concerns that denied privileges may be reported to the data bank. This study also involved a limited geographic scope and may not be generalizable to the United States as a whole. For example, in comparison with rural hospitals nationally, our study hospitals tended to be smaller, and fewer were nongovernment, for-profit institutions.[6] However, comparisons with earlier surveys show that the results of this study are very similar to those of other studies.[7,8]

Rural hospital administrators' estimation of the relative worth of the National Practitioner Data Bank and their reliance on it could increase as the number of years of information contained in the data bank grows. Estimates of its worth would also increase if it develops a strong reputation for completeness and accuracy. However, it will be important to monitor the effect of mandatory reporting to the data bank on disciplinary actions that hospitals take against members of their staff and on the behaviors of providers so that the data bank does not have a paradoxical effect on quality assurance. □

Acknowledgments

This research was primarily funded by the Health Resources and Services Administration's (HRSA's) Bureau of Health Professions (grant HRSA 93-1015 [P]); supplemental support was provided by HRSA's Office of Rural Health Policy (grant CSR 000007-03-0).

We wish to thank Fitzhugh Mullan, MD, former director of the Bureau of Health Professions; Idelle Price Smith, MSW, MPH, program management officer, Division of Quality Assurance; and Robert E. Oshel, PhD, research director, Division of Quality Assurance, for their support of this research.

References

1. Snelson E. Quality assurance implications of federal peer review laws. The Health Care Quality Improvement Act and the National Practitioner Data Bank. *Qual Assur Util Rev.* 1992;7:2–11.
2. *Electronic 1993 Department of Agriculture Update of Rural-Urban Continuum Codes.* Washington, DC: Agriculture and Rural Economy Division, Economic Research Service, US Dept of Agriculture; 1993.
3. *American Hospital Association Guide to the Health Care Field.* Chicago, Ill: American Hospital Association; 1993.
4. *Rand-McNally 1994 Road Atlas—United States, Canada, Mexico.* Skokie, Ill: Rand-McNally & Co; 1994.
5. Neighbor WE, Baldwin LM, West PA, Hart LG. *Experience of Rural Hospitals with the National Practitioner Data Bank.* Seattle, Wash: University of Washington; 1994. WAMI Rural Health Research Center Working Paper 29.
6. Hart LG, Amundson BA, Rosenblatt RA. Is there a role for the small rural hospital? *J Rural Health.* 1990;6:101–118.
7. *National Practitioner Data Bank: Usefulness and Impact of Reports to Hospitals.* Washington DC: US Dept of Health and Human Services; 1993.
8. Mullan F, Politzer RM, Lewis CT, Bastacky S, Rodak J, Harmon RG. The National Practitioner Data Bank: report from the first year. *JAMA.* 1992;268:73–79.

The Effectiveness of Family Planning Programs Evaluated with True Experimental Designs

Karl E. Bauman, PhD

ABSTRACT

Objectives. This paper describes the magnitude of effects for family planning programs evaluated with true experimental designs.

Methods. Studies that used true experimental designs to evaluate family planning programs were identified and their results subjected to meta-analysis.

Results. For the 14 studies with the information needed to calculate effect size, the Pearson *r* between program and effect variables ranged from −.08 to .19 and averaged .08.

Conclusions. The programs evaluated in the studies considered have had, on average, smaller effects than many would assume and desire. (*Am J Public Health.* 1997;87:666–669)

Introduction

Organized family planning programs have been implemented in developing countries for more than 3 decades. In 1990, they cost an estimated $4.5 billion worldwide.[1] This paper presents the first meta-analysis of family planning program evaluations that have used true experimental designs. Because none of the studies measured fertility or unwanted fertility, the primary ultimate outcomes targeted by the evaluated programs, this synthesis is limited to outcome variables believed to mediate the program–fertility relationship.

This analysis is restricted to evaluations that used true experimental designs because many family planning program evaluators and researchers who evaluate other types of human programs agree that properly implemented true experimental designs provide the strongest evidence for assessing program effects.[2–18] The distinguishing feature of the true experimental design is that the units of study—whether individuals, areas, or clinics—are randomly allocated to different treatment conditions. The strength of the properly implemented true experimental design, relative to all other research designs, is that random allocation of study units to comparison groups that did and did not receive the program or that received program variations can provide the most confident inference that an association between the program and outcome variables is causal rather than spurious.[19,20] When correlation coefficients or other statistics derived from true experimental

The author is with the Department of Health Behavior and Health Education, School of Public Health, and the Carolina Population Center, University of North Carolina at Chapel Hill.

