

Stephen C. Bayne · Gottfried Schmalz

Reprinting the classic article on USPHS evaluation methods for measuring the clinical research performance of restorative materials

Received: 15 July 2005 / Accepted: 4 October 2005 / Published online: 26 November 2005
© Springer-Verlag 2005

Abstract The original article published by Cvar and Ryge in 1971 on the US Public Health Service (USPHS) Guidelines is virtually inaccessible to current scientists, despite its remarkable impact on clinical dental research. The original article described all the pilot studies that led to the choices for the final USPHS guidelines. However, many of the important basic ideas expressed in the original article, such as evaluator calibration, have been overlooked in recent years. Challenges for effective clinical testing of restorative procedures and materials that were emphasized by those authors are even more relevant today. Therefore, it is totally appropriate to republish the original article by Cvar and Ryge in this issue of *Clinical Oral Investigations*. This preface to the republication of the original article provides key background information and references to contributions by the many now-famous clinical investigators who were involved with pilot studies. In addition, the USPHS recommendations are critically reviewed. Clinical evaluation of restorative procedures requires (a) choices of clinically relevant criteria, (b) assessment using simple nominal scales, (c) calibration of evaluators, (d) two independent evaluations, and (e) nonparametric statistic analysis that recognizes the patient (and not the restoration) as the independent variable. Only portions of those procedures are being preserved in current clinical investigations. USPHS criteria continue in use until today as part of routine clinical evaluation and as components of standards programs such as the ADA acceptance program. However, in addition, USPHS-like criteria have been appended over

the years to produce “modified USPHS guidelines.” These additional criteria include parameters such as postoperative sensitivity, fracture, interproximal contact, occlusal contact, and others. The combination of the original and modified USPHS criteria now have been accepted worldwide but are not necessarily uniformly applied. They constitute the foundation for current considerations of further development of clinical assessment methods for dental restorative procedures.

Keywords Clinical evaluation · Dental materials · Evaluator calibration · Restorative procedures · Standards

Justification for reprinting this classic article

Few if any methodological studies in the dental restorative materials field have been cited more often and had a greater scientific impact than the US Public Health Service (USPHS) Guidelines developed by Cvar and Ryge [9]. A simple search of PubMed (August 26, 2005) for (USPHS or Ryge) AND criteria AND (dental OR dentistry) produced 353 references. This paper spelled out the criteria and defined a system for the clinical evaluation of dental restorative materials. This evaluation system was also known as the “Ryge criteria,” of which the original categories were color match, cavosurface marginal discoloration, anatomic form, marginal adaptation, and caries. During the last 40 years, these criteria have been slightly modified by several authors, adjusting them to their special needs, and the list of criteria has been expanded to include other items of interest. The expanded list contains criteria for surface texture, postoperative sensitivity, proximal contact, occlusal contacts, fracture, and others. These modifications are explained and readily accessible in the current dental scientific literature. However, the original research report by Cvar and Ryge is very difficult to access.

There are only three remaining archived copies of this archived publication. New investigators around the world have essentially no access to its content, despite their interest in utilizing the USPHS system. They must rely on

S. C. Bayne (✉)
Department of Cariology, Restorative Sciences,
and Endodontics School of Dentistry, University of Michigan,
1011 North University, Room 2355,
Ann Arbor, MI 48109-1078, USA
e-mail: sbayne@umich.edu
Tel.: +1-734-6479396

G. Schmalz
Department of Conservative Dentistry,
University of Regensburg,
Franz-Josef-Strauß-Allee 11,
93053 Regensburg, Germany

secondary descriptions of this work or restatements of its content. In today's world, there is tremendous emphasis (and increasing governmental financial support) on clinical research. Evidence of this is demonstrated by the extraordinary multimillion-dollar recent expenditures by the National Institutes of Health (NIH) and National Institute of Dental and Craniofacial Research (NIDCR) on investment into and expansion of clinical research systems as part of the NIH Road Map [18].

Within this context and within these special discussions being currently held about improving clinical research techniques, we considered it extremely important to make this original article by Cvar and Ryge available once again to all clinical research investigators. The original research report was released in 1972 as a technical publication by the US Department of Health, Education, and Welfare. It is only appropriate that this research report be republished in a journal that is, by its name, devoted to "clinical oral investigations."

