A Meta-Analytic Approach to Determine the State of the Science on Implant Dentistry

Howard M. Proskin, PhD[1]/Robert L. Jeffcoat, PhD[2]/Annette Catlin[3]/
Jillian Campbell[4]/Marjorie K. Jeffcoat, DMD[5]

Meta-analyses have been widely used to estimate treatment effects in evidence-based dentistry. Few conferences, however, have used a systems approach to assure coherent data management and analysis. The purpose of this section is to describe the data management and statistical analysis for the State of the Science of Implant Dentistry (SSID) conference. This overview includes (a) *a description of widely used clinical trial designs for implant dentistry,* (b) *a description of data management procedures specifically designed for the SSID conference, and* (c) *a description of the data analysis methodology, including descriptive analyses and meta-analyses. The use of the systems approach facilitated data abstraction and analysis.* Int J Oral Maxillofac Implants 2007;22(suppl):11–18

DATA SOURCES AND METHODS

A conference on the State of the Science on Implant Dentistry (SSID) was undertaken to answer a small but significant group of questions concerning clinical success of endosseous root-form dental implants. After more than 3 decades of research and clinical experience, a substantial literature exists in this general subject area. It is sometimes difficult, however, to draw valid (ie, substantiated) conclusions from a body of experience with so much variability as to factors such as materials, methods, and population.

The structure of the 8 SSID questions (Table 1) deserves brief comment. Each deals with a single aspect of implant dentistry and can be answered by means of a single, carefully specified outcome measure. Each question was formulated with the following structure: "What, if any, effect does some treatment choice *x* have on outcome *S*, all other factors being equal?" By posing the questions in this way, attention was focused on first-order effects of 8 treatment options on implant success. This is a practical rather than a fundamental, limitation. The same methods could be applied directly to answer other first-order questions and, with only minor adjustments, to address the possibility of higher-order interactions (eg, "What is the combined effect of *x* and *y* on *S*?").

Table 1 The 8 Structured Questions Addressed by the SSID

1. What is the effect on outcomes of time-to-loading of a fixed or removable prosthesis placed on implant(s)?
2. Which hard tissue augmentation techniques are the most successful in furnishing bony support for implant placement?
3. In patients requiring single-tooth replacement, what are the outcomes of implant- as compared to tooth-supported restorations?
4. For teeth requiring endodontic treatment, what are the differences in outcomes of restored endodontically treated teeth compared to implant-supported restorations?
5. Does the type of implant prosthesis affect outcomes for the completely edentulous arch?
6. Does the type of implant prosthesis affect outcomes in the partially edentulous arch?
7. How do smoking, diabetes, and periodontal disease affect outcomes of implant treatment?
8. How does the timing of implant placement after extraction affect outcomes?

[1]President, Howard M. Proskin & Associates, Rochester, New York.
[2]Consultant, Vine Brook Research Corporation, Bryn Mawr, Pennsylvania.
[3]Data Analyst, School of Dental Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.
[4]Statistician, Howard M. Proskin & Associates, Rochester, New York.
[5]Dean, School of Dental Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Correspondence to: Dr Marjorie Jeffcoat, Morton Amsterdam Dean, University of Pennsylvania School of Dental Medicine, 240 S. 40th Street, Philadelphia, PA 19104. Fax: +215 573 4075. E-mail: jeffcoat@dental.upenn.edu

Table 2 Primary Outcomes and Treatment Factors by Question

Question	Outcomes				Treatment factors		
	Implant/ tooth survival	Implant/ tooth success	Prosthetic success	Graft success	Time to loading	Type of prosthesis	Grafting technique
1 - Effect of loading	x	x			x	x	
2 - Hard tissue augmentation	x	x		x	x	x	x
3 - Implant- vs tooth-supported (single tooth)	x	x	x			x	
4 - Endodontic treatment vs implant	x	x				x	
5 - Implant in completely edentulous arch	x	x	x			x	
6 - Implant in partially edentulous arch	x	x	x			x	
7 - Effect of systemic factors	x	x					
8 - Time from extraction to implant placement	x	x	x		x	x	

Although it would be quite simple to design *de novo* a study to answer any of the 8 questions, by assigning patients to 2 or more treatment groups (eg, x_1, x_2), recording the outcome y_i for each group, and using the appropriate statistical technique to compare the groups, the situation is more complicated when conclusions are to be drawn from past, published research studies. To cite just a few complications, investigators have used implants of various designs, enrolled patients according to varying criteria, made measurements of various kinds, and seen patients at variable recall intervals. The meta-analysis techniques employed for the SSID workshop permit such heterogeneous data sets to be combined in a meaningful and statistically defensible manner.