Requests for reprints should be sent to Karl E. Bauman, PhD, Department of Health Behavior and Health Education, School of Public Health, Rosenau Hall CB #7400, University of North Carolina, Chapel Hill, NC 27599-7400.

This paper was accepted June 28, 1996.

data gauge program effects, it often is unnecessary to introduce additional variables in analyses to determine whether the association is noncausal because those variables were balanced by the design.

Methods

An attempt was made to identify all family planning program evaluations that had used a true experimental design. For a study to be classified as having used a true experimental design, the units studied had to be randomly allocated to comparison conditions. The study and its findings had to be described in written form and could be published or unpublished. Studies that reported findings for program outcomes by December 31, 1992, were eligible.

The computer database POPLINE was searched to identify the studies. In addition, pertinent journals and books were reviewed, holdings of the Carolina Population Center Library were scanned, and an extensive review of the literature on techniques for family planning program evaluation was conducted. Furthermore, experts in family planning evaluation research were asked to identify evaluations that had used the design, and population scholars were asked to review the list of evolving studies to ensure completeness.

The analysis focuses mostly on outcome variables assumed to be primary determinants of fertility and unwanted fertility, such as clinic visits, contraceptive use, and contraceptive continuation. The studies did not measure fertility.

Some of the identified studies determined whether there were statistically significant program effects, but few considered effect size. Our concern is with magnitude of effect. From data provided in the reports of the studies, we calculated Pearson product–moment correlation coefficients (rs) to measure effect sizes for the associations between program and outcome variables. These coefficients can range from -1.00 to 1.00, 1.00 indicating a perfect correlation between the program and the desired effect. The r^2 value (sometimes called the coefficient of determination) can range from 0 (no relationship) to 1.00 (100% of the variation in the outcome variable attributable to the program). The pooled method was used to estimate standard deviations required for the calculations.[21] Meta-analysts sometimes use other statistics to determine magnitude of effect, such as d, which can be directly derived from r.[21] We prefer r because it is more widely known. Meta-analysis methodology is considered in detail in Cooper and Hedges.[22]

Within each of the 14 studies that presented the necessary data, effect sizes were calculated for relationships between each program comparison and selected outcome variables. All studies had multiple effect sizes because more than one program variation was evaluated, multiple outcome variables were measured, or more than one postprogram measurement of the outcome variable was obtained. The average effect size for each study and the average effect size across studies also were computed.

Results

Sixteen studies that used true experimental designs were identified.[23–38] Table 1 shows, for each study, (1) the country in which it was conducted, (2) the year of the evaluated program, (3) report authors, (4) the type of program evaluated, (5) the authors' conclusions (in terms of the outcome variables used in this analysis), and (6) the average effect size for the study.

In 13 of the 16 studies, the researchers concluded that program effects were positive. In 2 studies, the authors concluded that there were no significant program effects,[24,27] and, in 1 study, only unintended program effects were found.[29]

Effect sizes could not be calculated for the 2 studies that did not report the necessary data. Across the 14 studies that provided the necessary information, 51 correlation coefficients were calculated. These coefficients ranged from $-.10$ to .53. Effect sizes, which were averaged for each of the 14 studies, ranged from $-.08$ to .19.

The average r for all 14 studies, obtained by averaging the study averages, was .08 (SE = .28, $P > .05$). The average proportion of explained variance, obtained by averaging the r^2 value of each study, was .01; thus, on average, the programs accounted for no more than 1% ($r^2 \times 100 = 1.00\%$) of the variation in the outcome variables.

Discussion

Meta-analyses are limited to the studies that are found, and it is possible that, even with our intensive efforts, some studies were not found. We believe that number is negligible. We are unaware of other studies that meet our criteria for inclusion, and we assume that if they do exist there are so few that their addition to our data would have only a small influence, if any, on the effect size we found. Studies of other topics with significant effects are more likely to be published, and hence more readily found, than studies that do not produce significant effects.[39,40] To the extent this is true for family planning program evaluations using true experimental designs, our effect sizes may exaggerate what exists for all studies that have been conducted.

We discuss our findings relative to other considerations of the magnitude of family planning program effects. Attributions of effect have ranged from no impact for fertility[41] to large fertility declines in some developing countries.[39] Sometimes the argument focuses on the relative contributions of family planning services and socioeconomic development.[42] Heretofore, in spite of their methodological rigor, findings from evaluations using true experimental designs have not been considered in the debate.

