ACP
AMERICAN COLLEGE OF PROSTHODONTISTS®

A General Method for Describing Sources of Variance in Clinical Trials, Especially Operator Variance, in Order to Improve Transfer of Research Knowledge to Practice

David W. Chambers, EdM, MBA, PhD,[1] Casimir Leknius, BS Pharm, MA, MBA, DDS,[2] & Laura Reid, DDS[3]

[1]Professor, University of the Pacific Arthur A. Dugoni School of Dentistry, San Francisco, CA
[2]Professor Emeritus, University of the Pacific Arthur A. Dugoni School of Dentistry, San Francisco, CA
[3]Assistant Professor, University of the Pacific Arthur A. Dugoni School of Dentistry, San Francisco, CA

Keywords
Generalizability analysis; technology transfer; internal validity; external validity.

Correspondence
Dr. David W. Chambers, Pacific School of Dentistry, 2155 Webster Street, San Francisco, CA 94115. E-mail: dchambers@pacific.edu

This paper is based on presentations at the Invitational Conference on Technology Transfer, Kiawah Island, SC, July 2003, and the International Association for Dental Research, Honolulu, HI, March 2004.

Accepted February 11, 2008

doi: 10.1111/j.1532-849X.2008.00406.x

Abstract

Purpose: The purpose of this study was to demonstrate how the skill level of the operator and the clinical challenge provided by the patient affect the outcomes of clinical research in ways that may have hidden influences on the applicability of that research to practice. Rigorous research designs that control or eliminate operator or patient factors as sources of variance achieve improved statistical significance for study hypotheses. These procedures, however, mask sources of variance that influence the applicability of the conclusions. There are summary data that can be added to reports of clinical trials to permit potential users of the findings to identify the most important sources of variation and to predict the likely outcomes of adopting products and procedures reported in the literature.
Materials and Methods: Provisional crowns were constructed in a laboratory setting in a fully crossed, random-factor model with two levels of material (Treatment), two skill levels of students (Operator), and restorations of two levels of difficulty (Patient). The levels of the Treatment, Operator, and Patient factors used in the study were chosen to ensure that the findings from the study could be transferred to practice settings in a predictable fashion. The provisional crowns were scored independently by two raters using the criteria for technique courses in the school where the research was conducted.
Results: The Operator variable accounted for 38% of the variance, followed by Treatment-by-Operator interaction (17%), Treatment (17%), and other factors and their combinations in smaller amounts. Regression equations were calculated for each Treatment material that can be used to predict outcomes in various potential transfer applications. It was found that classical analyses for differences between materials (the Treatment variable) would yield inconsistent results under various sampling systems within the parameters of the study.
Conclusions: Operator and Treatment-by-Operator interactions appear to be significant and previously underrecognized sources of variance. It is suggested that variance estimates of factors thought to significantly influence the transfer of research findings to practice contexts and evidence of representative sampling across practice contexts be regularly included in reports of clinical trials.

Reports of a product's or procedure's good performance in a research context increase the expectation of similar good performance in various practice contexts, but significant research findings do not necessarily guarantee good clinical outcomes, nor do they explain what factors contribute to variations observed in practice contexts. To the extent that circumstances differ between product testing and product use, published results may be only generally accurate estimates of average practice outcomes. This is known as the issue of external validity in research design[1] or the technology transfer problem.[2] In this paper, we will use the term "research context" to designate the various factors that influence outcomes in a study reported in the literature and the term "practice context" to refer to those factors that exist in various clinical settings where the findings might be applied. Because the factors that represent potential sources of variation in outcomes may differ across these contexts, there will be a "transfer problem." A typical example is the case where research on a new material is performed under very tightly controlled circumstances and with carefully selected patients (leading to a very narrow range of variation in the reported outcomes and desirable statistical significance), but practitioners, following the research protocol as closely as possible, experience a wider range of variance in their outcomes. Significant advances have been made in research design, statistical rigor, and publication standards. These represent improvements in internal validity and are necessary but not sufficient for creating a scientific foundation for practice.