Source Review

The SSID meta-analysis was based on more than 1,300 articles* published in peer-reviewed journals. Responsibility for each of the 8 SSID questions was assigned to a reviewer and a co-reviewer with expertise in the respective area. The first task of these 2-person teams was to search the literature for potentially applicable publications. This done, they read and reviewed the articles for relevance and quality. A quality score was assigned to each study based on its design, which could be 1 of the following:

- **Randomized placebo-controlled double-blind clinical trial (highest):** Patients are randomly assigned to receive test or control therapy. Outcomes are assessed by clinicians without knowledge of the therapy (blinded). While such studies represent the gold standard, for example, in drug research, there are no such studies in the implant literature due to the impossibility of blinding both the practitioner and the patient.
- **Randomized controlled clinical trial:** Such trials are used when it is impossible to blind the patient to the study treatment.
- **Prospective study in sequential cases:** In prospective studies, individual patients are followed over time. The best of these studies use sequentially treated patients and account for subjects lost to follow-up. Prospective studies may employ no control group (an observational study), a concurrent control, or historical controls.
- **Cohort or retrospective study:** Cohort or retrospective studies look back in time at the outcome of patient therapy. The best of these studies use sequentially treated patients and account for subjects lost to follow-up. Retrospective studies are like prospective studies in that they may use no control, a concurrent control, or historical controls.
- **Case report (lowest):** A case report provides a narrative description of 1 case or a few cases sharing points of clinical interest. While case reports are often beneficial to clinicians and sometimes provide the earliest indication of important phenomena that deserve fuller study, they were not considered in the SSID analysis due to their lack of quantitative outcomes.

Outcomes

In order to perform a meta-analysis, the data extracted must be consistent. Table 2 is a matrix showing, for each question, the major outcomes assessed and treatment options considered. Further detail regarding the individual data sets may be found in each article.

*In the present article, these source documents are referred to as "articles," and the scientific works on which they report are referred to as "studies."

Implant survival and success were defined broadly to include as many manuscripts as possible. The defining characteristics for each category were:

- Implant survival
  - •Implant in the mouth and functioning
  - •No mobility (when measurement of this was possible)
  - •No pain
  - •No infection
- Implant success
  - •Implant n the mouth and functioning
  - •Bone loss less than 50%
  - •No pain
  - •No infection
  - •No mobility (when measurement of this was possible)

Abstraction

For every study that met the quality criteria, the reviewers were asked to provide

1. Publication data (title, authors, journal, date, volume, pages)
2. Type of study (see "Source Review" section)
3. Number of patients
4. Definition and size of treatment groups
5. Number of implants placed
6. Time of each observation (months after placement)
7. Condition of implants and prostheses at each observation

In addition, question-specific data were requested, such as auxiliary procedures performed (eg, endodontics, bone grafting) and type of prosthesis (eg, fixed, removable). Reviewers were free to record other relevant data and/or narrative comments in addition to this basic information.

It was permissible, and not uncommon, for an article to be abstracted by 2 or more teams. This happened, for example, when a study addressed more than 1 of the SSID questions, or when 1 of its treatment arms could be used as a historical control against which to compare the results of other studies. Such duplications were flagged and double-checked for validity as part of the analysis procedures.

DATA COLLECTION METHODS

Reviewer Data Entry

Each reviewer was asked to record and report findings using a spreadsheet-based data template designed and provided by the study coordinator. This approach had several desirable features:

- Data entered by the reviewer were transmitted directly and automatically to the analytical database, without the possibility of transcription error.
- Because data-entry forms were tailored to the specific requirements of each research question, the burden of completion was minimized, as was the possibility of confusion and misinterpretation.
- A modest degree of self-checking (validation) further reduced the likelihood of entry errors.
- Implementation on a commonly available spreadsheet platform (Microsoft Excel) allowed reviewers to accomplish their work using a variety of hardware.
- A stand-alone approach provided far more flexibility to the reviewers as to when, where, and how they carried out their assignments than an online, network-based architecture would have.
- Data could be checked, revised, and recast as often as desired prior to final submission.
- Forms were designed to permit free-form entry of any comments or qualifications that the reviewers felt necessary to supplement the basic quantitative data.