Rereading this 1972 article reveals that the clinical research challenges raised in the middle of the last century clearly are still valid and even more timely in the current climate favoring clinical research some 40 years later. Consider the emphases of the original authors: "Many researchers are acutely aware that clinical performance cannot be directly predicted from laboratory tests," or "In recent years, their task (of the dentist choosing a suitable material) has been complicated by the introduction of dozens of new restorative materials." Today, we can replace the word "dozens" by the word "hundreds". This makes the problem for understanding materials changes and differences even more relevant.

Our feeling is that this article may be regarded as the original basis for the scientific evaluation of the clinical performance of dental restorative materials. Our specific purposes in reprinting this original article are fourfold. First, this process will allow access by all clinical researchers to the original content. Second, it will help emphasize the key value of portions of the original system, such as "training and calibration," which have been all but forgotten. Third, it will stimulate modern discussion about the need for clinical research for everything in dentistry. Fourth, it will be a starting point for redefining the most suitable criteria for testing newer dental restorative materials.

Environment for the development of the USPHS guidelines

A great deal of understanding of the value of this "classic article" comes from appreciating the background and history of this particular research. Until this point in dental history, clinical research in restorative dentistry had not been organized. No one was sure what to measure or how to report their observations. Work toward these ends began for Cvar and Ryge in an earlier paper in August 1964, as noted by the authors in the "Acknowledgments" section of their original article. The environment for the beginnings

of this work was the Materials and Technology Branch, Division of Dental Health, which existed at the USPHS Hospital in San Francisco and which Ryge directed from 1964 to 1971. The first mention of this sort of clinical research effort was revealed in an early conference in 1965 by Ryge [11] and considered the dilemma for investigators. While there was interest in clinical research of restorative dental materials, there was very little history of systematic activity in clinical research. There was great uncertainty as to what categories of information to collect. No system of direct oral evaluation had been developed yet.

As with any new system or methodology or technique, Ryge recognized that this effort required very careful development and testing. In the last part of the 1960s, Ryge was in a unique place to begin to develop this work, the USPHS Hospital in San Francisco. After this groundbreaking effort, Gunnar Ryge moved on to the University of the Pacific and was succeeded at the USPHS (at the Presidio) by Joe Moffa. Moffa adopted and utilized a large part of this system in extensive clinical trials for the next 20 years until he retired and the hospital was closed. During that time, a number of remarkable individuals who would become famous in their own rights worked with Moffa, became calibrated, utilized the system in their own research projects, and perpetuated the effort. In light of Moffa's special role, it is only suitable to listen to him tell his view of this story. (personal communication; e-mail; 2004.09.10)

"It was a typical foggy San Francisco day in 1966 when Dr. Gunnar Ryge, as the newly appointed Director of the Materials and Technology Branch of the U.S. Public Service Health Dental Health Center, challenged a group of young clinical dentists, supportive staff, and a recent graduate from Ralph Phillip's biomaterials graduate program [Dr. Joe Moffa], with the seemingly impossible task of devising a system to quantify the clinical performance of dental restorative materials. [Early discussions included Dr. Bjorn Hedegard and Dr. Bruce E. Johnson.] The original team consisted of Dr. Gunnar Ryge, Dr. James McCune, Dr. Richard Webber, Dr. Rudolph Micik, Dr. Larry Gettleman, Mr. Jack Cvar, Ms. Peggy Benton, [Miss Mildred Snyder,] and Dr. Joseph P. Moffa."

"From the onset, as dental clinicians and from a purely empirical viewpoint, the group had little disagreement as to what constituted either an excellent clinical restoration or a defective restoration, but the group soon realized two major problems. First, the majority of clinical restorations fell between these two extremes and seemed to represent some sort of inexplicable continuous multi-dimensional variable. Second, the approach to the challenge of in-vivo measurement of clinical performance with hampered by the prejudices of prior in-vitro testing of the mechanical and physical properties of dental restorative materials. In the laboratory one could place a standardized specimen

in an Instron Universal Testing machine and obtain discrete numbers which were amenable to conventional parametric statistics. The group soon became aware that the clinical environment was not that amenable to the traditional parametric measurement systems.”

