is no way to directly answer this question. A design would need to take into account the facts that many water supplies are fluorinated, most patients have continuous fluoride exposure from foods they consume, there are impacts of patient behaviours and levels of risk are quite variable among patients. A balanced clinical trial to understand actual dental materials effects might require 10 000–20 000 patients and perhaps 50 000 restorations. A PBRN might have to follow 50 000–100 000 restorations to statistically prove any cause and effect between restoration fluoride-release and reduced caries risk. As an alternative, one can inferentially determine the answer by observing the patterns of failures and conclude that the effect of fluoride-release from dental materials could not be great because the failure rates of those restoration types seems to be high compared to all other things (20).

The ultimate solution is to conduct many small trials that include careful risk-analysis determination. Risk analyses should then be combined into a predictive model. With the input from a new trial's results, the predictive model would provide an estimate of the long-term outcomes and an idea of the failure curve for that particular pool of restorations. For the next 5–10 years, many small trials are needed to establish this baseline.

Correlation of laboratory and clinical testing

Despite all the energy invested into laboratory testing, there are no tests that are truly predictive of long-term clinical performance. There are a couple

of tests that seem to have limited correlation with short-term performance and those will be addressed.

Correlation of laboratory properties with clinical performance measures

What is sought is a series of laboratory tests that predicts the clinical outcomes and parallels the set of clinical performance parameters. One of the traditional methods in doing this is to correlate microleakage tests from thermal cycling with the clinical performance that is measured in terms of resistance to marginal staining or resistance to dental caries. There is absolutely no correlation between these laboratory and clinical events. This has been studied for more than 80 years using a wide range of approaches and combined events with no success at all. A recent symposium at the American Association for Dental Research (AADR) in Orlando (21) concluded that there is no future in continuing these tests. All attempts were fruitless and this was a wasted research effort. Examine for a moment some of the reasons why this absence of correlation occurs.

Clinical events are not nearly as well defined as one might like. *What is marginal staining?* It could be caused by penetration of small molecules associated with foods that leave residual stains. Even a well-bonded margin can have small molecule diffusion along its boundary. If you use iodine as the tracer atom to detect leakage, it can leak directly through the enamel and dentin using available nano-channels. It diffuses more readily along restoration margins through micro-channels associated with the bonded layers or ones related to restorative dentistry damage from cavity preparation. If a larger tracer molecule is selected, it may no longer diffuse through tooth structure, but could still invade micro-spaces. If you wait for too short a period of time, you might not detect any diffusion or may not encounter degradation changes of the materials that could contribute to diffusion opportunities. Not all products degrade at the same rates or with exactly the same mechanisms. If you run microleakage tests and perceive one ranking order for a collection of materials in 1-week testing time, the order could be different for 1 month or 6 months. For each set of laboratory experimental conditions, there is no confirming clinical information that shows that the assumptions about clinical conditions represent the average conditions included in the laboratory test. Finally, and probably

most importantly, the rate of change in a laboratory test is generally an accelerated one. We have no idea of what clinical time period might be represented by the test. Thus, we find that nothing is predictive. Ideally, a laboratory result for a 1-week simulation could predict the relative acceptability of a specific clinical event, occurring over a time period of 3 years.

Criteria for success for clinical and/or laboratory testing

Laboratory tests can produce numerical or categorical results. For the test to have meaning, one must assign an outcome level for the test that is associated with a clinical outcome that is acceptable. Consider a discrete example. One might state that for microleakage testing, if less than 20% of the restorations showed leakage at the dentin–enamel junction, after a 1-year duration using silver-nitrate staining methods, then this would correspond to less than 5% failure rate in 5 years caused by secondary caries for class II restorations in first molars. This connection would state the criteria for success for the laboratory test of interest and its corresponding clinical outcome.

In all cases, we must also consider time-dependent responses. The clinical result considered above (less than 5% secondary caries in 5 years) might convert into 30% secondary caries in 10 years and 75% secondary caries in 15 years. Refer back to the failure curve presented earlier. Early failure levels cannot be linearly extrapolated to predict later failure levels because the failure curve is a reverse s-shape. With this in mind, understand that each particular dental product has its own reverse s-shaped curve. When you test a cluster of related products and rank order the results for a particular time point that does not mean that the same order occurs at a later time. If the results are good, you can probably expect that the reverse s-shaped failure curves are shifted to the right for everything, but the actual order for the individual curves will be unknown.

Quite frequently in publications of laboratory test results, an author infers that the order of results indicates the goodness to badness of the clinical performance. This is often wrong for two important reasons. First, all results could be actually acceptable. Second, the rank order could shift as a function of time. For reporting results, always interpret the goodness of the results by stating the criterion for success. If the criterion was to have less than 20% microleakage in 1

week and all products did that, then they are all successful. You cannot state that one is better than another because there is no evidence in clinical trials or any expectation that the order will occur in that same sequence over all time points.

