

INTERRUPTED TIME SERIES DESIGNS IN HEALTH TECHNOLOGY ASSESSMENT: LESSONS FROM TWO SYSTEMATIC REVIEWS OF BEHAVIOR CHANGE STRATEGIES

Craig R. Ramsay

University of Aberdeen

Lloyd Matowe

Kuwait University

Roberto Grilli

Agenzia Sanitaria Regionale

Jeremy M. Grimshaw

University of Ottawa

Ruth E. Thomas

University of Aberdeen

Abstract

Objectives: In an interrupted time series (ITS) design, data are collected at multiple instances over time before and after an intervention to detect whether the intervention has an effect significantly greater than the underlying secular trend. We critically reviewed the methodological quality of ITS designs using studies included in two systematic reviews (a review of mass media interventions and a review of guideline dissemination and implementation strategies).

Methods: Quality criteria were developed, and data were abstracted from each study. If the primary study analyzed the ITS design inappropriately, we reanalyzed the results by using time series regression.

Results: Twenty mass media studies and thirty-eight guideline studies were included. A total of 66% of ITS studies did not rule out the threat that another event could have occurred at the point of intervention. Thirty-three studies were reanalyzed, of which eight had significant preintervention trends. All of the studies were considered “effective” in the original report, but approximately half of the reanalyzed studies showed no statistically significant differences.

Conclusions: We demonstrated that ITS designs are often analyzed inappropriately, underpowered, and poorly reported in implementation research. We have illustrated a framework for appraising ITS designs, and more widespread adoption of this framework would strengthen reviews that use ITS designs.

Keywords: Professional practice, Cross-sectional studies, Regression analysis, Time factors

The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Executive Health Department. The guidelines dissemination and implementation review was funded by the UK NHS R&D Health Technology Assessment Programme. However, the views expressed are those of the authors and not necessarily those of the funding body.

Reliable and valid information on the effectiveness and cost-effectiveness of different interventions is required if policy makers are to make decisions based on these interventions. Rigorous research can help provide such information for policy makers; however, many existing studies are of poor quality (6).

The randomized controlled trial (RCT) is seen as the criterion standard methodology for the evaluation of health care interventions (17). However, there are many interventions for which it is impossible or impractical to use RCTs. For national interventions such as mass media campaigns (35) or within local contexts, the policy makers may wish to assess the impact of specific policies such as the dissemination of clinical guidelines (49). Thus, there is increasing interest in the use of quasi-experimental designs (18).

A common quasi-experimental design used to evaluate such interventions is a before-and-after study design. That is, some measure of compliance or outcome is taken before the intervention and then the same measure is taken after the intervention has occurred. Any changes are then inferred to have occurred because of the intervention. For example, the Working Party of the Royal College of Radiologists undertook a before-and-after study evaluating the impact on general practice of disseminating the Royal College of Radiologists booklet *Making the Best Use of a Department of Radiology* to general practitioners (58). The booklet provided guidelines for radiological investigations and was introduced on January 1, 1990. In the study, the number and type of referrals were measured for the year before and the year after the introduction of the guidelines. There was a reduction in radiological requests after the guidelines were introduced, and it was concluded that the guidelines had changed practice. Several questions remained unanswered. Was the number of referrals already decreasing before the intervention? Did the major National Health Service reforms introduced during 1989–90 impact upon the number of referrals (23;60)? Did the referrals stay at this lower level throughout 1990, or did they begin to return to the original level? These questions demonstrate the potential weaknesses of the before-and-after design (49).

Interrupted time series (ITS) designs can be used to strengthen before-and-after designs (15). In an ITS design, data are collected at multiple instances over time before and after an intervention (interruption) is introduced to detect whether the intervention has an effect significantly greater than the underlying secular trend. There are many examples of the use of this design for the evaluation of health care interventions (16;22;45;67).