A research claim can be highly significant, in the statistical sense, even though it accounts for a small proportion of the variation in outcomes. Variation in outcomes means the scatter or distribution of outcomes and is reported in clinical research as standard deviation (technically, variance is the squared standard deviation). Large variance is a problem in practice because of the annoyance caused by unpredictability and because large variation increases the chances of failure (regardless of average values). Practitioners could use information on both the significance of hypotheses and the magnitude of importance of sources of variance. When reading the literature, practitioners review inclusion and exclusion criteria for patients in journal articles to gauge the approximate similarity between the research sample and their own practice contexts. But such descriptions do not support any estimate of the extent to which differences among patients was either a small or a large source of variation in the outcomes reported. It would be helpful to practitioners if the research literature reported findings in clinical terms (in the same millimeter or bond strength units of practice rather than *p*-values or correlation coefficients). It would also be helpful to know the likely variation in outcomes for the practice context with one treatment, one operator, and one patient.

In classical approaches to clinical research, variation due to sources other than the products and procedures being studied is regarded as "error," and efforts are taken to minimize it. This may have the unfortunate consequence of distorting or covering information that potentially affects outcomes in practice contexts that are not the same as the context where the research was conducted. A single skilled operator/evaluator is often used in clinical trials, thus making "operator effects" a hidden variable. Similarly, patient selection criteria may mask differences across this dimension, making expected outcomes in practice better or worse or more or less predictable in range than those reported in the literature. This issue is especially important when there are patterns of interaction between an investigational treatment and the operator or factors associated with different patients. If there is an interaction effect (e.g., when Treatment A works best for Patient Type X and Treatment B works best for Patient Type Y), there is danger of reporting superior performance for one or the other treatment depending on which patients are sampled or of reporting "no difference" on average across all types of patients because the differential advantages cancel each other out. In many cases there is insufficient information given in published studies to permit sound expectations about how the studied variables will perform when transferred to the practice context.

Strictly speaking, research findings are valid only for contexts that are the same or "substantially similar." A case can be made that sampling in clinical trials should reflect potential practice contexts rather than power considerations in the research context. The value of clinical research would also be enhanced if the major sources of variance that affect the primary results of the study could be reported as guides to assist practitioners interested in applying the findings. Although it is not possible to identify all sources of variance or to express them in terms that exactly match all practice contexts, a first approximation could be made using the major categories of variation due to the treatment, the operator, and the patient.

Although there have been studies reporting operator variance,[3–9] this effect is normally treated as a special class of error. It is not regarded as valuable in "explaining" outcomes, and its interaction with treatment effects is seldom reported because of insufficient statistical methods. Differences among patients are more apt to be recognized as meaningful sources of variance in outcomes, but such differences are seldom reported systematically within a single study.

The purpose of this study was to demonstrate how the skill level of the operator and the clinical challenge provided by the patient affect the outcomes of clinical research in ways that may have hidden influences on the applicability of that research to practice.

Materials and methods

This study was conducted in a laboratory setting on mannequins, and it involved two provisional impression materials, two levels of operator skill, and two teeth of differing "ease of treatment." To maintain consistency in this paper, the three sources of variance explored will be designated Treatment (referring to the interim prosthesis material used), Operator (referring to the ability level of students who performed the restorations), and Patient (referring to the type of tooth on which the interim prosthesis was fabricated). It is not intended that the designation "Treatment" include all characteristics of all possible treatments or that the measured Operator or Patient factors capture all variation of this type; however, since the total variance in outcomes always equals 100%, any variance that is not completely captured by an effective operationalization of the sources in question will appear in the "error" variance instead.

The Treatment source of variance was represented by two provisional impression materials in common use. Students who fabricated the provisional crowns had previous experience with both materials and would use both in the clinical years of their program. Because this was a study of the effects of variability in technology generally and not of specific products, the materials were identified as T_X and T_Y. Their composition and handling properties are summarized in Table 1.