“The group adopted the approach that this challenge was no different from the solution of any other so-called complex ‘insoluble’ problem—identify the elements, break them down, and solve each element individually. Thus, for subjective clinical assessment of restorations which were either excellent or grossly defective, the key was to identify the individual components of that total subjective judgment. After some minor bickering as to appropriate terminology there was unanimity that color match, marginal discoloration, marginal integrity, anatomic form, and dental caries represented the five multi-dimensional parameters which were the major influences on our clinical judgment of a restoration’s success or failure. It would appear that the team’s first success was that it had broken down the subjective clinical judgments to a very basic nominal scale.”

“Having identified the major components of clinical judgment which impact upon a restoration’s clinical performance, the team’s next challenge was to draw upon joint clinical experiences and create a descriptive clinically relevant scale of increased severity within each nominal parameter. In order to reduce the possibility of recorder error, phonetic names were attributed to each of the scale units, i.e. alfa, bravo charlie, delta, etc. [These names are part of the US Air Force system of stating alphabetic letters during radio communications. Alfa is NOT a misspelling of the Greek letter, “α.”] The measurement system which finally evolved implied increased severity of each nominal class and was based upon an ordinal or ranking system.”

“Although Alfa, Bravo and Charlie were easy to pronounce... they were only just names and not numbers. A Bravo restoration wasn’t twice as bad as an Alpha and Charlie wasn’t three times as severe as Alpha restoration. We couldn’t assign a number 1 to Alfa, 2 to a Bravo, and a 3 to a Charlie and calculate means, standard deviations, etc. These were no longer nice neat numbers which could be analyzed parametrically. The group found themselves suddenly in the then unfamiliar world of non-parametric statistics and had to rely on their statistician, Mr. Jack Cvar, to bring light to this brave new world.”

“They realized quite early that in order to use any measuring system in a reliable way, it was important that all prospective clinical evaluators be calibrated systematically. The purpose of the calibration procedure was two fold. First, calibration should eliminate

candidates who lacked the visual acuity, discrimination, and/or familiarity with the scale. Second, calibration should prevent individual drift in judgmental assessment over time.”

“In retrospect, I know I [Dr. Joe Moffa] speak for all the members of the initial team who were challenged by Dr. Gunnar Ryge’s acute perception for the need for a quantitative measure of clinical performance. The team was motivated by his persistence, guidance and overall enthusiasm and were proud to be there when the page was blank. We fully realize that much more is still to be written in the unending quest for measures of clinical performance.”

Overview of the article

This description by Moffa of the events is crucial in understanding the core value of this publication. The USPHS guidelines exist as a “system of clinical evaluation steps” that (a) defines key intraoral events to be measured for any clinical trial, (b) describes or ranks the key clinical stages of change, and (c) provides a calibration system for evaluators who might be involved in clinical trials using the system. The actual article carefully documents all of the stages in the development of the guidelines. As the authors note, “Further experience with the rating scales in actual clinical studies led to the consolidation of anterior and posterior criteria, which had been developed separately, and to the deletion of certain rating scales which failed to yield useful information. The rating scales which were finally adopted are for color match, cavosurface marginal discoloration, anatomic form, marginal adaptation, and caries.” At one point, the authors suggest that the system could have many applications “... including assess[ment of] the work of dental students... [or] comparing two different dental materials or two different dental procedures involving the same patient.” In addition to the information about ranking, the authors offered numerous comments on appropriate methods of devising clinical trials and applying the rating scales.

To make the rating process work, it was and still is crucial to train and continually calibrate examiners. This appears to have been lost as part of the process over many years since the USPHS guidelines were originally published. While some might argue that it is not necessary to calibrate trained clinicians, it has constantly been demonstrated that there is wide variability in the diagnosis of dental problems because of differences in perception and importance among individuals [7]. This uncertainty is clearly a challenge for ratings such as detection of caries. For training, it is therefore necessary that research teams have a set of models or photographs that guide them in the calibration process. Calibration should have a minimum performance expectation such as 85% correct judgment in the calibration phase. Clinical trials should include a declaration of the training and calibration processes, as well as record keeping for those processes.