An advantage of an ITS design is that it allows for the statistical investigation of potential biases in the estimate of the effect of the intervention. These potential biases include:

- **Secular trend** – the outcome may be increasing or decreasing with time. For example, the observations might be increasing before the intervention; hence, one could have wrongly attributed the observed effect to the intervention if a before-and-after study was performed.
- **Cyclical or seasonal effects** – there may be cyclical patterns in the outcome that occur over time.
- **Duration of the intervention** – the intervention might have an effect for the first three months only after it was introduced; data collected yearly would not have identified this effect.
- **Random fluctuations** – these are short fluctuations with no discernible pattern that can bias intervention effect estimates.
- **Autocorrelation** – this is the extent to which data collected close together in time are correlated with each other. The autocorrelation ranges between -1 and 1 . A negative autocorrelation suggests that outcomes taken close together in time are likely to be dissimilar. For example, a high outcome is followed by a low outcome that is then followed by a high outcome and so on. In contrast, a positive autocorrelation suggests that outcomes measured close together in time are similar to each other. For example, a high outcome is followed by another high outcome. Ignoring autocorrelation can lead to spuriously significant effects (18).

To investigate such biases, researchers need to use appropriate statistical methods. For example, time series regression techniques (25) or autoregressive integrated moving average models (ARIMA) (10).

Studies using ITS designs are increasingly being considered for inclusion in systematic reviews (7); however, there has been no published research to assess the methodological quality of ITS studies. We, therefore, undertook a critical review of the methodological quality of ITS studies included in two systematic reviews. The first review evaluated the effectiveness of mass media interventions targeted at improving the utilization of health services (35) and the second review evaluated the effectiveness of different clinical guideline dissemination and implementation strategies (36).

METHODS

Inclusion Criteria

All ITS studies in the mass media review were included. One study from the guideline review was excluded, because it was also in the mass media review (48). Both reviews used the Effective Practice and Organisation of Care (EPOC) Cochrane Group definition of an ITS design (7): (i) there were at least three time points before and after the intervention, irrespective of the statistical analysis used; (ii) the intervention occurred at a clearly defined point in time; (iii) the study measured provider performance or patient outcome objectively.

Data Abstraction

Quality criteria were developed from two sources; from the EPOC data collection checklist (5), and from the classification of threats to validity identified by Campbell and Stanley (15). These quality criteria were agreed by consensus among all the authors in the study and are displayed in Box 1. At least two reviewers independently abstracted the data from each study.

Data Analysis

Descriptive statistics for each review were tabulated. If studies had not performed any statistical analysis or had used inappropriate statistical methods, we reanalyzed the results (where possible). For the purpose of reanalysis, data on individual observations over time were derived from tables of results or graphs presented in the original study. Data from graphs was obtained by scanning the graph as a digital image and reading the corresponding values from the images (35).

Statistical Method for Reanalysis of Primary Studies

Time series regression was used to reanalyze the results from each study. The best fit preintervention and postintervention lines were estimated by using linear regression, and autocorrelation was adjusted for by using the maximum likelihood methods where appropriate (25). First-order autocorrelation was tested for statistically by using the Durbin-Watson statistic, and higher-order autocorrelations were investigated by using the autocorrelation and partial autocorrelation function.

Two effect sizes were estimated (see Figure 1). First, a change in the level of outcome at the first point after the introduction of the intervention was estimated. This strategy was performed by extrapolating the preintervention regression line to the first point postintervention. The difference between this extrapolated point and the postintervention regression estimate for the same point estimated the change in level. Further mathematical details are available from the authors on request. Second, a change in the slopes of the regression lines was estimated (calculated as postintervention minus preintervention slope). We

Box 1. Quality Criteria for ITS Designs**1. Intervention occurred independently of other changes over time**

- DONE The intervention occurred independently of other changes over time
- NOT CLEAR Not specified (will be treated as NOT DONE if information cannot be obtained from the authors)
- NOT DONE Reported that intervention was not independent of other changes in time

2. Intervention was unlikely to affect data collection

- DONE Reported that intervention itself was unlikely to affect data collection (for example, sources and methods of data collection were the same before and after the intervention)
- NOT CLEAR Not specified (treated as NOT DONE if information cannot be obtained from the authors)
- NOT DONE Intervention itself was likely to affect data collection (for example, any change in source or method of data collection reported)