Two commercially available prepared mannequin teeth, both examples of tooth #8, were selected to represent the Patient source of variance. This variance was created by preparing one with working space and definition of chamfer that would be easy to restore (P_E) and one that would be difficult (P_D). These are shown in Figure 1.

The Operator variable was represented by dental students of high and low laboratory skill ability—31 students who had

Table 1 Working characteristics of the Treatment factor (two materials for fabricating temporary crowns) used in study of transfer from the research context to the practice context

	Material X	Material Y
Advantages	Good recovery memory	Longer working time
	Automixed ingredients resulting in greater material consistency	Can add or subtract material
	Standardized working and setting times	
	Very little shrinkage	
	Sets on bench	
Disadvantages	Brittle, will break under pressure	Variability in consistency because of individual mixing
	Difficult to add or subtract material	High shrinkage, requiring accommodations
		Sets on tooth
		Poor recovery memory

completed their preclinical restorative technique courses but had not yet entered clinic completed the project. From this set, 25 were selected prior to further analysis based on their grade-point averages (GPAs) in preclinical technique courses so as to form a group of high-skill-ability students (O_H) and a group of low-ability-level students (O_L) by removing 6 students with average GPAs.

Each student prepared eight provisional crowns in a balanced order (Latin squares design with random starting points). These included two crowns each with a combination of Treatment and Patient. The research design represented a three-factorial (T, O, P), fully crossed, random-effects model. The design was fully crossed in the sense that each value of Treatment was studied with each value of Operator and each value of Patient, each value of Operator was measured with each value of the other factors, etc. The design was random in the sense that the levels of each factor were selected "randomly," or more technically, selected to represent the general range of available Treatment, Operator, and Patient factors found in the practice context. This was not a study to demonstrate that a particular product has superior performance characteristics (which would be a fixed-effects model). Each provisional crown was evaluated by two of the authors using the criteria employed in the Preclinical Fixed Prosthodontics Technique course at the school where the research was performed. An eight-point scale was used, with values of 2 representing minimally clinically acceptable interim prostheses, and 4 representing serviceable or "typical" quality.

Classical statistical analysis, regression analysis, and generalizability analysis[10,11] were performed. Many factors can cause an observed score (outcome level of performance) to be high or low. The technique known as generalizability analysis can be used to estimate the amount of variance from each source or combination of sources. This procedure was developed in the 1950s and has been used most extensively in the field of psychometrics (testing and measurement). It goes beyond traditional ANOVA methods by identifying all measured sources of variance (including interactions). The results of this analysis are independent of the effects of sample size and are expressed in the same units that practitioners would use (millimeters, degrees, etc.).

Generalizability analysis can be useful as a preliminary step in classical studies. If it is determined that a factor is of no theoretical interest and does not interact with other factors under consideration, cases can be collapsed across that factor. In the present study, there were no interactions between either first and second replication by students or the two scorers and the three primary factors of Treatment, Operator, and Patient. Consequently, averages were taken on replications and raters to provide more stable estimates rather than treating these as separate and meaningful sources of variation.

Results

The analysis of results will be presented in four sections: (1) classical reporting of results, (2) demonstration that the sampling is representative, (3) estimation of variance components, and (4) estimation of level of performance and predictability for transfer to various practice contexts.

Classical analysis

It is standard in the dental literature where products and procedures are investigated to focus on the differences (variance) between the products or procedures and to regard factors such as operator and patient as sources of error. In the classical analysis, the test of significance is derived by dividing the variance attributed to treatment by variance attributed to other sources. In the present study, the F-value for a one-way ANOVA for Treatment is 11.060 (df = 1.98), statistically significant at $p < 0.001$. Material X is better than Material Y, but only in a general way and only in the context of typodont teeth prepared by 25 particular dental students as described above. We have not

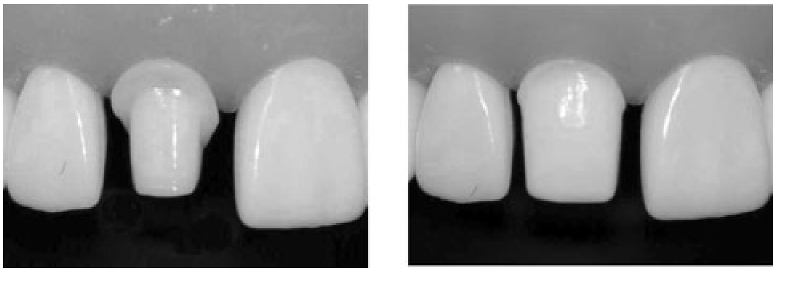

Figure 1 Provisional crown preparations—conventional preparation on left (P_E), conservative preparation on right (P_H).