The debate often focuses on fertility. For the one study with a measure most closely resembling fertility (pregnancy 12 months after the program), the effect size (r) was .01.[27] All of the other effect sizes involved variables assumed to intervene between a program and fertility, such as clinic use, contraceptive prevalence, and contraceptive continuation. The correlation between mediators and dependent variables is imperfect[43]; thus, for the subjects included in this meta-analysis, the program effect size for fertility must have been smaller than the .08 average we found. Because fertility was not measured in these other studies, a more exact estimate of the smaller program effect for fertility cannot be determined.

Most behavioral scientists would consider our average effect size of .08 to be small.[21] Still, small effects in large populations can yield substantial reductions in the total numbers of births. Estimates of the number of births averted cannot be derived from these studies, however, because number of births was not measured. In any case, effect sizes were smaller than desired by many family planning program advocates, and it appears that the fertility for the samples studied derived largely from factors other than the types of family planning programs evaluated with true experimental designs.

We would have preferred more than 14 studies from which to generate effect sizes. We would have been more concerned about this possible limitation, however, if the effect sizes had varied

TABLE 1—Selected Characteristics of True Experimental Studies of Family Planning Program Effects

Country (Year)	Author(s)	Program	Author Conclusion	*r*
Brazil (1981)	Foreit & Foreit[23]	Supervisory visit frequency	No effect for number of new clients, returns	.09
Colombia (1969)	Simmons[24]	Home visit, mailed pamphlet	No effect for clinic attendance	.04
Colombia (1979–1981)	Gomez[25]	Health promoter distribution	Increased oral contraceptive use	.09
Hong Kong (1966)	Population Council[26]	Home visit, visitor qualification	Increased clinic attendance	.14
Hong Kong (1968/69)	Chan[27]	Home visit	No effect for IUD retention	.02
Korea (1966/67)	Yang[28]	Mother education class	IUD prevalence increased in low prevalence areas; termination decreased with special education	.02
Korea (1967)	Bang[29]	Early clinic return schedule	Increased IUD removal and expulsion	−.08
Mexico (1986/87)	Macorra et al.[30]	Supermarket condom location	Increased condom sales	. . .[a]
Nepal (1973)	Gubhaju et al.[31]	Single-purpose field worker	Increased pill continuation	.08
Nigeria (1984–1986)	Omu et al.[32]	Postpartum education	Increased sterilization and contraceptive use	.19
Peru (1985–1987)	Zambrano et al.[33]	Physician frequency at clinic	Increased visits and IUD insertions	. . .[a]
Philippines (1973)	Phillips et al.[34]	Motivator incentive and at-large affiliation	Increased motivator performance	.16
Sri Lanka (1983–1985)	Vidyasagara et al.[35]	Satisfied user–midwife team	Increased IUD acceptors, no effect for termination	.14
Taiwan (1964)	Freedman & Takeshita[36]	Home visit, wife vs couple involved, mailing	Home visit increased contraceptive acceptance; wife or couple involvement equally effective; mailing ineffective	.05
Taiwan (1971)	Chang et al.[37]	Field worker incentive	Increased contraceptive acceptance	.07
Thailand (1969/70)	Rosenfield & Limcharoen[38]	Midwife or physician pill prescription	Increased acceptance and continuation with midwife prescription	.09

Note. IUD = intrauterine device.
[a]Coefficient could not be calculated from information provided.

more across studies. More studies would have allowed determining effect sizes for different types of programs, outcomes, and geographical regions. We also would have preferred true experiments that assessed fertility effects and more types of programs evaluated with true experiments.

True experiments have been conducted on a wide range of family planning programs and countries. Moreover, such studies have been spread quite evenly throughout the 30 years of organized family planning programs and have included programs implemented as recently as the late 1980s. However, many types of programs have not been evaluated with these designs, and, in most countries, family planning programs have not been evaluated with a true experimental design. We assume that the effect sizes we present are reasonable approximations rather than perfect indicators of what might have been found if a more representative set of evaluations had been available. Assessment of this assumption must await the findings of additional true experiments.

Most of the programs evaluated with true experimental designs can be viewed as special family planning activities or components that are added to basic clinic services that provide contraceptives (e.g., home visitation or follow-up) or as activities that are varied within a basic clinical program (e.g., frequency of supervision or field-worker incentive). Therefore, the findings of this meta-analysis can best be generalized to such conditions. True experimental designs have not been used to evaluate programs that provide only contraceptives through basic clinical services, and no study has used a control group for which program-preferred contraceptives were completely unavailable. No comprehensive national program or program that simulates national-level intervention has been evaluated with a true experimental design. These characteristics of existing family planning program evaluations limit the extent to which the findings of this meta-analysis can inform the debate about the more general impact of family planning programs on fertility in developing countries.