3. The primary outcome was assessed blindly or was measured objectively

- DONE Stated explicitly that primary outcome variables were assessed blindly *or* outcome variables are objective e.g., length of hospital stay, drug levels assessed by a standardized test
- NOT CLEAR Not specified (treated as NOT DONE if information cannot be obtained from the authors)
- NOT DONE Outcomes were not assessed blindly

4. The primary outcome was reliable or was measured objectively

- DONE Two or more raters with agreement $\geq 90\%$ or kappa ≥ 0.8 *or* outcome assessment is objective, e.g., length of hospital stay, drug levels assessed by a standardized test
- NOT CLEAR Reliability not reported for outcome measures obtained by chart extraction or collected by an Individual (will be treated as NOT DONE if information cannot be obtained from the authors)
- NOT DONE Two or more raters with agreement $< 90\%$ or kappa < 0.8

5. The composition of the data set at each time point covered at least 80 % of the total number of participants in the study

- DONE Data set covers 80–100% of total number of participants or episodes of care in the study
- NOT CLEAR Not specified (will be treated as NOT DONE if information cannot be obtained from the authors)
- NOT DONE Data set covers less than 80% of the total number of participants or episodes of care in the study

6. The shape of the intervention effect was prespecified

- DONE A rational explanation for the shape of intervention effect was given by the author(s)
- NOT CLEAR Not specified
- NOT DONE Any of the conditions above are not met

7. A rationale for the number and spacing of data points was described

- DONE Rationale for the number of points stated (e.g., monthly data for 12 months postintervention was used because the anticipated effect was expected to decay) *or* sample size calculation performed
- NOT CLEAR Not specified
- NOT DONE Any of the conditions above are not met

8. The study was analyzed appropriately using time series techniques

- DONE ARIMA models were used *or* time series regression models were used to analyze the data and serial correlation was adjusted/tested for
- NOT CLEAR Not specified
- NOT DONE Any of the conditions above are not met

ITS, interrupted time series; ARIMA, autoregressive integrated moving average.

classified each study intervention as “significant” if either of these effect sizes were statistically significant.

RESULTS

Twenty ITS studies were included from the mass media review (8;9;11;18;24;27;39–42;46–48;51;53;55;61;64;69;71), and thirty-eight ITS studies were included from the guidelines

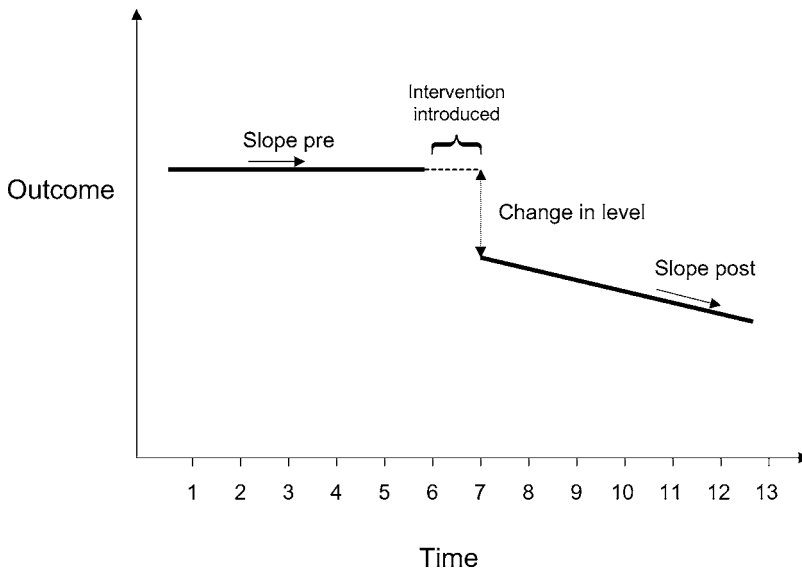


Figure 1. The effect sizes estimated by time series regression analysis of an interrupted time series design.

review (1–4;12–14;16;22;26;28–33;37;38;43–45;50;52;54;56;57;59;62;63;65–68;70; 72–75). Eighteen studies were published before 1990 and forty studies after 1990. The characteristics of the ITS studies are displayed in Table 1. Both reviews had similar numbers of preintervention data points; however, the average ratio of postintervention points to preintervention points indicated that the guideline studies tended to collect more postintervention points than preintervention points within each study. The time interval between data points varied, with “monthly” the most common in both reviews.