Figure 1 is a representative data entry screen for Question 1. The user could only type in cells with white backgrounds; the rest of the form (eg, titles, labels) was protected to prevent accidental entries that might corrupt the data. The workbook[†] was organized into multiple identical worksheets or "tabs," one per manuscript reviewed. Notice that several of the column heads in Fig 1 are specific to Question 1; the workbooks provided to reviewers for other questions were similar in appearance, but slight variations were made to accommodate the peculiarities of each research question.

A number of built-in Excel features (eg, cell protection, data validation) were utilized to make the worksheets easy and reliable to use. In addition, custom-embedded software, invisible to the user, automated such higher-level tasks as generating additional fresh worksheets and pre-entering group and time-period data based on the study being reviewed. Although the worksheets could have been even more strongly protected against modification, reviewers were given the option of expanding the prescribed structure if they thought it appropriate or necessary to capture

[†]An MS Excel workbook includes 1 or more worksheets and charts plus any shared programs, forms, images, and menus that accompany them. Individual sheets are accessible in most versions by selecting a labeled "tab." For this project, all the articles cited in response to 1 question were collected in a single workbook, 1 article per worksheet.

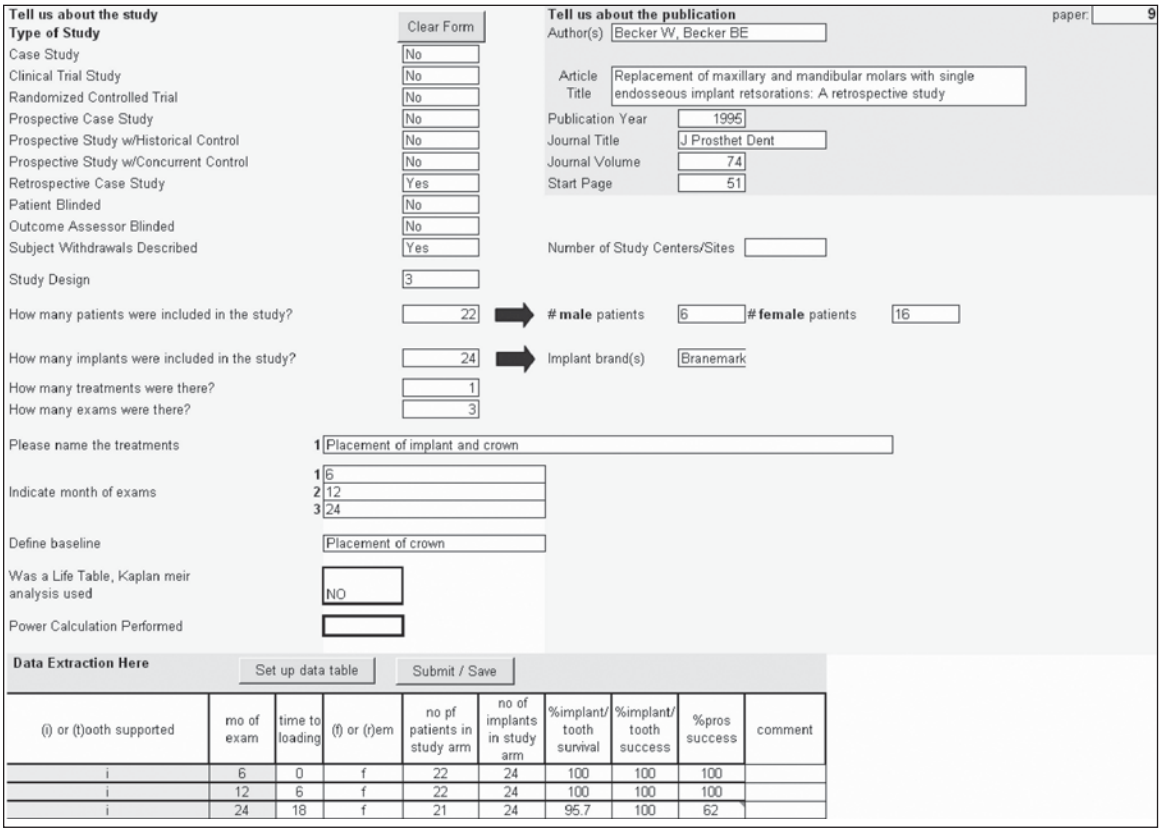

Fig 1 Sample data abstraction screen.

subtleties of particular articles. Some made extensive use of this flexibility, and as long as they did not omit or relocate the required fields, the downstream data extraction steps were not affected.

