

An Evaluation of the Quantity and Quality of Empirical Research in Three Pastoral Care and Counseling Journals, 1990-1999: Has Anything Changed?*

Kevin J. Flannelly, Ph.D.
Associate Director of Research
The HealthCare Chaplaincy
307 E. 60th Street
New York, NY 10022

Clarence Liu, M.Div.
Chaplain for Pacific Health Ministry
Director of Bereavement Services
Hospice Hawaii
Honolulu, Hawaii

Julia E. Oppenheimer, B.A.
Research Assistant
The HealthCare Chaplaincy
307 E. 60th Street
New York, NY 10022

Andrew J. Weaver, M.Th., Ph.D.
Director of Research
The HealthCare Chaplaincy
307 E. 60th Street
New York, NY 10022

David B. Larson, M.D., M.S.P.H.
President
International Center for the Integration of Health and Spirituality
Rockville, MD

This article summarizes a review of all articles published in *Pastoral Psychology*, *The Journal of Religion and Health*, and *The Journal of Pastoral Care* between 1990 and 1999, identifying a total of 737 scholarly articles, of which 165 (22.4%) were research studies. The proportion of research studies, especially quantitative studies, increased significantly between the first and second half of the study period ($p < .05$). There was a significant positive correlation between compliance with three out of four criteria of internal validity. Three of five criteria of external validity were also positively related to one another. Compared to previous research using identical criteria to assess quantitative studies in the same journals in 1980-1989, the 1990-1999 sample showed improved compliance with respect to specifying the sampling method ($p < .001$), reporting the response rate ($p < .05$), and discussing the limitations of research studies ($p < .001$). However, the overall findings suggest that many researchers in the field do not have a sophisticated knowledge of statistical sampling, statistical analysis, or research design. Several recommendations for increasing the quality of quantitative research are offered.

The publication of research findings is an essential part of any clinical profession, and the quality of such research has been considered an indicator of the development of a discipline.^{1,2,3} As in other profes-

The Journal of Pastoral Care & Counseling, Summer 2003, Vol. 57, No. 2

The HealthCare Chaplaincy is deeply grateful to the Fannie E. Ripple Foundation for its generous funding that made this research possible.

*John Gartner, David B. Larson and Carole D. Vachar-Mayberry, "A Systematic Review of the Quantity and Quality of Empirical Research Published in Four Pastoral Counseling Journals: 1975-1984," *The Journal of Pastoral Care*, 1990, Vol. 44, No. 2, pp. 115-123.

*Thomas St. James O'Connor, Fred Konig, Elizabeth Meakes, Kelly McLarnon-Sinclair, Katherine Davis and Vanessa Loy, "Quantity and Rigor of Qualitative Research in Four Pastoral Counseling Journals," *The Journal of Pastoral Care*, 2001, Vol. 55, No. 3, pp. 271-280.

*Andrew J. Weaver, "Has there been a Failure to Prepare and Support Parish-Based Clergy in Their Role as Frontline Community Mental Health Workers: A Review," *The Journal of Pastoral Care*, 1995, Vol. 49, No. 2, pp.128-147.

sions, original research in the field of pastoral care is necessary to inform professional practitioners and guide clinical practice.^{4,5,6}

A 1991 survey of members at all levels of the American Association of Pastoral Counselors (AAPC) found that approximately two-thirds of those who responded considered research in pastoral care to be either "important" or "very important."⁷ Roughly the same percentage of survey respondents said they had taken one or more courses in statistics. One is left to ponder whether the AAPC members who thought research was important were those who had taken statistic courses, since Henderson *et al.*⁸ did not conduct analyses to examine this potential relationship.

A review of the literature on pastoral care published during the 1980s indicated that the pastoral care literature contained proportionately less quantitative research than that published in either nursing-home care or psychiatry.⁹ While one may question the meaningfulness of such a comparison between fields, the study also found pastoral care research to be less sophisticated than the research published in these fields. Sophistication was assessed in terms of nine criteria, including seven measures of internal and external validity.^{10,11} As Gartner *et al.*¹² concluded, "... the overall methodological quality of research in pastoral counseling journals is not satisfactory" (p.122). The authors went on to offer several recommendations for improving the quality of research in the profession.