The quality criteria for the ITS studies are shown in Table 2. Thirty-eight (66%) studies did not rule out the threat that another event could have occurred at the same time as the intervention. Reporting of factors related to data collection, the primary outcome, and completeness of the data set were generally done in both reviews. No study provided a justification for the number of data points used or a rationale for the shape of the intervention effect.

Table 1. Characteristics of Included Interrupted Time Series Studies

	Mass media review	Guidelines review
Number of included studies	20	38
Median number of preintervention points	9	10
Median number of postintervention points	6	12
Ratio of postintervention points to preintervention points – mean (SD)	0.9 (0.8)	1.9 (2.0)
<i>Time interval between points</i>		
5 days	0	1
1 week	3	3
1 month	9	25
2 months	3	0
3 months	1	5
1 year	4	4

Table 2. Quality Criteria for ITS Studies

Threat to validity	Mass media review (n = 20)	Guidelines review (n = 38)
<i>Intervention independent of other changes</i>		
Done	7	13
Not clear	11	24
Not done	2	1
<i>Intervention unlikely to affect data collection</i>		
Done	19	35
Not clear	1	3
Not done	0	0
<i>Blinded assessment of primary outcome</i>		
Done	17	23
Not clear	3	15
Not done	0	0
<i>Reliable primary outcome measure</i>		
Done	18	24
Not clear	2	14
Not done	0	0
<i>Completeness of data set</i>		
Done	16	37
Not clear	4	0
Not done	0	1
<i>Shape of intervention effect specified</i>		
Done	0	0
Not clear	20	38
Not done	0	0
<i>Rationale for sample size</i>		
Done	0	0
Not clear	0	0
Not done	20	38
<i>Data analyzed appropriately</i>		
Done	5	16
Not clear	0	0
Not done	15	22

Thirty-seven of the ITS studies in the reviews were analyzed inappropriately, of which thirty-three could be reanalyzed. Four “inappropriately analyzed” studies could not be reanalyzed because of the poor quality of the graphs in the primary papers. The results of the reanalyzed studies are shown in Table 3. Eight of the studies had statistically significant preintervention trends (3;11;21;30;52;57;61;75). The mean autocorrelations in the reanalyzed ITS studies were small. All of the interventions in each of the reanalyzed studies were termed “effective” in their original reports, but approximately half of these studies showed no statistically significant differences in slope and level when reanalyzed.

Table 3. Characteristics of Reanalyzed Studies

	Mass media review	Guidelines review
Number of reanalyzed studies	15	18
Significant preintervention trend	3 (20%)	5 (28%)
Autocorrelation coefficient, mean (SD)	-0.18 (0.32)	0.06 (0.40)
Original result significant	15 (100%)	19 (100%)
Original result overturned	7 (47%)	8 (44%)

DISCUSSION

We have demonstrated that ITS designs are being increasingly used to evaluate health care behavior change interventions. Many of these studies had apparent threats to their internal validity that made interpretation of the results unreliable. The design of the studies was often poorly reported and analyzed inappropriately leading to misleading conclusions in the published reports of these studies.

There were no substantial differences between the characteristics of the studies in the two reviews. The mass media studies tended to have smaller postintervention phases. Most of the studies in both reviews had short time series. Short time series suggest that model fitting is performed with less confidence, standard errors are increased, type I error is increased, power is reduced, and hence, failure to detect autocorrelation or secular trends.

Over 65% of the studies were analyzed inappropriately. Most of these studies performed a *t*-test of the preintervention points versus postintervention points. The *t*-test would give incorrect effect sizes if preintervention trends were present and would decrease the standard error if positive autocorrelation were present (19;49). The *t*-test, therefore, should not be used to analyze results from an ITS design.