Suspected correlations of laboratory and clinical performance

In the oral environment, there are always many things occurring simultaneously to a restoration. Mechanical stresses occur cyclically during intercuspation, swallowing, or chewing. Thermal stresses occur cyclically during opening and closing one's mouth. Water and other ions are being actively exchanged with restorations and tooth structure as a function of changing temperatures, moisture contents, saliva flows, or bacterial activity periods. These are just a few of the multi-factorial events. To attempt to understand these, laboratory testing has moved more towards laboratory simulation that includes mechanical, temperature and pH cycling as a quasi-representation of the oral condition. Tests attempt to mimic expected outcomes of clinical trials for approximately 3 years. Unfortunately, there is no agreement at this point about which types of equipment and combination of conditions one should use for testing.

Connect individual properties into an equation for predicting failure

No matter what the status of laboratory testing, the goal should be to move towards a model of risk. There are so many different factors involved that it is difficult for a dentist to absorb information in a clinically useful way other than being able to calculate the risk for failure.

By way of example, consider a second maxillary molar tooth which has lost two cusps and is to be restored with only a limited amount of remaining gross-retention, no pins and amalgam. If the decision for treatment is based on the fact that this tooth is not expected to survive more than a few more years in service and that the patient cannot afford other potentially long-lasting alternatives, then the risk of clinical survival might be 50% for a 3-year outcome. If that is acceptable for that patient and dentist, then this choice for treatment is a good one. If the desire is for 5–10 years of survival, then this is not a good choice. If it were easy to combine the knowledge of the five

factors contributing to survival into a real mathematical equation and quickly calculate the risk, then this would become a useful process for planning treatment and understanding performance.

Evidence-based dentistry and biomaterials research

Introduction to evidence-based dentistry

Evidence-based dentistry (EBD) has long been a goal but only recently has come into vogue as a guiding principle for much of the treatment or future research that is being planned. Ernie Cochrane [after whom the Cochrane Collaboration (22) was named] had been active in questioning the real basis of knowledge that we had for treatment. In the mid-1990s, when EBD moved to the forefront, there was an estimate that only around 8% of all of the dentistry practiced in the world was based on legitimate scientific knowledge. Mostly, it was a best guess (called clinical judgement) or was simply conforming to the standard of practice (what we have always done). The good news is at least 8% of practice includes evidence. The bad news is that we do not really know which part it is.

Shifting the concern to EBD does not mean that evidence magically appears. Information for decision making that would be considered as acceptable evidence actually turns out to be quite complicated. Ideally, evidence from a large number of long-term clinical trials would produce good meta-analyses that would clearly point the way for choices in treatments. However, the knowledge base is almost non-existent. There were only five meta-analyses published for anything concerning restorative dental materials by the year 2004. One of those meta-analyses examined the clinical trials for posterior composite wear (23).

Meta-analyses depend on randomized controlled clinical trials (RCTs) and controlled clinical trials (CCTs) but most of the clinical research in dentistry involves poor CCTs at best. Most are unique enough that it is hard to combine their results into a meaningful meta-analysis.

In lieu of good clinical information, by far, most of the information about restorative dentistry is embedded in a collection of huge numbers of laboratory tests. This is emphasized by the fact that ≥ 600 publications for each popular dental material product has appeared in the literature, in response to excessive testing, during

the last couple of decades (24). Over the years, only about 5–10% of all the information about restorative dental materials has been associated with clinical research (25). Clinical trials are small, short term and limited in terms of products tested.

Laboratory research data are of limited value for the many reasons previously presented. Most data do not correlate with clinical outcomes, as just discussed. Most tests do not strictly follow the same set of specifications or standards and this limits the comparability of the results. Most tests are performed on existing dental materials products, which are poorly characterized in terms of chemical composition and microstructures, so little structure–property understandings are forthcoming.

Another way to determine relative acceptability of products is simply to obtain feedback from clinicians in active practice. *What materials are preferred for handling reasons and associated with the lowest number of clinical problems?* Questionnaire information from clinicians becomes a poor-man's practice-based research network (PBRN). Nothing is truly known about the clinicians or patients. Yet, this type of survey research has become popular and is included in *Clinical Research Associates (CRA) Newsletter*, *The Dental Advisor (TDA)*, *Reality* and the *American Dental Association (ADA) Professional Products Review*.

Individuals involved in laboratory research, clinical research, discussions with clinicians in practice and reports in publications develop impressions about the relative acceptability or desirability of products and provide expert opinions (opinion leaders) to the profession in commentaries, editorials, continuing dental education courses or as consultants. Opinion about the data is easy to collect. It also tends to drive decision making in much of dentistry. It has been stated anecdotally that 85% of all the initial decisions for material selection and use in the United States is based on suggestions from the *CRA Newsletter*. That is not evidence! It has value, but it has no relationship to the long-term clinical performance information that is desired.