Data Validity Review

When a workbook was returned by a reviewer, it was scrutinized by an analyst to verify its formal (as distinct from scientific) validity. Checks included:

- Was the critical data structure intact? Were all key data elements in the proper location within the worksheets?
- Were data entries of the correct type? Were they consistent both with study standards (eg, time reported in months) and with other available information on the manuscript (eg, summary statistics)?
- Was there embedded evidence of errors in interpretation or transcription (eg, increasing survival over time, abnormally high or low success rates)?

Questions were referred to the reviewer for resolution. Following corrections, the data for each question were reformatted, reprinted, and sent out for final checks by the respective reviewer.

Data Extraction

The data were extracted from the Excel workbooks using SAS Release 9.1 (SAS Institute, Cary, NC). The SAS dynamic data exchange (DDE) feature was used to acquire data directly from the Excel workbooks. For each workbook (which generally contained information on many studies), 4 SAS data sets were created:

- **Article descriptions:** This data set included author, publication year, manuscript title, journal, total number of implants in the study, and other study descriptor variables.
- **Examination timepoints:** This data set included the number of examinations and the timepoints in the study (expressed in months) at which those examinations took place.
- **Treatment descriptions:** This data set included a list of conditions (termed "treatments") that were evaluated and compared in each study.
- **Extraction:** This data set included the outcome data for each treatment for each timepoint, including information such as the implant survival rate, the implant success rate, and the prosthetic success rate (as relevant for each research question).

Data Management

The examination timepoint and extraction data sets were merged to ensure that all examination months mentioned in the study description were represented in the extraction data set. The treatment description data set was similarly used to ensure that all treatments in the study were represented in the extraction data. Any discrepancies were sent to the reviewer for resolution.

For each study treatment, it was assumed that the cumulative implant survival rates, implant success rates, and prosthetic success rates recorded in the extraction data set represented the actual respective rates among all implants (or prostheses) for that treatment in the study.

In many of the Excel sheets the number of implants (or prostheses) for each study treatment was not explicitly given. When this occurred, the estimates of those numbers were obtained using the following imputation process:

- If there was only 1 treatment in a study, the total number of implants (or prostheses) for that treatment was imputed as the number presented in the study description. If the study description did not provide this information, the total number of implants (or prostheses) for that treatment was imputed as the number recorded at the earliest examination.
- If there were 2 or more treatments in a study, the total number of implants (or prostheses) for each treatment was imputed as the number recorded for each treatment at the earliest examination. If this information was not recorded, the total number of implants (or prostheses) for each treatment was imputed as the total number of implants (or prostheses) presented in the study description divided by the number of treatments, rounded down to the nearest whole number.
- If a study related to endodontics and the number of prostheses was not recorded anywhere in the Excel file used for data extraction, then the number of prostheses was assumed to be equal to the number of patients in the relevant treatment group.

All imputed data were entered into an Excel file and, after verification, were converted into an SAS data set using DBMS/Copy version 8.0.

Data Preparation

All questions that arose based on the SAS data sets described were resolved through collaboration with the data quality analyst. They were then merged into a master SAS data set for summarization and analysis.

Table 3 Definitions of Examination Timepoints

Value	Range included
Month 6	4 – 8 months
Month 12	12 – 18 months
Month 24	24 – 30 months
Month 36	36 – 42 months
Month 48	48 – 54 months
Month 60	60 – 66 months
Month 72	72 – 78 months
Month 120	120 – 126 months
Month 180	180 – 186 months

Table 4 Quality Categories

Category	Defining property
Fair	Retrospective study
Average	Prospective case study
Good	Prospective with historical controls
Better	Prospective with concurrent controls
Best	Double-blind randomized controlled trial
Unknown	None of the above

Table 5 Study Size Categories

Category	Defining property
Small	30 or fewer implants (or prostheses)
Medium	31 to 100 implants (or prostheses)
Large	101 to 1,000 implants (or prostheses)
Very large	More than 1,000 implants (or prostheses)

In the master data set, a variable was created that assigned an examination timepoint category to each examination in each study. Assignment was based on the month recorded for each examination, rounded to the nearest integer. In those instances where the examination had been reported in terms of a time range, the midpoint of that range was used. The possible values for this variable, and the defining range for each value, are shown in Table 3.