Table 2 Confirmation of representative sampling of variables used in study of transfer from research to practice contexts

	Mean (SD)
Treatment	
Material X	12.1 (0.04)
Material Y	69.3 (0.19)
Operator	
High-skill students	11.7 (0.07)
Low-skill students	82.5 (0.07)
Patient	
Easy preparation	14.7 (0.06)
Hard preparation	72.0 (0.16)

Mean ratings for material and teeth by 12 faculty members with reference to potential practice contexts: 1 = best practical value, 100 = worst practical value in use. GPA of students converted to same scale.

yet addressed the transfer question of how these materials can be expected to perform in other contexts.

Sampling validity

For a research study to be of clinical value, there must be a correspondence between the research context and the practice context to which it will be transferred. In the present study, the authors chose Treatment products they believed would produce performance levels near the top and near the bottom of the range likely to be encountered among products in common use in practices. The Patient teeth were fabricated to similarly represent a difficulty level near the top and near the bottom of what would be encountered in practice. Confirmation of this sampling as representative of a realistic range in practice was made by asking 12 faculty members who teach in courses where provisional crowns are taught to rate the materials and the teeth. The results are shown in Table 2 and graphed in Figure 2. The third factor, Operator ability, was sampled by selectively reducing the available dataset so that the technique-course GPAs of student operators would match the distributions on the Treatment and Patient factors.

The three factors studied in this research have similar distributions: normally distributed around two points approximating the interquintile range of the factors as they would be expected to occur in the population of applications in practice. In other words, differences in proportion of variance attributed to each source are not likely to be an artifact of sampling differences.

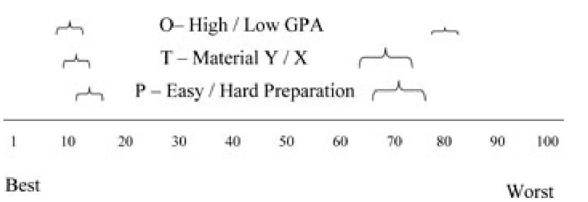

Figure 2 Graphic display of values of Treatment, Operator, and Patient variables in a study relative to their distribution in the transfer domain. Treatment and Patient variables rated by 12 faculty members, and the Operator variable being student grade-point average, converted to the same scale used in rating technology and situation; brackets representing standard deviations.

Estimations of variance components

Table 3 contains the estimates for variance components for the three sources measured in this study and their interactions. The estimated variance components are expressed in the units of the scoring scale used. These are also shown as percentages of the total variance. The remaining column in this table reports "unexplained" variance components attributed to various sources.

The data in this table show that the major source of variance, the factor contributing most to differences between high and low scores on the fabricated provisional crowns, was Operator (student technical skill level), with 38%. Other large sources of variance were the interaction between Treatment and Operator (17%) and the type of material used, Treatment (17%). Differences involving tooth morphology, Patient, contributed less to differences in judged serviceability of the provisional crowns. The unexplained variance is composed of equal parts three-way interaction and "residual error" (analysis not shown). Seventy-six percent of the variance in provisional crown quality in such a laboratory setting is explained by considering only the three factors of Technology, Operator, and Patient, and their interactions.

Figure 3 is a Venn diagram representation of the relative proportions of variance attributed to Technology, Operator, and Patient, and their interactions in fabrication of provisional crowns in a laboratory setting. The sizes of the circles and their segments are proportional to the variance each source represents of the total, 100%.