In their responses to Gartner *et al.*,¹³ Rector¹⁴ and Schmidt¹⁵ both criticized the authors' interpretations of the study's findings. Despite his reservations about the authors' interpretations, Schmidt¹⁶ said: "The article has appropriately challenged our methodological rigor whether we pursue the quantitative or qualitative factors which govern our theory and practice" (p.129). Rector¹⁷ added that a stronger commitment needed to be made to providing pastoral counselors with the training required to conduct sound research.

⁴*Ibid.*

⁵Frederic C. Craigie, Jr., Ingrid Y. Liu, David B. Larson, and John S. Lyons, "A Systematic Analysis of Religious Variables in *The Journal of Family Practice*, 1976-1986" *The Journal of Family Practice*, 1988, Vol. 27, No. 5, pp. 509-513.

⁶Judith E. Thomson, "Pastoral Counseling in a Hospice: Have You Considered It?" *The Journal of Pastoral Care*, 1994, Vol. 48, No. 1, pp. 75-79.

⁷Daniel C. Henderson, John Gartner, and Floyd A. Chambers, "The Knowledge and Use of Quantitative Research by Pastoral Counselors," *Pastoral Psychology*, 1991, Vol. 40, No. 1, pp. 39-45.

⁸*Ibid.*

⁹Gartner *et al.*, *op.cit.*

¹⁰Donald T. Campbell, "Factors Relevant to the Validity of Experiments in Social Settings," *Psychological Bulletin*, 1957, Vol. 54, No. 4, pp. 297-312.

¹¹Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago, IL: Rand McNally, 1966).

¹²Gartner *et al.*, *op.cit.*

¹³Gartner *et al.*, *op.cit.*

¹⁴Lallene J. Rector, "A Response to: A Systematic Review of the Quantity and Quality of Empirical Research Published in Four Pastoral Counseling Journals: 1975-1984," *The Journal of Pastoral Care*, 1990, Vol. 44, No. 2, pp. 123-126.

¹⁵William S. Schmidt, "A Response to: A Systematic Review of the Quantity and Quality of Empirical Research Published in Four Pastoral Counseling Journals: 1975-1984," *The Journal of Pastoral Care*, 1990, Vol. 44, No. 2, pp. 126-129.

¹⁶*Ibid.*

¹⁷Rector, *op. cit.*

¹⁸Rector, *op. cit.*

Given Rector's¹⁸ and Schmidt's¹⁹ recognition of the short-comings of research in the field, and their support for its improvement, we wondered if the research had improved during the 1990s with respect to Gartner *et al.*'s criteria.²⁰ To find out, we conducted a review of the same journal literature published between 1990 and 1999, comparable to their review of the literature published between 1980 and 1989. We hypothesized that the proportion of research articles—particularly, quantitative research—would increase during the 1990s, in light of Gartner *et al.*'s findings.²¹ We similarly hypothesized that the sophistication of published research would increase over time.

Methods

The authors examined all articles published in *Pastoral Psychology*, *The Journal of Religion and Health*, and *The Journal of Pastoral Care* between 1990 and 1999. Gartner *et al.*²² examined articles published between 1980 and 1989 in these same three journals and the *The Journal of Pastoral Counseling*. We did not include the *The Journal of Pastoral Counseling* in the present study because it became an annual publication in the 1990s, publishing a single issue of roughly six to twelve articles each year, which dealt with a specific topical theme.

All articles that cited at least one reference were categorized as scholarly articles and counted, except editorials, commentaries, rejoinders, news stories, literary works (*e.g.* poems), letters to the editor, and regular columns by the same writer. By counting only scholarly articles, we reduced the article base somewhat, compared to Gartner *et al.*,²³ as we discuss later. Scholarly articles that contained method and results sections were further classified as research articles.

All research articles were photocopied and read by two researchers who served as judges to independently classify them according to the coding system described below. Research articles were defined as those containing a statement of purpose and a description of research methods and results, regardless of whether such topical headings were present.^{24,25,26,27} Initial agreement in the classification of articles as research or non-research was high—approximately 95%.