In addition to the timepoint category, the master data set recorded which examination was the last examination reported for each study. The SSID reviewer for each research question had the option of imposing an earlier boundary on these Last Examination designations. The Last Examination variable was left blank if it fell outside of the boundary.

The master data set also included variables categorizing the quality and size of each study. Quality categories were assigned as shown in Table 4, and study size was categorized as shown in Table 5.

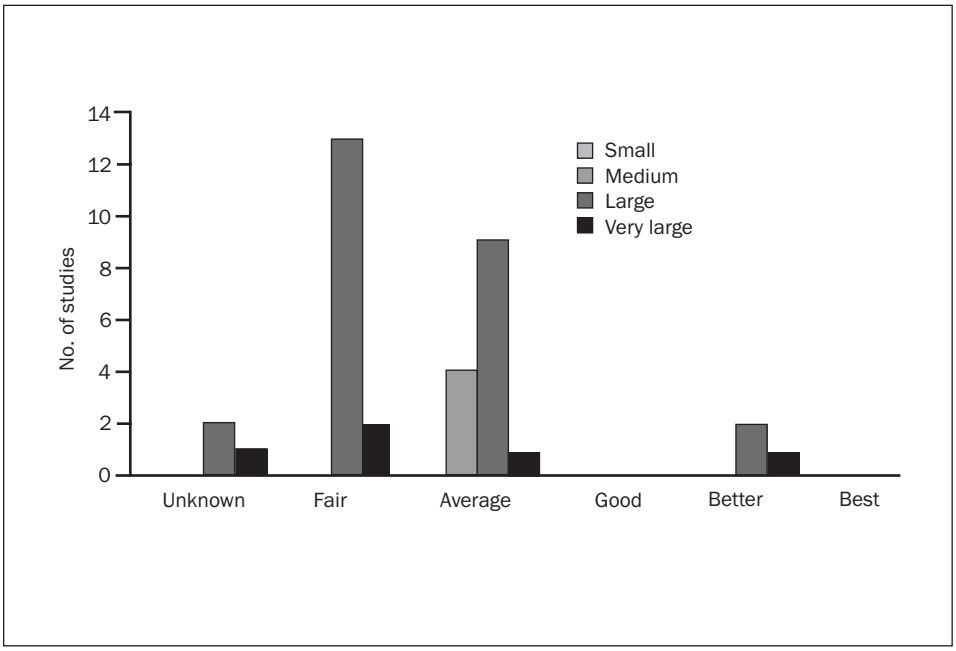

Fig 2 Graph showing the quantity and quality of the studies used in the analyses.

Once constructed, the master data set was audited to ensure that the information it contained was complete and accurate. This auditing process involved both hand-checking and the use of programmed edit checks using SAS. Upon successful completion of this audit, the master data set was saved both as an SAS data set and as a tab-delimited ASCII file. The latter was created using the export procedure in SAS.

DATA ANALYSIS

All analyses were performed using the master data sets described. Separate analyses were performed for each research question.

SAS was employed to produce bar charts showing the distribution of the size and quality of the studies included in the review and a distribution of the publication dates of the included articles.

The ASCII version of the master data set was read into R (version 2.2.1, R Foundation for Statistical Computing, Vienna, Austria) and was analyzed using a meta-analysis function library created by Howard M. Proskin and Associates for this purpose.[‡]

The SSID reviewer for each research question was responsible for deciding which examination time-point categories were relevant to that particular question. Given this information, a separate meta-analysis was performed for the studies in each of those categories. In addition, a meta-analysis was also performed on the designated last examination data for all studies included in the review.