Table 3 Estimated variance components, proportions of variance, and error variance for quality of provisional restorations in transfer study with Treatment (T), Operator (O), and Patient (P) as factors

	Est σ^2	% Var	Est (Δ)
T	0.516	17.1	
O	1.148	38.1	1.148
P	0.126	4.2	0.126
T × O	0.498	16.5	0.498
T × P	0.011	0.4	0.011
O × P	0.133	4.4	0.133
TOP, e	0.580	19.3	0.580
Σ Est σ^2	3.012		
Est (Δ)			2.496
Est CI_{95}			3.097

Est σ^2 are the estimates of variance for each component, corrected for sample size, in the research study. % Var are the same values expressed as proportions of the total variance. Est (Δ) are the estimates of rating scale "error" variance for each component in a standardized practice context with either technology, one randomly selected operator, and one randomly selected tooth. Each component of "error" variance is expressed in units on the rating scale used, and they are additive. There is no "error" variance component for T, because that is the variable of interest in such a study.

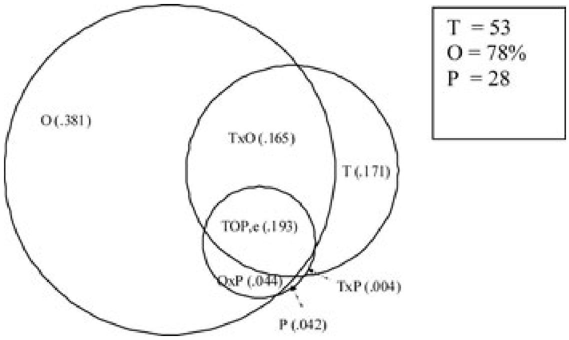

Figure 3 Graphic display of three components of variance in technology transfer model based on proportion of variance in outcome attributed to each source (T = Treatment, O = Operator, P = Patient).

Figure 4 graphically displays some of the interaction effects found in this study. Interactions appear in such graphs as lines that are not parallel. Of significance are the Treatment-by-Operator interaction and the Treatment-by-Patient interaction. In both cases, Treatment X is generally superior; however, had the sampling involved only students of lower ability or the more difficult teeth, the differences between the materials would not have reached the conventional $p < 0.05$ level of significance. A minor three-way cross-over interaction was found, indicating a situation where preference for Treatment would differ depending on the combination of Operator skill and difficulty level of the Patient for which the provisional crown needs to be fabricated.

Estimating level of performance and predictability

In estimating performance in practice from the clinical literature, dentists can typically do no better than anchoring their predictions in the reported means and standard deviations and then making general and subjective allowances for the degree they consider the circumstances in the study might resemble their own circumstances. Estimates of level of performance might be too high or too low; range of results (predictability) might be larger in the research context or larger in the practice context. If sampling in the research study is representative of potential practice contexts and the variance can be estimated, it is possible to combine this information to give improved estimates of performance and predictability for transfer from research to practice contexts.

This is accomplished by means of traditional regression analysis. We calculate a projected quality outcome score for Treatment X using the data available regarding the effects of Operator skill and Patient difficulty when using this material. We also calculate a separate projected quality score for Treatment Y. Separate regression equations are necessary because an Operator cannot use both materials at the same time; the Treatment-by-Operator interaction source of variance is missing for the same reason. In this case, the calculated regression equations for the two materials studied are

$$\text{TreatmentX} : \text{Outcome} = 2.409 + 0.841\,\text{O} - 0.780\,\text{P}$$
$$\text{TreatmentY} : \text{Outcome} = -0.570 + 0.758\,\text{O} - 0.180\,\text{P}.$$

It is apparent from inspection of these equations that, other things being roughly equal, Treatment X is expected to produce better results (higher intercept term). It is also apparent that this advantage for Treatment X is improved in the hands of better Operators and eroded as the difficulty of the preparation increases (greater negative coefficient for the P term). To project performance in a particular practice context, values unique to the practice context are substituted for the variables O and P in these equations. The use of regression equations to project expected outcomes in a practice context is meaningful only