Many of the studies were underpowered. It is difficult to derive a formula for the sample size (34). As a rule of thumb, if one collects ten pre- and ten post-data points, then the study would have at least 80% power to detect a change in level of five *standard deviations* (of the pre-data) only if the autocorrelation is greater than 0.4 (20). This score is a large effect, and the results from the two reviews suggested that the autocorrelation was around the order of 0.1; therefore, a study with ten pre- and post-data points was likely to be underpowered. In addition, it is beneficial to have a long preintervention phase thereby increasing power to detect secular trends.

This study showed that many ITS studies had analytical shortfalls. In such instances, reviewers may need to consider reanalyzing. The data may be obtained directly from the authors, but if this is not possible, the reviewers can often scan the graphs from the original paper onto a computer as described in the methods section. We recommend reviewers use time series regression techniques for reanalyzing ITS studies. The main advantages of this technique is, first, that it can be performed using most standard statistical packages and, second, that it can estimate the effect sizes more precisely than other time series techniques when the series are short (20;34).

The lack of a randomized control group in an ITS design requires the investigators to be critical of the internal validity of the study. With this in mind, we developed a checklist for reviewers of such designs. This checklist encompassed the major threats to validity of ITS designs as proposed by Campbell and Stanley (15). Generally, the greatest threat to validity is that an event other than the control of the researchers occurred at the same time as the intervention, thereby making causal inferences impossible. It was disappointing, therefore, to note that this threat was not ruled out explicitly in 66% of the studies.

Implications for Researchers

ITS designs may appear superficially simple; however, we have demonstrated that they are often analyzed inappropriately, underpowered, and poorly reported in the implementation research field. To improve the quality of analysis and reporting, we suggest that researchers consider the list of quality criteria described in Box 1.

Implications for Systematic Reviewers

ITS designs are a useful and pragmatic evaluative tool when RCTs are not feasible. It is possible to consider this often neglected study design in systematic reviews and in so doing draw both qualitative and quantitative results on intervention effects. In this study, we have

illustrated the usefulness of a framework for appraising ITS designs, and more widespread adoption of such a framework would strengthen reviews that use ITS designs.

Implications for Policy Makers

Policy makers may find ITS designs a useful way to assess the impact of specific policies that could remain unassessable otherwise. In addition, these designs can often be performed inexpensively from routine data collection.

REFERENCES

1. Avorn J, Soumerai SB, Taylor W, et al. Reduction of incorrect antibiotic dosing through a structured educational order form. *Arch Intern Med*. 1988;148:1720-1724.
2. Bareford D, Hayling A. Inappropriate use of laboratory services: Long term combined approach to modify request patterns. *BMJ*. 1990;301:1305-1307.
3. Barnett GO, Winickoff R, Dorsey JL, Morgan MM, Lurie RS. Quality assurance through automated monitoring and concurrent feedback using a computer-based medical information system. *Med Care*. 1978;16:962-970.
4. Berbatis CG, Maher MJ, Plumridge RJ, Stoelwinder JU, Zubrick SR. Impact of a drug bulletin on prescribing oral analgesics in a teaching hospital. *Am J Hosp Pharm*. 1982;39:98-100.
5. Bero L, Freemantle N, Grilli R, et al. Checklist for data abstraction in EPOC reviews. Aberdeen: EPOC(mimeo); 1997.
6. Bero L, Grilli R, Grimshaw JM, et al. Closing the gap between research and practice: An overview of systematic reviews of interventions to promote implementation findings by health care professionals. *BMJ*. 1998;317:465-468.
7. Bero L, Grilli R, Grimshaw JM, et al. The Cochrane Effective Practice and Organisation of care Group (EPOC) Module. In: *The Cochrane Library*. Oxford: Update Software, 2002 Issue 1.
8. Blohm M, Herlitz J, Hartford M, et al. Consequences of a media campaign focusing on delay in acute myocardial infarction. *Am J Cardiol*. 1992;69:411-413.
9. Bonerandi JJ, Grob JJ, Cnudde N, Enel P, Gouvernet J. Campaign of early detection of melanoma in the Provence-Alpes-Cote-d'Azur area 1989. Lessons of an experience. *Ann Dermatol Venerol*. 1992;119:105-109.
10. Box EP, Jenkins GM. *Time series analysis*. London: Holden-Day; 1976.
11. Brasca A, Pigliacampo J, Pollastri E, et al. An awareness campaign on the diagnosis and prevention of colon and rectal cancer. Planning and evaluation of its results. *Acta Gastroenterol Latinoam*. 1987;17:325-329.
12. Brook RH, Williams KN. Effect of medical care review on the use of injections: A study of the New Mexico Experimental Medical Care Review Organization. *Ann Intern Med*. 1976;85:509-515.
13. Brooks RJ. Implementing guidelines for eye care of diabetic patients: Results from an HMO intervention study. *Am J Managed Care*. 1996;2:365-369.
14. Brufsky JW, Ross-Degnan D, Calabrese D, Gao X, Soumerai SB. Shifting physician prescribing to a preferred histamine-2-receptor antagonist. Effects of a multifactorial intervention in a mixed-model health maintenance organization. *Med. Care* 1998;36:321-332.
15. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally; 1966.
16. Clarke JA, Adams JE. The application of clinical guidelines for skull radiography in the accident and emergency department: Theory and practice. *Clin Radiol*. 1990;41:152-5.
17. Cochrane AL. Effectiveness and efficiency: Random reflections on health services. London: British Medical Journal for Nuffield Provincial Health Trust; 1989.
18. Cook TD, Campbell DT. *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin Company; 1979.
19. Crosbie J. Interrupted time-series analysis with brief single-subject data. *J Consult Clin Psychol*. 1993;61:966-974.
20. Crosbie J. Interrupted time-series analysis with short series; why it is problematic; how it can be improved. In: Gottman JM, ed. *The analysis of change*, New Jersey: Lawrence Erlbaum Associates; 1995.