Statistical Methods

For each meta-analysis performed, point estimates of implant survival, implant success, and/or prosthetic success from each study were used to construct a 95% confidence interval for each proportion. The Wilson score method was used for this purpose.[3,4]

For each meta-analysis, it was necessary to decide whether to employ a fixed-effect or a random-effects meta-analytic model. A random-effects model is appropriate when a pooled estimate of effect is to be derived from a collection of studies for which heterogeneity of effects is indicated. In the absence of apparent heterogeneity, a fixed-effect model is considered appropriate. For each analysis, preliminary chi-square tests were performed in order to investigate the homogeneity of reported success rates across the studies reported.

[‡]The methods employed in the meta-analyses performed for the SSID workshop were similar to those reported in previous publications.[1,2]

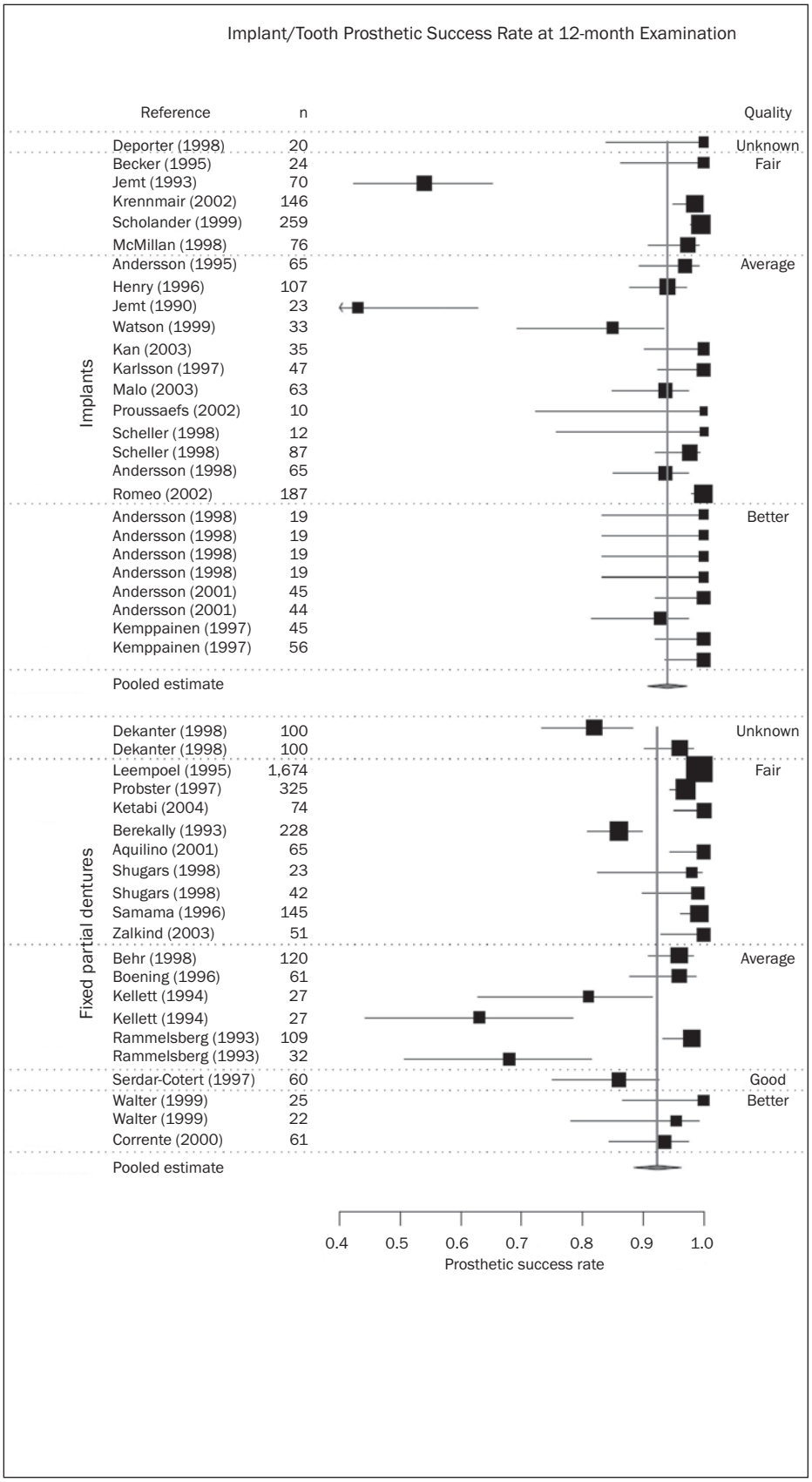

Fig 3 Forest plot showing the detailed question-specific results for each of the studies used in the meta-analysis, together with pooled estimates derived from that analysis.