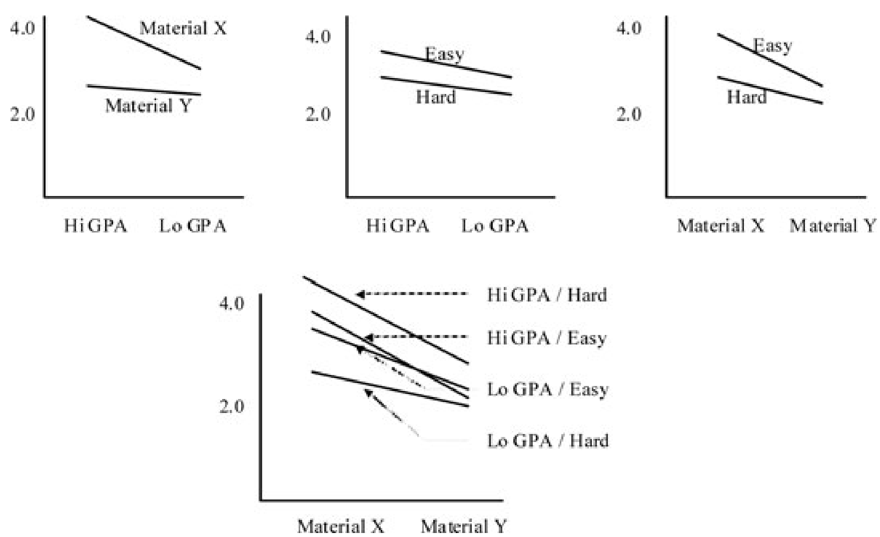

Figure 4 Graphs of two- and three-way interactions involving Treatment, Operator, and Patient in study of transfer. The vertical axis is quality rating of temporary crown, where 2 = minimally acceptable for clinical use and 4 = average serviceability; the two Treatments are labeled "Material X" and "Material Y;" "Hard" and "Easy" refer to tooth preparation, Patient; "High GPA" and "Low GPA" refer to the Operator source of variance.

when the sources of variance in the research context have been sampled to represent the same relative variance expected in the practice context.

Estimating level of predictability is accomplished directly from the data in Table 3, the estimated variance components. All the sources of variance (weighted by number of replications) are added, except for the sources of interest (Treatment in this case). This represents the expected variation in outcomes when using the chosen Treatment. For example, a single Operator fabricating a provisional crown on a single tooth, each selected randomly from a population characterized as was done in the study, would have a variance of 2.496. Multiplying this variance by 1.96 gives the traditional 95% confidence interval, a range in which 95 of each 100 samples would likely contain the true average performance level.

Discussion

This paper has sketched a general approach to bridging the gap of transferring knowledge generated through research into practice. Representative sampling and generalizability analysis to estimate components of variance provide estimates of expected levels of performance and predictability in various individual practice contexts.

It was observed that the three factors of Treatment, Operator, and Patient captured more than three-quarters of the variance in this study. In particular, differences between Operators and interactions between Treatment and Operator were important sources of variance. These variables are seldom studied in dentistry and are seldom measured and reported separately in studies of dental materials or procedures. Consequently, the profession has missed an opportunity to better understand potentially powerful approaches to improved performance. The fact that the combination of only three factors and their interactions accounted for 76% of the variation in this study suggests that further investigations of this type have potential for explaining practical differences in product and process results.

Applying new Treatments in clinical situations where research conditions are not representative of the potential practice contexts is liable to be misleading, and misleading without warning. Standards for management of randomization are described in the CONSORT statement (http://www.consort-statement.org). The CONSORT statement and conventional practice limit the management of randomization, however, to undifferentiated, unmeasured elements that enter statistical analysis as residual error.[12] There is no corresponding standard for random selection of factors known to affect variation in performance and predictability. Traditional research designs "control" conditions to both favor the hoped-for outcome and minimize error variance. While contributing to internal validity, such practices cloud the prospects for meaningful estimates of the transfer value of research results.