Research articles were further categorized as either qualitative or quantitative research. Studies that presented at least one descriptive or inferen-

¹⁸Schmidt, *op. cit.*

²⁰Gartner *et al.*, *op. cit.*

²¹Gartner *et al.*, *op. cit.*

²²Gartner *et al.*, *op. cit.*

²³Gartner *et al.*, *op. cit.*

²⁴Laura T. Flannelly, Kevin J. Flannelly, and Andrew J. Weaver, "A Systematic Review of Research Using Religious and Spiritual Variables in Three Major Oncology Nursing Journals: 1990-1999," *Oncology Nursing Forum*, (in press).

²⁵Andrew J. Weaver, Laura T. Flannelly, and Kevin J. Flannelly, "A Review of Research on Religious and Spiritual Variables in Two Primary Gerontological Nursing Journals: 1991-1997," *Journal of Gerontological Nursing*, 2001, Vol. 27, No. 9, pp.47-54.

²⁶Gartner *et al.*, *op. cit.*

²⁷Andrew J. Weaver, Laura T. Flannelly, Kevin J. Flannelly, Harold G. Koenig, and David B. Larson, "An Analysis of Research on Religious and Spiritual Variables in Three Mental Health Nursing Journals, 1991-1995," *Issues in Mental Health Nursing*, 1998, Vol. 19, No. 3, pp. 262-276.

tial statistic were classified as quantitative,^{28,29,30,31,32} while articles that did not contain a numerical presentation of findings were classified as qualitative.^{33,34} Articles that used phenomenological or ethnographic approaches or various qualitative techniques^{35,36} were classified as quantitative if they met the above criteria, to be consistent with Gartner *et al.*³⁷

Following the methods used by Gartner *et al.*,³⁸ we coded the quantitative studies on several criteria of internal and external validity. Four criteria of internal validity were used, including the three that Gartner *et al.*³⁹ used. These were: (1) the presence of a control group; (2) the presentation of reliability statistics for the measures used; (3) the presentation of descriptive or inferential statistics; and (4) the design and analysis with respect to the statistical control of extraneous variables or multiple independent variables. A study was coded as containing inferential statistics if it contained at least one inferential statistic. Two statistical experts agreed on the coding of the fourth criteria: 0 = no statistical controls (*e.g.* simple correlation, t-tests, one-way Analysis of Variance—ANOVA); 1 = analysis of two independent variables (*e.g.*, 2 X 2 ANOVA) or controlling for one variable other than the independent variable (*e.g.* Analysis of Covariance—ANCOVA); 2 = statistical analysis and control of the effects of three or more independent variables (*e.g.* Multiple Regression).

Five measures of external validity were recorded for all quantitative studies, including the four used by Gartner *et al.*⁴⁰ The five criteria were: (1) the sample size; (2) whether the sampling method was specified; (3) whether a random sample was used; (4) whether the response rate was specified; and (5) whether the study took measures at one or more than one point in time (*i.e.*, repeated measures). A log transformation of sample size was used in the correlation analyses. Initial agreement between the judges on the coding of all criteria of internal and external validity ranged between 96% to 99%.

In addition, we recorded (1) whether hypotheses were explicitly stated; (2) whether the limitations of the study's findings were discussed; and (3) the type of study it was (survey, experiment, *etc.*). The first two measures were also taken by Gartner *et al.*⁴¹ Initial agreement on coding these three measures ranged from 95% to 98%.

²⁸Flannelly, Flannelly and Weaver, *op. cit.*

²⁹Gartner *et al.*, *op. cit.*

³⁰Weaver, Flannelly, Flannelly, Koenig and Larson, *op. cit.*

³¹Weaver, Flannelly and Flannelly, *op. cit.*

³²Kimberly A. Sherrill, David B. Larson, and Mary A. Greenwood, "Is Religion Taboo in Gerontology?" Systematic Review of Research on Religion in Three Major Gerontological Journals, 1985-1991," *American Journal of Geriatric Psychiatry*, 1993, Vol. 1, No. 2, pp. 109-117.

³³Flannelly, Flannelly and Weaver, *op. cit.*

³⁴Weaver, Flannelly and Flannelly, *op. cit.*

³⁵Robert C. Bogdan and Sari K. Biklen, *Qualitative Research for Education: An Introduction to Theory and Methods*, Second Edition (Boston, MA: Allyn and Bacon, 1992).