What has been missing for the most part over the years is a well-defined set of models or photographs for training and calibrating individuals from diverse clinical research teams to arrive at the same level of judgment so that results of clinical trials might truly be comparable. This is a major deficit for current clinical research efforts. Part of this problem might disappear if adequate and similar controls were used routinely in clinical trials. Yet controls rarely exist for so much of the current clinical research. Nagging problems of the expenses have all but eliminated inclusion of controls. One of the great strengths of the USPHS guidelines has been that if investigators are adequately calibrated then controls theoretically might not matter.

During the last 15 years, the American Dental Association (ADA) has continually evolved a series of ADA Acceptance Program Guidelines for clinical trials for such things as bonding systems and posterior composites [1–6]. These guidelines rely on the USPHS categories as the primary information about clinical performance. However, even these ADA guidelines have not defined a requirement for training and calibration.

Another part of the original USPHS guidelines was a requirement for at least dual examination with a process of resolving differences when they arose. Clinical trials during the early days diligently adhered to calibration and dual examination [16]. Costs and inconvenience largely have driven out this process as well. Recalls tend to be done by only one evaluator and perhaps reviewed later by a second using the photographic record of the patient appointment. Using an evaluator who is trained and calibrated produces an 85% likelihood that the rating is correct. Under the original system using two trained evaluators, the likelihood of correct ratings would become at least 97.75% [$85\% + (85\% \times 15\%)$].

Cvar and Ryge included cautions in the “Appendix” to their article “that teeth of any given patient could not be treated as independent of one another” and “that some method must be devised to represent each patient by a single score....” Clinical research trials of recent time have never really dealt with this problem. All restorations are treated as independent. While ADA Guidelines for clinical trials discourage the use of more than two test restorations per patient, there is still strong potential for biasing the results.

Cvar was quick to point out that categories of evaluation included ratings or rankings that should only be analyzed as nonparametric data. There was no presumption that the changes from alfa-to-bravo or bravo-to-charlie could be considered equal in difference. There was no presumption that changes must occur in a single direction (e.g., from alfa-to-bravo-to-charlie), and reverse changes have been reported [17]. In fact, it has not been uncommon for categories such as color matching to go the opposite direction. A bravo could change into a charlie. In the original design, results were reported as percentage of alfa, bravo, and charlie ratings.

Often, the ratings (alfa, bravo, and charlie) have been abbreviated as A, B, and C, respectively. For some categories such as caries, there is no intermediate rating. The patient

either has caries associated with the restoration (C) or does not (A). The original Ryge category of caries used only the ratings of alfa (A) and bravo (B) to designate this, but most investigators have changed this choice to alfa (A) and charlie (C). The latter parallels the meaning of other categories in which charlie (C) means clinically unacceptable.

While it might seem that a charlie (C) rating should dictate immediate replacement of a restoration, that does not happen immediately under many circumstances. Generally, the clinical trial team determines the risk of the failure to the patient. Under some circumstances such as a color change to charlie (C), there is no risk, and so the restoration may not be replaced until the end of the study. This begs another interesting question. Are the USPHS categories all equal in weight or affect in decision-making? The answer is clearly no. Therefore, there is no practical way to pool results across categories or at least none has been shown to be possible to date.

In light of the many possible outcomes, an ADA panel of consultants and advisors developed the ADA Acceptance Program Guidelines for clinical trials. The approach was to choose to define acceptable outcomes as less than a certain percentage of C or charlie scores (e.g., <10% charlie at 3 years) in certain categories only. Other categories might be tracked but might not lead to any decision about overall restoration acceptability. Many investigators have chosen to report the same data alternatively in publications in terms of the percentage of restorations still showing clinical excellent performance (i.e., percent alfa or %A) after various periods of time. Since almost all restorations at baseline start as alfa, the second approach presumably tracks the changes as trends toward failure. There is no right or wrong reporting method, but it would help if current clinical researchers would adopt a more standardized approach.

Variants and modified versions of the USPHS guidelines

Almost as soon as the original articles about USPHS guidelines were published, Ryge set out to consider expanding the number of rating levels and testing the system for applications beyond clinical research per se. Two noteworthy variations arose. One included an additional classification of delta (D) for some of the guidelines [12, 13, 15] A second variation proposed utilizing the same schema for evaluating clinical practice quality [14] but shifting to the ratings of romeo (R), sierra (S), tango (T), and victor (V) [12, 13, 15]. For the most part, these tightly paralleled the A, B, C, and D system.