Two specific cases of the problem of nonrandom sampling of sources of variances will be mentioned here. First is the problem of "hidden factors." When product or process research is performed by a single operator, there is no report of Operator variance. It is likely that most published research overestimates the performance level of products and procedures, because the skill level of the Operator and care taken are at a higher level than would normally occur in a practice context. The current research suggests that this is an important source of variance. Further, published research almost certainly overestimates the predictability of product and process use in applied settings by failing to measure Operator variance. Failure to measure significant sources of variance will not normally detract from the statistical significance of findings in the research context, but it will compromise the accuracy of estimates regarding use of the research in practice.

The second problem involves "fixed effects." The current study assumed a random-effects model. Treatment, Operator, and Patient were chosen in the study to be representative of potential practice contexts. When effects are fixed (as in a study of specific products or procedures) different statistical analyses are required. Without taking this distinction into account, statistical analysis overestimates the variance attributed to product or process and underestimates the variance attributed to other sources.[11]

In the current report, it is assumed that the practice context of interest would involve only one Treatment, performed by one Operator, in one Patient. The generalizability analysis in this study made necessary adjustments for the multiple Treatments, Operators, and Patients estimated in the research context. Classical research does not make such an adjustment, with the attendant consequence that estimates of predictability (standard deviation in the practice context) are usually optimistic.

The estimates of various sources of variance produced by generalizability analysis are not dependent on sample size. This means that these estimates can be used to project expected outcomes in various settings, by reintroducing the sample size numbers for each source of variance unique to the *particular practice context*. A manager who is interested in estimating the average and variance of a clinic with 500 patients per month and four dentists could use features of the generalizability approach (not described here[11]) for that circumstance. For the one Treatment, one Operator, one Patient case, the 95% confidence intervals on outcome are projected to be ± 3.097 (from Table 3); the 95% confidence intervals for the hypothetical clinical case with four Operators and 500 Patients is projected to be ± 0.314.

The generalizability technique is potentially of value to researchers in designing efficient experiments. Based on variance estimates from pilot studies, a researcher may be able to determine whether it is better to use 1 Operator and 40 Patients or 4 Operators and 10 Patients. In the research reported in this study it would appear that the multiple Operator research design is superior, because the Operator source of variance is appreciably larger than the Patient source of variance.

It is currently an accepted standard in reporting the results of clinical trials that inclusion and exclusion criteria be described for subjects in the study. It could also prove valuable to adopt the practice of routinely reporting the various components for Patients as well. The data supporting this type of analysis already exist in clinical trials; it would only be necessary to perform and report the generalizability analysis. Making a general practice of reporting the variance associated with Operators would be similarly valuable. This would only be possible, however, where multiple Operators are used. The results of this paper suggest that both regularly using multiple Operators and regularly performing tests to estimate variance attributed to this source and reporting it would be beneficial. Designing research that makes it possible to identify and analyze interaction effects is more difficult, but it should be done until it becomes clear that interaction effects can be safely ignored as a significant source of variation for various types of research. An additional standard that may improve the quality of published reports of clinical trials is to include, in the "Materials and methods" section, information supporting the extent to which the sampling of cases in the research context is representative of the population in the practice context.

It is assumed that performance and predictability are cardinal characteristics that determine practice innovations, although Chambers[13] has shown that cost, speed, and ease of use may be stronger determinants of product adoption. Chambers[14] has demonstrated a simple, general approximation method for estimating transfer parameters based on the published literature. The potential has also been demonstrated that two Treatments may be equivalent in both mean and variance while one would be preferred because it is more robust, less subject to Operator and other interactions. This is likely to happen where one procedure or product has a large variance, since likelihood of failure is a function of both average and variance. There is a small body of literature in medicine[15,16] showing that physicians may prefer "fuzzy" to precise findings in supporting their adoption of emerging technologies.