³⁶O'Connor *et al.*, *op. cit.*

³⁷Gartner *et al.*, *op. cit.*

³⁸Gartner *et al.*, *op. cit.*

³⁹Gartner *et al.*, *op. cit.*

⁴⁰Gartner *et al.*, *op. cit.*

⁴¹Gartner *et al.*, *op. cit.*

Results

A total of 165 (22.4%) research studies were identified out of the 737 scholarly articles that were published in the three journals from 1990 through 1999. One hundred and thirty-six (18.4%) of the studies were classified as quantitative and 29 (3.9%) were classified as qualitative. Two articles that initially appeared to be research were excluded from the study. They both claimed to have conducted statistical tests but neither contained any study results.

Qualitative and Quantitative Research

Ten of the studies we classified as quantitative only used qualitative techniques or explicitly said they used a qualitative methodology,⁴² but they met our criteria for quantitative research because they presented statistical results. As O'Connor *et al.*⁴³ noted in their review of the same journals between 1993 and 1997, a number of studies employed both qualitative and quantitative methods.

More than eight out of 10 research articles published in these three journals during the last decade were quantitative studies by our definition. On average, 82.7% of the research published each year was quantitative, and there was no statistically significant change in the percentage of quantitative and qualitative studies between 1990 and 1999. As seen in Table 1, there were significant increases in the proportion of research studies, ($t = 2.67$, $p < .05$, one-tailed), especially quantitative studies ($t = 2.67$, $p < .05$, one-tailed), between the first and second half of the study-period, confirming our first hypothesis. The increases tended to be linear for both research articles in general ($r = .61$) and for quantitative research, in particular ($r = .58$), although neither trend was statistically significant.

TABLE 1
Percentage of Qualitative, Quantitative, and All Research
Articles From 1990 to 1999

Time-Period	Qualitative Research	Quantitative Research	All Research
1990-1994	3.4	16.2	19.6
1995-1999	4.2	20.2*	24.4*

*Significant increase between time-periods, $p < .05$ (one-tailed)

Table 2 shows the different kinds of research methods employed by the qualitative and quantitative research we examined. The vast majority of quantitative research used survey methods, whereas the most common method used in qualitative research was interviewing. Since O'Connor *et al.*⁴⁴ presented a thorough assessment of the quality of qualitative research published in recent years in the three journals covered in the current study, the qualitative studies were not included in further analyses.

⁴²O'Connor *et al.*, *op. cit.*

⁴³O'Connor *et al.*, *op. cit.*

⁴⁴O'Connor *et al.*, *op. cit.*

Table 2
Types of Research Methods used in Qualitative and Quantitative Studies

Method	Qualitative		Quantitative	
	N	%	N	%
Case Study	5	17.2%	—	—
Focus Group	2	6.9%	—	—
Observation	3	10.3%	—	—
Interview	13	44.8%	6	4.4%
Program Evaluation	1	3.5%	2	1.5%
Content Analysis	—	—	1	0.8%
Delphi	—	—	1	0.8%
Literature Review	—	—	4	2.9%
Experiment	—	—	2	1.5%
Quasi-experiment	—	—	2	1.5%
Survey/Questionnaire	5	17.2%	118	86.7%
Total	29	100.0%	136	100.0%

Internal Validity of Quantitative Studies

Table 3 shows the percentage of different kinds of studies that employed statistical controls in their designs and analyses. Only those studies that conducted statistical tests (*i.e.*, those that used inferential statistics) are presented in the table. Overall, most studies only used statistics that did not control for the potential influence of predictive variables other than the independent variable on the observed effects. Simple correlation coefficients, *t*-tests and one-way ANOVAs were the most commonly used statistical tests,⁴⁵ with 69.5% of the 97 studies that conducted statistical analyses employing one of these three tests. The few studies we found that controlled for one variable other than the independent variable used 2X2 factorial ANOVA (3 studies) or ANOVA (4 studies). Most of the studies (11 out of 13) that attempted to control for multiple independent variables used multiple regression.

TABLE 3

Percentage of Quantitative Studies Utilizing Statistical Controls in Their Analyses^a

	Controlling for:					
	No Controls		One Variable		Two or More	
	N	%	N	%	N	%
Survey	74	81.3%	5	5.5%	12	13.2%
Quasi-Experiment and Experiment	1	33.3%	2	66.6%	0	—
Other	2	66.6%	0	—	1	33.3%
Total	77	79.4%	7	7.2%	13	13.4%

^aThe sample includes only those studies that used inferential statistics

⁴⁵George A. Ferguson and Yoshio Takane, *Statistical Analysis in Psychology and Education*, Sixth Edition (New York, NY: McGraw Hill, 1989).