A great hoax perpetrated by the desire for EBD is citing stray research articles, as evidence for a particular choice in dental treatment. EBD is not the process of finding a research article that agrees with your opinion or practice patterns. EBD is a process of 'analysing all the existing good literature'. It is entirely parallel to the scientific method as it asks a question, examines the existing literature, determines method for accepting or

excluding the literature based on its relative value, analyses the accepted information, makes a judgement based on that information and the cycle continues as new information is generated. When you consider this loop as an endless cycle, it is easy to see why a conclusion drawn as EBD in 1995 might be considered erroneous once new information has been reported and a re-analysis is conducted in 2005. Perhaps, a better way to define evidence-based dentistry is 'best-evidence-based decision making'.

In 2007, it is still fair to say that far less than 10% of all of the information desired for clinical decision making is really known. *Should one be discouraged by that situation? What is a practical target?* Start by considering the process of EBD collection. If you need around 5–10 years of clinical testing to collect information sufficient for an effective meta-analysis, then one immediately sees that perhaps, the best one can do is provide evidence for practice choices, as they existed around 10 years ago. Information that reflects clinical trials and which is older than 10 or more years would be useful information for EBD determinations. You can infer that about 50% of practice of dentistry could have a strong evidence base. In the meanwhile, the newer techniques, thoughts, or treatments will not have a strong evidence base, yet. These will require time to become fully documented as best choices. Other 50% of the literature would simply be suggestive and not demonstrative.

Therefore, dentistry has a long way to go to reach the 50% of EBD target value, as it has to develop a strong foundation of clinical trial information and extract that information in a manner that can be easily used by individuals in clinical practice. All these are challenges for the future.

The Cochrane Collaboration was initiated in the early 1990s as a standardized approach in analysing the health-care literature to determine what published information could be connected together to answer a question. The process is elegant and thoughtful. The problem is that literature forms a limited data set for the process. If a couple of bad articles exist in the literature but are accepted as the sole source for an EBD review, then the recommendation at the end of the process could be a gross error. This seems to have happened quite often and that is why many of the Cochrane Collaboration reports are strongly disputed. Another outcome that leads to the same situation is that no published materials are accepted for consideration and

no answers arise for important clinical questions. This creates a 'void' for many clinicians – and so, they quickly retreat to the newsletters or opinion leaders for suggestions.

Our dental literature is quite limited in providing excellent publications. As a back-of-the-envelope calculation, consider that (i) less than 20% of all research actually gets submitted for publication, (ii) journals typically have acceptance rates of only 20–50% and (iii) of the published material, Thompson Scientific reports less than 1% of all articles are ever cited. The last statement agrees well with the anecdotal statements of readers, who conclude that only 5% of the literature is good and only 1% is excellent. This calculation translates into a dataset for EBD that includes only $20\% \times 50\% \times 1\% = 0.1\%$ of the evidence being valuable, with most the evidence being only laboratory data.

Analysis of clinical research publication

There is a strong desire for clinical information that is so poorly met by the existing literature, that readers may tend to accept poor clinical research, in spite of obvious faults or major design problems. To see if one has the practical know-how of extracting and arranging the information from the jungle of irrelevant data, consider any example of a recently published clinical research studies. Any example will do. Go to MEDLINE (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) and enter the search string 'clinical AND trial AND dental AND restoration'. Choose any article that is available via Open Access so that you can inspect the PDF file. Quickly scan the article. Now, ask the following questions. *Is the trial of sufficient size, to answer the main research question (objective or purpose) in a statistically meaningful way? Is the population of patients involved described in detail, as far as their gender, age, tooth being restored and restoration size is considered? Is the population of patients or restorations, representative of the average population of interest, for the research question? If multiple clinicians are involved, are they being calibrated for restoration placement and evaluation procedures? Are any conclusions carefully qualified in terms of the limitations of the study? How much do you think it would cost to conduct the study? What would it have cost to conduct the study you might like to have seen done? What would you say about the relative value of the results in the clinical trial, for decision making for patient care?*

Conclusion

Laboratory testing of current biomaterials products constitutes the largest part of information reported in the dental literature. Publications in this regard are not comprehensive in terms of properties or extent of available products. There are no known strong correlations between laboratory testing and short-term or long-term clinical performance. Laboratory testing should always include a criterion for acceptability. Clinical research reports in the dental literature for biomaterials are extremely limited in number and value. At this point, the evidence base for clinical performance is scant.