In the early 1980s, workers at several universities [Indiana University (Ralph W. Phillips, Marjorie Swartz, Jim Setcos), University of North Carolina (Karl F. Leinfelder, Duane F. Taylor, Aldridge D. Wilder, Jr., Harald O. Heymann, Stephen C. Bayne), University of Texas at San Antonio (E. Steven Duke), and others elsewhere (David J. Eick, Mal Jendresen, Armand Lugassy)] began to extend the number of USPHS categories of direct evaluation. Instead of the original five categories (caries, color

match,...), there was interest in parameters such as occlusion, postoperative sensitivity, fracture, retention, and others. More and more clinical trials began to report an expanded list of clinical performance evaluations. For lack of any better title, these modified lists became known as the modified USPHS guidelines. For the most part, the criteria for the original five categories remained exactly the same and those categories were included, but now there were as many as nine new categories as well that mimicked the original rating system in design [alfa (A), bravo (B), and charlie (C)]. Unfortunately, authors from different clinical research teams did not always use the same definitions for assigning these new ratings. Thus, it became almost a requirement to declare the categories and define the ratings as part of all publications. This is still common today.

Early adoption of the USPHS guidelines

The core group of contributors to the original USPHS system, as well as the few dental clinical research centers for investigating restorative dental materials (Presidio, Indiana, University of North Carolina, Oregon, and others) immediately adopted the original USPHS guidelines for their work. There is early evidence of this in the literature. Mal Jendresen and Ralph Phillips probably did the very first official clinical trial based on the new guidelines while studying zinc oxide-eugenol (ZOE) cements that were being modified to develop intermediate restorative materials (IRM) [10]. Many others were exploring new types of materials evaluated with the system [8]. During this same time period, Joe Moffa succeeded Gunnar Ryge at the Presidio in San Francisco and began to significantly expand the clinical research operation. Literally dozens of clinical trials were conducted by Moffa over the next 10–15 years utilizing the USPHS criteria.

While Ryge and Cvar did publish several articles on the system and statistical evaluation of clinical data obtained by the use of the rating scales, few articles were published of any of the comparative pilot studies from the original article. Clinical findings were presented annually at the International Association for Dental Research (IADR) meetings between 1965 and 1970 on spherical, fine grain, and micro amalgam alloy systems and some of the very early composite materials. Spherical alloys, composites, and new cements were being introduced to the market, with little more than laboratory properties being used to promote their acceptance.

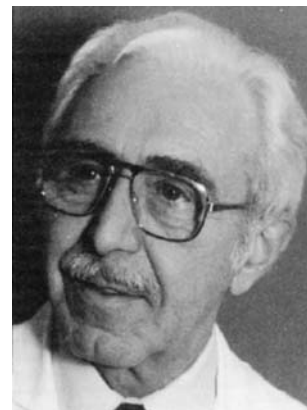
Where are these early pioneers and adopters now?

Jack Cvar was a statistician and not a dentist. He only published six articles that are referenced in PubMed and passed away years ago from being a very heavy smoker.

The rest of the team were young Commissioned Officers with the US Public Health Service and were selected to staff the new branch. Dr. Richard Weaver is now the Associate Director for the Center for Educational Policy and Research at the American Dental Education Association in Washington DC. Dr. Joe Moffa is retired in living in Fair Oaks Ranch in Texas.