The control groups in true experiments may include some family planning program input and thereby yield an underestimate of program impact. Moreover, national populations might not be completely represented by evaluations using true experimental designs. These possible limitations are shared with many other evaluations that involve other types of research designs.

In addition to the finding of a much smaller effect size than family planning program proponents would like to see, an observation from this meta-analysis is that although there has been a large investment in family planning programs and there is consensus that true experiments would most rigorously evaluate their effects, only 16 completed evaluations used the design. Moreover, as discussed earlier, the existing studies have not addressed many

of the important questions about family planning programs and their effects. Why are there so few true experiments to inform family planning policy and programming? That there are 16 studies is proof that true experimental designs can be used for this purpose. That methodological, practical, and ethical barriers necessarily preclude their use has been challenged, and it has been shown that many of the limitations attributed to randomized experiments apply equally to the less rigorous designs that are used much more often.[2] Family planning policies and programs could be better informed if relevant family planning program models were more frequently evaluated with true experimental designs. □

Acknowledgments

Funding for this research was provided by the US Agency for International Development under the Evaluation of Family Planning Impact Project (contract DPE-3060-C-00-1054-00).

This research was presented at the annual meeting of the Population Association of America, April 1995, San Francisco, Calif.

Dr David Guilkey, Dr Chirayath M. Suchindran, Dr Amy O. Tsui, and Ms Claire I. Viadro made significant contributions to this paper.