TABLE 4

Percentage of Quantitative Studies Meeting Each of Three Criteria of Internal Validity

Methodology	N	Control Group		Reliability Statistics		Inferential Statistics	
		N	%	N	%	N	%
Survey	118	15	12.7%	44	37.3%	91	77.1%
Quasi-Experimental and Experimental	4	1	25.0%	2	50.0%	3	75.0%
Other	14	0	—	0	—	3	21.4%
Total	136	16	11.8%	46	33.8%	97	71.3%

Table 4 gives the percentage of studies meeting the other three criteria of internal validity. The use of control groups was low, even among the quasi-experimental and experimental studies. Only one experiment used a true control group, and three of the four used within-subject, pretest/posttest designs in which subjects served as their own controls. Although the use of reliability statistics was higher among survey studies, it was still relatively low. The most commonly met criterion was the use of inferential statistics. Among the 97 studies that used inferential statistics, the most frequently used ones were, respectively, one-way ANOVAs (27.4% of studies), t-tests (24.2%), correlations (15.8%) and chi-squares (6.3%). These are unduplicated counts. For example, if a study used t-tests as post-hoc tests following an ANOVA they were counted as using ANOVA, but not counted as using t-tests. And, if they used t-tests and correlations they were counted as using t-tests, but not correlations. Of course, this ignores the fact that t is actually used to determine the significance of r .⁴⁶

A 3 (criteria) by 2 (time-periods) Multivariate Analysis of Variance (MANOVA)^{47,48} conducted on the percentages for the total sample, found no significant increase in the percentage of studies meeting any of the criteria (*i.e.*, having a control group, reliability statistics, inferential statistics, and statistical controls) over time. However, univariate analyses found significant differences in the percentage of studies meeting each criterion ($F = 98.3$, $p < .001$). Orthogonal contrasts confirmed that significantly more studies used reliability and inferential statistics than used control groups ($F = 169.0$, $p < .001$) and that significantly more studies reported inferential statistics than reliability statistics ($F = 109.5$, $p < .001$).

External Validity of Quantitative Studies

Sample size varied widely, within and across the various types of studies. One study analyzed more than 42,000 chaplain interventions and another examined religious themes in over 17,000 articles published in medical journals. The sample size of survey studies ranged from four to over 5000

⁴⁶Jacob Cohen and Patricia Cohen, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Hillsdale, NJ: Lawrence Erlbaum, 1975).

⁴⁷Robert J. Harris, *A Primer of Multivariate Statistics*, Second Edition (New York, NY: Academic Press, 1985).

⁴⁸Barbara G. Tabachnick and Linda S. Fidell, *Using Multivariate Statistics*, Third Edition (New York, NY: HarperCollins, 1996).

participants, with the median being 160 participants. Apart from survey research, the median sample size of the quantitative studies was 40.

The overall percentage of studies that specified sampling methods and response rates were 62.5% and 42.9%, respectively (see Table 5). Naturally, these percentages were somewhat higher among the survey studies, among which 65.3% specified the sampling method and 51.7% specified the response rate. The six studies in the "Other" category that reported response rates included a study that used the Delphi technique⁹ and five that conducted interviews. Less than a quarter of the 118 survey studies used random samples. Three of the four quasi/experimental studies used repeated measures in the form of pretest/posttest designs. As with the assessments of internal validity, MANOVA found no increase in the percentage of studies meeting the external validity criteria over time, but significant differences were found among the criteria themselves ($F = 42.2, p < .001$). Orthogonal contrasts showed that studies met the sampling method and response rate criteria more often than they met the other two criteria of internal validity ($F = 59.1, p < .001$), and that significantly more studies used random samples than used repeated measures ($F = 11.0, p < .05$).