In the SSID review, significant heterogeneity was frequently indicated among studies for every review question, so it was decided to employ a random-effects meta-analytic model throughout. The exceptions to this were cases where either *(a)* only a single study was to be employed in a particular overview or *(b)* all of the success rates for studies in the particular overview were exactly the same.

Generalized estimating equations were employed to obtain pooled success rates in those meta-analyses performed based on the random-effects model. This method accounts for between-study variability and was shown in a simulation study to perform better than other approaches used for pooling proportions in this setting.[1,2] For those instances where a fixed-effect meta-analytic model was employed, pooled Wilson score confidence intervals were employed to estimate the overall success rates.

Although comparisons between study treatments are suggested by the meta-analytic overviews obtained for each of those treatments, this issue is ideally addressed by obtaining an overview from only those studies in which *both* of the treatments were reported. Unfortunately, such data were not available in most of the articles employed in the overview, since the comparisons relevant to the SSID research questions were rarely the original focus of those publications. However, in some instances, a direct comparison of treatments based on data within individual articles was possible, and where this occurred, additional meta-analytic overviews were performed pooling these comparisons across studies. Such analyses investigated the risk difference between the treatments, obtained by subtracting the reported survival (or success) rate for 1 treatment from that for the other. The pooled estimate and variance for the risk difference were based on the DerSimonian-Laird random effects model,[5,6] which yielded *P* values and confidence intervals for the pooled risk difference based on a normal approximation.‡

Graphic and Tabular Display

To the extent possible, a common format was used to display the results of the analyses performed for each of the 8 SSID research questions. The materials included:

- A graph demonstrating the quantity and quality of the studies employed in the analyses. Figure 2 is an example of such a display. The x axis represents the quality of a study according to the categories described previously, and the y axis represents the number of implants reported for each study. The height of each vertical column represents the number of studies for each combination of quality and size.
- A forest plot showing the detailed, question-specific results for each of the studies employed in the meta-analysis as well as the pooled estimates derived from that analysis. Figure 3 is an illustration of such a display. For each individual study, this plot presents the observed success rate (the horizontal position of the center of the square) and the 95% confidence interval constructed for that success rate (the endpoints of the horizontal line drawn through the square). Both reference the scale at the bottom of the plot. The size of the square is proportional to the number of units (eg, implants, prostheses) included in the analysis.

The pooled estimates derived from the meta-analysis are shown at the bottom of each plot by diamond-shaped symbols, where the pooled estimate of the success rate is indicated by the horizontal position of the diamond and the vertical gray line. The 95% confidence interval for the pooled success rate is indicated by the width of the diamond. Confidence intervals are truncated at 100%.

In each forest plot the studies are grouped in ascending order by quality category. Apart from this use in the display, quality scores were not used in the statistical procedures reported here.

CONCLUSION

The SSID meta-analysis was facilitated by clearly posed research questions, a strongly structured data-reporting mechanism, ongoing involvement of domain experts, and rigorous statistical methods tailored to the specific nature of the questions. Despite a paucity of controlled trials, and substantial heterogeneity among the published studies, the procedure resulted in valid and potentially useful answers to questions of importance to implant dentistry.

REFERENCES

1. Zhou XH, Brizendine EJ, Pritz MB. Methods for combining rates from several studies. Stat Med 1999;18:557–566.
2. Suri RS, Clark WF, Barrowman NB, et al. Diabetes during diarrhea-associated hemolytic uremic syndrome. Diabetes Care 2005;28:2556–2562.
3. Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. Statistics in Medicine 1998;17:857–872.
4. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Statistical Science 2001;16:101–133.
5. DerSimion R, Laird N. Meta-analysis in clinical trials. Controlled clinical trials 1986;7:177–188.
6. Bohning D, Malazahn U, Dietz E, Schlattmann P. Some general points in estimating heterogeneity variance with the DerSimonian-Laird estimator. Biostatistics 2002;3:445–457.