References

1. Bayne SC, Thompson JY. Biomaterials. In: Roberson TM, ed. *Sturdevant's art and science of operative dentistry*. 5th ed, Chapter 4. St Louis: Mosby; 2006:135–242.
2. Bayne SC, Thompson JY. *Biomaterials science*. Chapel Hill: Brightstar Publishing; (1980, 1988, 1989, 1990, 1991, 1992):586 pages. [Electronic Full-Text: 1995, 2001, 2004].
3. Mjør IA. Minimum requirements for new dental materials. *J Oral Rehabil*. 2007;34:907–912.
4. Ryge G. Clinical evaluation of adhesive restorative materials. In: Austin RH, Wilsdorf HGF, Phillips RW, eds. *Adhesive restorative dental materials II. Proceedings of a workshop held at the University of Virginia*, Charlottesville, Virginia, December 1966. National Institute of Dental Research, Public Health Service Publication No. 14941965; 1966:8–9.
5. Cvar J, Ryge G. Criteria for the clinical evaluation of dental restorative materials. US DHEW Document, US PHS 790244, Printing Office, San Francisco, 1972:1–42 (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.
6. Bayne SC, Heymann HO, Sturdevant JR, Wilder AD, Sluder TB. Contributing co-variables in clinical trials. *Am J Dent*. 1991;4:247–250.
7. Jokstad A, Bayne S, Blunck U, Tyas M, Wilson NH. Quality of dental restorations – FDI Commission Project 2-95. *International Dent J*. 2001;53:117–158.
8. Leinfelder KF, Taylor DF, Barkmeier WW, Goldberg AJ. Quantitative wear measurement of posterior composite resins. *Dent Mater*. 1986;2:198–201.
9. Kusy RP, Leinfelder KF. *In situ* replication techniques I. Preliminary screening and the negative replication technique. *J Dent Res*. 1977;56:925–932.
10. Marshall GW, Marker VA, Bayne SC. Replication technique for monitoring intra-oral processes with the SEM. *Quintessence Int*. 1978;9:1–3.
11. Kidd EA. Diagnosis of secondary caries. *J Dent Educ*. 2001;65:997–1000.
12. Bayne SC, Schmalz G. Reprinting the classic article on USPHS evaluation methods for measuring the clinical research performance of restorative materials. *Clin Oral Investig*. 2005;9:209–214.
13. Bayne SC, Swift EJ, Wilder AD, Sturdevant JR, Heymann HO. Post-operative sensitivity, retention, and sclerosis levels in Class V clinical trials. *J Dent Res*. 2003;82(Spec. Iss. B):0909.
14. Baelum V, van Palenstein Helderma W, Hugoson A, Yee R, Fejerskov O.. A global perspective on changes in the burden of caries and periodontitis: implications for dentistry. *J Oral Rehabil*. 2007;34:872–906.
15. Heymann HO, Bayne SC, Sturdevant JR, Wilder AD, Roberson TM. The clinical performance of CAD/CAM generated ceramic inlays. *J Am Dent Assoc*. 1996;127:1171–1181.
16. Taylor DF, Bayne SC, Sturdevant JR, Wilder AD. Restoration width and complexity effects on posterior composite wear. *J Dent Res*. 1989;68(A):186 (Abstr. 35).
17. Bayne SC. Biomaterials: where are we and where are we going? Proceedings of distinguished professor's symposium. *J Dent Ed*. 2005;69:571–585.
18. Wilder AD, May KN, Bayne SC, Taylor DF, Leinfelder KF. Seventeen-year clinical study of ultraviolet-cured posterior composite class I and II restorations. *J Esthet Dent*. 1999;11:135–142.
19. Mjør IA. Practice based dental research. *J Oral Rehabil*. 2007;34:913–920.
20. Mjør IA. Glass ionomer cement restorations and secondary caries: a preliminary report. *Quintessence Int*. 1996;27:171–174.
21. Dental Materials Group. *Marginal quality of dental restorations*. Sequence 26, Orlando, FL: AADR Annual Session, 2006. [http://iadr.confex.com/iadr/2006Orld/techprogram/session_15371.htm].
22. The Cochrane Collaboration, Website, August 2006. [<http://www.cochrane.org/>].
23. Taylor DF, Bayne SC, Leinfelder KF, Davis S, Koch GG. Pooling of long term clinical wear data for posterior composites. *Am J Dent*. 1994;7:167–174.
24. Bayne SC. Product-testing impact for popular dental materials on biomedical publication levels. *J Dent Res*. 2006;84:1658.
25. Bayne SC, Marshall SJ. 20-year analysis of basic vs applied DMGIADR research abstracts. *J Dent Res*. 1999;78:130 (Abstr. 196).

Correspondence: Stephen C. Bayne, Department of Cariology, Restorative Sciences, and Endodontics, School of Dentistry, University of Michigan, 1011 North, University Avenue, Ann Arbor, MI 48109-1078, USA.
E-mail: sbayne@umich.edu