This research may also have implications for evidence-based dentistry, as the potential has been demonstrated that Operator and Treatment-by-Operator variability can be major sources of variance in dental outcomes. For example, a procedure or product demonstrated to be effective on average in a set of studies may still not be the Treatment of choice in some Patients or for some Operators. Research in the social sciences would support the expectation that Operator and Treatment-by-Operator interactions will be larger sources of variation than will Treatment alone.[17,18] This would be a concern where Patients and Operators are not explicitly studied as factors.

Conclusions

A laboratory experiment involving fabrication of provisional crowns was performed with three measured sources of variance (Treatment, Operator, Patient) sampled to standardize relative variance and to represent a balanced range of differences in potential applications. Regression and generalizability analysis were used to generate estimates of expected performance level and predictability in potential transfer situations for the two materials tested (Treatment). Generalizability analysis revealed that, in this case, Operator variance accounted for 38% of the variance, while Treatment-by-Operator and Treatment variance each accounted for another 17% of the variance.

Research studies should be designed to measure potentially significant sources of variance. Regularly reporting proportions of variance attributed to significant sources of variance and regression equations relating these sources of variance to characteristics likely to be found in practice would improve the accuracy of clinicians' estimates of level of performance and predictability when applying the results of research to practice. Confidence in the applicability of reported research findings would further be improved by inclusion of information about the extent to which the research sample is representative of the population in practice contexts to which the findings of the research are to be transferred.

References

1. Campbell DT, Stanley JC: Experimental and quasi-experimental designs for research on teaching. InGage NL (ed): Handbook of Research on Teaching. Chicago, IL, Rand McNally, 1963, pp. 171-246
2. Chambers DW: Some contributions to the challenge of measuring quality of technology. Int J Tech Trans Commerc 2003;2:384-398
3. Bouillaguet S, Degrange M, Cattani M, et al: Bonding to dentin achieved by general practitioners. Schweiz Monatsschr Zahnmed 2002;12:1006-1011
4. Jacobsen T, Soderhold KJ, Yang M, et al: Effect of composition and complexity of dentin-bonding agents on operator variability—analysis of gap formation using confocal microscopy. Eur J Oral Sci 2003;111:523-528
5. Finger WJ, Balkenhol M: Practitioner variability effects on dentin bonding with an acetone-based one-bottle adhesive. J Adhes Dent 1999;1:311-314
6. Miyazaki M, Onose H, Moore BK: Effect of operator variability on dentin bond strength of two-step bonding systems. Am J Dent 2000;13:101-104
7. Re D, Augusti D, Spreafico D, et al: Effect of operator variability on dentin adhesion [abstract]. J Dent Res 2004;120
8. Sano H, Kanemura MF, Inai N, et al: Effect of operator variability on dentin adhesion: students vs. dentists. Dent Mater J 1998;17:51-58
9. Shortall AC, Harrington E, Patel HB, et al: A pilot investigation of operator variability during intra-oral light curing. Br Dent J 2002;193:276-280
10. Brenner RL: Elements of Generalizability Theory. Iowa City, IA, ACT Publishing, 1983
11. Cronbach LJ, Glezer GC, Nanda H, et al: The Dependability of Behavioral Measures: Theory of Generalizability for Scores and Profiles. New York, Wiley, 1972
12. Jokstad A, Esposito M, Coulthard P, et al: The reporting of randomized controlled trials in prosthodontics. Int J Prosthodont 2002;15:230-242
13. Chambers DW: Lessons from badly behaved technology transfer to a professional context. Intl J Tech Trans Commerc 2005;4:63-78
14. Chambers DW. Snowballs in hell. J Am Col Dent 2002;69:3-5
15. Christensen-Szalanski JJ, Bushyhead JB: Physicians' misunderstanding of normal findings. Med Dec Making 1983;3:169-175
16. Colombotos J: Information transfer in dentistry: dilemmas and contradictions. J Am Col Dent 1989;56:54-58
17. Kortez D, Stenches B, Klein S, et al: The Vermont portfolio assessment program. Educ Meas Issues Prac 1994;13:5-18
18. Shavelson RJ, Baxter GP, Gao X: Sampling variability of performance assessments. J Educ Meas 1993;30:215-232