References

1. Lande RE, Geller JS. Paying for family planning. *Popul Rep.* 1991; series J(39).
2. Bauman KE, Viadro CI, Tsui AO. Use of true experimental designs for family planning program evaluation: merits, problems, and solutions. *Int Fam Plann Perspect.* 1994;20:108–113.
3. Bennett EA, Lumsdaine AA. Field trial designs in gauging the impact of fertility planning programs. In: Bennett EA, Lumsdaine AA, eds. *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs.* New York, NY: Academic Press Inc; 1975.
4. Brass W. *Comments on Comparison Strategies for the Evaluation of Family Planning Impact. Methods of Measuring the Impact of Family Planning Programmes on Fertility: Problems and Issues.* New York, NY: United Nations; 1978.
5. Bulatao RA, Lee RD. An agenda for research on the determinants of fertility in the developing countries. In: Bulatao RA, Lee RD, eds. *Determinants of Fertility in Developing Countries.* New York, NY: Academic Press Inc; 1983.
6. Cuca R, Pierce CS. *Experiments in Family Planning: Lessons from the Developing World.* Baltimore, Md: Johns Hopkins University Press; 1977.
7. Fisher AA, Laing J, Stoeckel J. Guidelines for overcoming design problems in family planning operations research. *Stud Fam Plann.* 1985;16:100–105.
8. Hermalin AI, Chandrasekaran C. Overview. In: Chandrasekaran C, Hermalin AI, eds. *Measuring the Effect of Family Planning Programs on Fertility.* Dolhain, Belgium: Ordina Editions; 1975.
9. Hermalin AI. Effects on fertility. In: Ross J, ed. *International Encyclopedia of Population.* New York, NY: Free Press; 1982.
10. Hernandez DJ. The impact of family planning programs on fertility in developing countries: a critical evaluation. *Soc Sci Res.* 1981;10:32–66.
11. Hernandez DJ. *Success or Failure? Family Planning Programs in the Third World.* Westport, Conn: Greenwood Press; 1984.
12. Lloyd CB, Ross JA. *Methods for Measuring the Fertility Impact of Family Planning Programs: The Experience of the Last Decade.* New York, NY: Population Council; 1989.
13. Sherris JD, London KA, Moore SH, Pile JM, Watson WB. The impact of family planning programs on fertility. *Popul Rep.* 1985;XIII:J733–J772.
14. Population Information Program. Operations research: lessons for policy and programs. *Popul Rep.* 1986;XIV:J815–J852.
15. Reynolds J. Evaluation of family planning program performance: a critical review. *Demography.* 1972;9:69–86.
16. Ross JA, Lloyd CB. Methods for measuring the fertility impact of family planning programmes: the experience of the last decade. In: Phillips JF, Ross JA, eds. *Family Planning Programs and Fertility.* New York, NY: Oxford University Press Inc; 1992.
17. Schultz TP. *Effectiveness Evaluation of Family Planning: Case Study Taiwan.* Santa Monica, Calif: Rand Corp; 1972.
18. Srinivasan K. Recent developments in the evaluation of the demographic impact of family planning programmes. Presented at the International Population Conference, August 1977, Mexico City, Mexico.
19. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research.* Boston, Mass: Houghton Mifflin Co; 1963.
20. Wells HB. Principles of experimental design. In: *The Methodology of Measuring the Impact of Family Planning Programmes on Fertility.* New York, NY: United Nations; 1979.
21. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
22. Cooper H, Hedges LV. *The Handbook of Research Synthesis.* New York, NY: Russell Sage Foundation; 1994.
23. Foreit JR, Foreit KG. Quarterly versus monthly supervision of CBD family planning programs: an experimental study in Northeast Brazil. *Stud Fam Plann.* 1984;15:112–120.
24. Simmons AB. Information campaigns and the growth of family planning in Colombia. In: Stycos JM, ed. *Clinics, Contraception, and Communication.* New York, NY: Appleton-Century-Crofts; 1973.
25. Gomez F. Community-based distribution: the case of Colombia. In: Wawer M, Huffman S, Cebula D, Osborn R, eds. *Health and Family Planning in Community-Based Distribution Programs.* Boulder, Colo: Westview Press; 1985.
26. Population Council. Hong Kong: an evaluation of field workers and decision-making in family planning programs. *Stud Fam Plann.* 1968;30:7–12.
27. Chan KC. Hong Kong: report of the IUD Reassurance Project. *Stud Fam Plann.* 1971;2:225–233.
28. Yang JM. *Studies in Family Planning and Related Programs in Rural Korea. Social Evaluation and Research Activities in Korea.* Seoul, Korea: Korea Sociological Association; 1972.
29. Bang S. *Can IUD Retention Be Improved by Prompt Check-Up Visits? Population and Family Planning in the Republic of Korea.* Seoul, Korea: Ministry of Health and Social Affairs; 1970.
30. Macorra L, Roca R, Townsend J. *Marketing of Condoms in Supermarkets: Cash Registers versus Regular Shelves as Point of Sale (Final Technical Report).* New York, NY: Population Council; 1987.
31. Gubhaju BB, Tuladhar J, Pande BR, Stoeckel J. Workshop-conference on population, family planning, and development in Nepal, August 1975, Berkeley, Calif.
32. Omu AE, Weir SS, Janowitz B, Covington DL, Lamptey PR, Burton NN. The effect of counseling on sterilization acceptance by high-parity women in Nigeria. *Int Fam Plann Perspect.* 1989;15:66–71.
33. Zambrano MR, Turner EM, Tafar RM, Lopez SO, Foreit JR, Garcia-Nunez J. *An Experiment to Improve the Efficiency, Effectiveness and Cost-Effectiveness of the IUD Insertion and Medical Back-Up Component of a CBD Program in Lima, Peru (Final Report).* Lima, Peru: Instituto Peruano de Paternidad Responsable; 1987.
34. Phillips JF, Silayan-Go A, Pal-Montano A. An experiment with payment, quota, and clinic affiliation schemes for lay motivators in the Philippines. *Stud Fam Plann.* 1975;6:326–334.
35. Vidyasagara NW, Abeywickrama D, Wickramasuriya K, de Silva SV. *Operations Research Study to Increase IUCD Acceptance in Sri Lanka.* Colombo, Sri Lanka: Family Health Bureau, Ministry of Health and Family Planning Association of Sri Lanka; 1985.
36. Freedman R, Takeshita JY. *Family Planning in Taiwan: An Experiment in Social Change.* Princeton, NJ: Princeton University Press; 1969.
37. Chang MC, Cernada GP, Sun TH. A fieldworker incentive experimental study. *Stud Fam Plann.* 1972;3:270–272.
38. Rosenfield AG, Limcharoen C. Auxiliary midwife prescription of oral contraceptives: an experimental project in Thailand. *Am J Obstet Gynecol.* 1972;114:942–949.
39. Smith ML. Publication bias and meta-analysis. *Eval Educ.* 1980;4:22–24.
40. White KR. The relation between socioeconomic status and academic achievement. *Psychol Bull.* 1982;91:461–481.
41. Pritchett LH. Desired fertility and the impact of population policies. *Popul Dev Rev.* 1994;20:1–55.
42. Bongaarts J. Population policy options in the developing world. *Science.* 1994;263:771–776.
43. Bongaarts J, Potter RG. *Fertility, Biology, and Behavior: An Analysis of the Proximate Determinants.* New York, NY: Academic Press Inc; 1983.