Many of the related players who conducted early clinical research trials based on these methods are retired or consultants but some are still involved in research. Here is some information on some of the group. Dr. Mal Jendresen is retired and living in Sausalito, CA. Dr. Larry Gettleman is a professor at the University of Louisville and actively involved in the clinical research of maxillofacial materials. Dr. David J. Eick is the Chair of the Department of Oral Biology at UMKC in Kansas City, MO. Dr. Al Guckes is the Assistant Dean for Predoctoral Education at the University of North Carolina. Dr. Karl Leinfelder is retired as a Professor Emeritus from the University of Alabama and an Adjunct Professor at the University of North Carolina and currently living in Chapel Hill, NC. Dr. Jim Setcos is living in England. Dr. Steven Duke is Director of the Central Texas VA Health Care System in Georgetown, TX. Dr. Jack Mitchem is retired but is still active as an Adjunct Professor at the Oregon Health Science Unit in Portland, OR. Dr. Ralph Phillips passed away in 1991 before retirement from Indiana University Dental School. Ms. Marjorie L. Swartz is retired and living in Indiana. Dr. P.O. Glantz is retired and living in Malmo, Sweden. Dr. Ivar Mjor is a professor in Operative Dentistry at the University of Florida in Gainesville, FL. Dr. Naim Wilson is Dean of the GKT Dental Institute in London.

Dr. Gunnar Ryge (shown in the image below) passed away in 1991 after a brilliant career in dental research that was completed at the University of the Pacific. However, his influence in this process of dental clinical research will be forever honored for the scientific method that he used to devise the USPHS guidelines in his work with Jack Cvar.



Acknowledgements We are sincerely indebted to Dr. Joe Moffa, Dr. Richard Webber, Dr. Mal Jendresen, Dr. Albert Guckes, and Dr. Larry Gettleman for providing background and sharing their own personal experiences in the preparation of this special introduction.

References

1. ADA Council on Scientific Affairs (1989) American Dental Association acceptance program guidelines: composite resins for posterior restorations. American Dental Association, Chicago, pp 1–13
2. ADA Council on Scientific Affairs (1996) American Dental Association acceptance program guidelines: restorative materials. American Dental Association, Chicago, pp 1–14
3. ADA Council on Scientific Affairs (1998) American Dental Association acceptance program guidelines: products for treatment of dentinal hypersensitivity. American Dental Association, Chicago, pp 1–15
4. ADA Council on Scientific Affairs (1998) American Dental Association acceptance program guidelines: home use tooth-whitening products. American Dental Association, Chicago, pp 1–14
5. ADA Council on Scientific Affairs (2001) Revised American Dental Association acceptance program guidelines: dentin and enamel adhesives. American Dental Association, Chicago, pp 1–9
6. ADA Council on Scientific Affairs (2001) American Dental Association acceptance program guidelines: composite resins for posterior restorations. American Dental Association, Chicago, pp 1–12
7. Bader JD, Shugars DA, Rozier G, Lohr KN, Bonito AJ, Nelson JP, Jackman AM (2001) Diagnosis and management of dental caries. *Evid Rep Technol Assess* 36:1–4
8. Cantor R, Webber RL, Stroud L, Ryge G (1969) Methods for evaluating facial prosthetic materials. *J Prosthet Dent* 21: 324–332
9. Cvar J, Ryge G (1971) Criteria for the clinical evaluation of dental restorative materials. US DHEW Document, US PHS 790244, Printing Office, San Francisco, pp 1–42
10. Jendresen MD, Phillips RW (1968) A clinical comparison of four posterior intermediate restorative materials. Technical report SAM-TR Dec, pp 1–10
11. Ryge G (1966) 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 8–9, 1965. National Institute of Dental Research, Public Health Service Publication No 1494
12. Ryge G (1977) Development of clinical testing of materials. In: RG Craig (ed) Dental materials review. University of Michigan Press, Ann Arbor, pp 192–204
13. Ryge G (1980) Clinical criteria. *Int Dent J* 35:347–358
14. Ryge G, DeVincenzi RG (1983) Assessment of the clinical quality of health care. Search for a reliable method. *Eval Health Prof* 6:311–326
15. Ryge G, Snyder M (1973) Evaluating the clinical quality of restorations. *J Am Dent Assoc* 87:369–377
16. Ryge G, Jendresen MD, Glantz PO, Mjor I (1981) Standardization of clinical investigators for studies of restorative materials. *Swed Dent J* 5:235–239
17. Wilder AD, May KN, Bayne SC, Taylor DF, Leinfelder KF (1999) Seventeen-year clinical study of ultraviolet-cured posterior composite Class I and II restorations. *J Esthet Dent* 11:135–142
18. Zerhouni E (2003) Medicine. NIH Roadmap. *Science* 302: 63–72