TABLE 5
Percentage of Quantitative Studies Meeting Each of Four Criteria of External Validity

Methodology	N	Specified				Repeated Measures		Random Sample	
		Sampling Method		Response Rate		N	%	N	%
		N	%	N	%				
Survey	118	77	65.3%	61	51.7%	6	5.1%	28	23.7%
Quasi-Experimental and Experimental	4	1	25.0%	-	-	3	75.0%	0	-
Other	14	7	50.0%	6	42.9%	2	14.3%	2	14.3%
Total	136	85	62.5%	67	42.9%	11	8.1%	30	22.1%

Correlations Between Validity and Other Criteria

There was a positive correlation between compliance with three out of four criteria of internal validity: the use of reliability statistics, inferential statistics and statistical controls (r 's = .43 to .70, $p < .001$). Three of five measures of external validity were also positively related to one another: specifying the sampling method, stating the response rate, and using a random sample (r 's = .35 to .41, $p < .001$). Only one significant correlation was found between the four criteria of internal validity and the five criteria of external validity—this being a positive correlation between degree of statistical control and the specification of the sample method used ($r = .23, p < .01$). No other statistically significant correlations among the eleven criteria were found, and contrary to our second hypothesis, no increase in compliance with any of the eleven criteria was observed over time.

⁹Gordon F. West and JoLynne Reynolds, "The Applicability of Selected Rational-Emotive Therapy Principles for Pastoral Counseling," *The Journal of Pastoral Care*, 1997, Vol. 51, No. 2, pp.187-194

Validity and Research Sophistication: 1980s vs 1990

Table 6 compares the percentage of quantitative studies we found that met Gartner *et al.*'s⁵⁰ major criteria of validity and research sophistication during the 1990s to that found by Gartner *et al.*⁵¹ during the 1980s. Frequency counts were calculated from the percentages presented by Gartner *et al.*⁵² in order to perform chi-square analyses on the data. These analyses revealed significant improvements in "research methods" compliance during the 1990s on three of the seven criteria listed in the table. These were: (a) specifying the sampling method used ($X^2 = 31.0, p < .001$), (b) reporting the response rate ($X^2 = 4.37, p < .05$), and (c) discussing some of the limitations of the study ($X^2 = 27.7, p < .001$). The frequency data needed to conduct a chi-square on the use of inferential statistics could not be discerned from Gartner *et al.*'s⁵³ presentation of their results.

TABLE 6

Percentage of Studies Meeting Gartner *et al.*'s Validity Criteria in the 1980s and the 1990s

Criteria	Gartner <i>et al.</i>	Current Study
Internal Validity		
Reliability	40%	34%
Control Group	4%	12%
External Validity		
Sampling Method	20%	63%**
Response Rate	33%	43%*
Repeated Measures	7%	8%
Research Sophistication		
Hypotheses Stated	25%	32%
Limitations Stated	4%	58%**

* $p < .05$

** $p < .001$

Discussion

Our findings indicate that the percentage of research articles published in *Pastoral Psychology*, *The Journal of Religion and Health* and *The Journal of Pastoral Care* increased significantly between 1990 and 1999 and that the increase tended to be linear. This trend is mainly due to a significant increase in the percentage of quantitative studies that were published, supporting our first hypothesis. The percentage of quantitative studies published even during the first half (1990-1994) of the years in our sample (16.2%) would also seem to be substantially higher than that found by Gartner *et al.*⁵⁴ during the 1980s (5.3%). However, our findings of 16.2% quantitative studies in 1990-1994 and 20.2% in 1995-1999 are calculated on a more restrictive article base than Gartner *et al.*⁵⁵ used. We restricted our count of articles to those that contained at least one reference—what we refer

⁵⁰Gartner *et al.*, *op. cit.*

⁵¹Gartner *et al.*, *op. cit.*

⁵²Gartner *et al.*, *op. cit.*

⁵³Gartner *et al.*, *op. cit.*

⁵⁴Gartner *et al.*, *op. cit.*

⁵⁵Gartner *et al.*, *op. cit.*

to as scholarly articles. This approach eliminated from our base-count, such articles as personal reflections, which are common in the *Journal of Pastoral Care*, but do not ordinarily appear in scientific journals in other disciplines, except perhaps as editorials. If we had strictly adhered to Gartner *et al.*'s⁵⁶ procedure of counting all articles as our base, we would have found a relatively lower proportion of quantitative studies. Restricting the base to scholarly articles probably provides a somewhat fairer representation of the discipline compared to the rate of research in other scientific disciplines, which mainly publish scholarly articles. Regardless of this definitional difference, a distinct increase in quantitative articles was observed across the 1990s.

We were not able to detect any increase in the research sophistication of the quantitative studies published between 1990 and 1999. Sample size varied widely, control groups were rarely used, and reliability statistics were reported by roughly one-third of the studies. The use of inferential statistics was the most commonly met criteria of internal validity, with over two-thirds of the studies using them. However, among the 97 studies that used inferential statistics, less than 14% of them controlled for two or more variables in their analyses.

In examining the correlations among the criteria of internal validity and those among the criteria of external validity, one sees two separate clusters of measures, both of which relate to the study statistics employed. In the case of internal validity, the three criteria dealing with statistical analyses are all related to each other: the use of 1) reliability statistics, 2) inferential statistics and 3) statistical controls. But they are not related to the other criteria of internal validity—the use of a comparison or control group. Thus, it appears that researchers who know about one aspect of statistical analysis tend to know about the other two that are among our criteria of internal validity. This appears to be consistent with Henderson *et al.*'s⁵⁷ findings that 63% of the AAPC members they surveyed had taken one or more statistics courses. The use of control groups is conceptually somewhat different than the other criteria because it falls squarely within the realm of experimental design,^{58,59,60,61} and Henderson *et al.*⁶² found that only 49% of the AAPC members they surveyed had taken a course in research design. Although some might not think control or comparison groups are needed for surveys—which made up the bulk of the quantitative studies—their use would allow researchers to better understand and interpret their findings. If one uses a comparison group and one has the advanced statistical and design skills needed, the survey results can be analyzed as a quasi-experimental design.⁶³

In the case of external validity, the three criteria dealing with sampling are all related (*i.e.* specification of sampling method, use of a random sample, specification of the return rate), but they are unrelated to the other two criteria: sample size and repeated measures. Once again, then, it seems that researchers who tend to know one aspect of sampling also know about the others. The lack of correlation between sample size and the other three cri-

⁵⁶Henderson *et al.*, *op. cit.*

⁵⁷Campbell, *op. cit.*⁵⁸Campbell and Stanley, *op. cit.*

⁵⁹Allen L. Edwards, *Experimental Design in Psychological Research*, Fifth Edition (New York, NY: Harper & Row, 1985).

⁶⁰Geoffrey Keppel, *Design and Analysis: A Researcher's Handbook* (Englewood Cliffs, N.J.: Prentice Hall, 1973).

⁶¹Henderson *et al.*, *op. cit.*

⁶²Campbell and Stanley, *op. cit.*

⁶³Gartner *et al.*, *op. cit.*

teria relating to sampling is not surprising, however, since the other criteria are unrelated to sample size on a more practical level, since a researcher may report or may not report those three measures regardless of the size of the sample that is taken. Like the use of control groups, the use of repeated measures falls within the realm of research design. And, even those knowledgeable about statistics may have only a limited understanding of experimental or quasi-experimental design. As noted above, although almost two-thirds of AAPC members surveyed by Henderson *et al.* had taken one or more statistics courses, less than half had taken a course in research design.

Because our definition of quantitative studies was identical to that of Gartner *et al.*,⁶⁴ we were able to directly compare our results to theirs with respect to their measures of validity and research sophistication. The comparison of the 1980 data with the 1990 data partially supports our second hypothesis, that research sophistication would increase over time. Although no increase in sophistication was found between 1990 and 1999, significant increases were found between the 1980s and the 1990s on three of Gartner *et al.*'s⁶⁵ measures: 1) specification of the sampling method, 2) reporting of the study's response rate, and 3) discussion of the limitations of the study. The first two represent improvements in aspects of external validity compared to the earlier findings of Gartner *et al.*⁶⁶

The studies in our sample used a variety of research methods (see Table 2). The vast majority were surveys, however, and only a handful used experimental (*e.g.* clinical trials) or quasi-experimental designs. The dearth of experimental or, more specifically, clinical trial studies could partly account for the low percentage of studies that stated specific hypotheses. Nevertheless, hypothesis testing can be performed through surveys, but this was done in less than 32% of the surveys we examined. In reading the studies in our sample it was obvious that many of them were exploratory in nature. They were designed to examine possible relationships between variables, without trying to predict (or hypothesize about) what those relationships would be. In such cases, one would not expect the statement of hypotheses. Thus, the low rate of hypothesis testing may reflect the research sophistication of the field, but not the research sophistication of the researchers themselves. Exploratory studies can be initially important but with time lead to more focused hypotheses-testing. Unfortunately, the latter is not yet happening very frequently.

The overall findings suggest that many researchers in the field do not have a very sophisticated knowledge of statistical sampling or statistical analysis. As noted in the preceding paragraph, strides have been made in this area since the 1980s, but no progress was seen across the 1990s. And, our analyses could find no improvement in the sophistication of statistical analyses and research design. As for progress in teaching about research, we know of no recent study comparable to Henderson *et al.*'s 1991 survey⁶⁷ that has tried to assess pastoral counselors' knowledge of statistics and research design. Of course, a follow-up of Henderson *et al.*'s⁶⁸ study is needed.

Research provides an empirical basis on which to guide a discipline, replac-

⁶⁴Gartner *et al.*, *op. cit.*

⁶⁵Gartner *et al.*, *op. cit.*

⁶⁶Henderson *et al.*, *op. cit.*

⁶⁷Henderson *et al.*, *op. cit.*

⁶⁸Gartner *et al.*, *op. cit.*

ing opinion and anecdote with objective, replicable scientific findings. The value of this guidance, however, depends upon the quality of the research. Although we found some improvement in the methodological rigor of the research published in pastoral care, more progress needs to be made. Along with Gartner *et al.*,⁶⁹ and Henderson *et al.*,⁷⁰ we offer several recommendations for increasing the quality of quantitative research in the discipline.

Although Henderson *et al.* found that roughly two-thirds of AAPC members had taken one or more statistics courses, we believe more pastoral counselors and chaplains would benefit from a better understanding of statistics. To this end, we believe training and educational programs should put greater emphasis on statistics in their curriculum—offering at least one course that includes sampling methodology and inferential statistics. A course in research design should also be required. To enhance the master's and doctoral degree student's awareness of the field and clinical relevance of such research, the students should be strongly urged to: (1) design and conduct their own studies, as part of supervised, individual or group projects, or (2) work on on-going research projects.

Of course, our approach is limited in the sense that we did not attempt to assess the kinds of topics investigated in the studies we found, or to evaluate the value of their findings to the field. Neither did we assess the research methodology of the qualitative studies published during the study period. But O'Connor *et al.*⁷¹ presented such an assessment for qualitative research published in the same three journals between 1993 and 1997.

As this review reveals, researchers themselves should become better acquainted with research design and more sophisticated statistical procedures. The statistical software that is needed is readily available, but one must learn the logic of design and analysis beyond t-tests, one-way ANOVAs, and chi-squares, to be better able to analyze and interpret the data one intends to collect. Like Gartner *et al.*,⁷² we recommend interdisciplinary collaboration as a worthwhile alternative to mastering statistics and design. Finally, we agree with Henderson *et al.*⁷³ that institutions employing pastoral counselors, chaplains, or other clergy should encourage and support research as part of their regular duties.

Pastoral care is at a professional crossroad,⁷⁴ but without quality research available to guide clinical practice, professionals will continue to be informed by opinion and anecdote. In order to promote a constructive dialogue regarding religious variables in health-care settings, we need to encourage assessment of the beneficial, neutral, and occasionally harmful role of religion among patients and providers.⁷⁵ Building a credible empirical basis for pastoral care will mean an increased impact of pastoral research on the general scientific community, as well as more effective spiritual care by clinical staff. However, in order to fully share the benefits of clinical pastoral training, pastoral professionals at the master's and doctoral levels should develop a knowledge base that will enable them to understand and conduct quantitative as well as qualitative research. ♣

⁶⁹Henderson *et al.*, *op. cit.*

⁷⁰Henderson *et al.*, *op. cit.*

⁷¹O'Connor *et al.*, *op. cit.*

⁷²Gartner *et al.*, *op. cit.*

⁷³Henderson *et al.*, *op. cit.*

⁷⁴Henderson *et al.*, *op. cit.*

⁷⁵Craigie *et al.*, *op. cit.*