21. Del Mar CB, Green AC, Battistutta D. Do public media campaigns designed to increase skin cancer awareness result in increased skin excision rates? *Aust N Z J Public Health*. 1997;21:751-754.
22. Dempsey CL. Nursing home-acquired pneumonia: Outcomes from a clinical process improvement program. *Pharmacotherapy*. 1995;15:33S-38S.
23. Department of Health and the Welsh Office. *General practice in the National Health Service: A new contract*. London: HMSO; 1989.
24. Domenighetti G, Luraschi P, Casabianca A, et al. Effect of information campaign by the mass media on hysterectomy rates. *Lancet*. 1988;2:1470-1473.
25. Draper N, Smith H. *Applied regression analysis*. New York: Wiley; 1981.
26. Elam K, Taylor V, Ciol MA, Franklin GM, Deyo RA. Impact of a worker's compensation practice guideline on lumbar spine fusion in Washington State. *Med Care*. 1997;35:417-424.
27. Eppler E, Eisenberg MS, Schaeffer S, Meischke H, Larson MP. 911 and emergency department use for chest pain: Results of a media campaign. *Ann Emerg Med*. 1994;24:202-208.
28. Everitt DE, Soumerai SB, Avorn J, Klapholz H, Wessels M. Changing surgical antimicrobial prophylaxis practices through education targeted at senior department leaders. *Infect Control Hosp Epidemiol*. 1990;11:578-583.
29. Fowkes FG, Evans RC, Williams LA, et al. Implementation of guidelines for the use of skull radiographs in patients with head injuries. *Lancet*. 1984;2:795-796.
30. Fowkes FG, Hall R, Jones JH, et al. Trial of strategy for reducing the use of laboratory tests. *BMJ Clin Res Ed*. 1986;292:883-885.
31. Fraser GL, Wennberg DE, Dickens JD Jr, Lambrew CT. Changing physician behavior in ordering digoxin assays. *Ann Pharmacother* 1996;30:449-454.
32. Gama R, Nightingale PG, Ratcliffe JG. Effect of educational feedback on clinicians' requesting of cardiac enzymes. *Ann Clin Biochem*. 1992;29:224-225.
33. Gortmaker SL, Bickford AF, Mathewson HO, Dumbaugh K, Tirrell PC. A successful experiment to reduce unnecessary laboratory use in a community hospital. *Med Care*. 1988;26:631-642.
34. Gottman JM. Time series analysis: A comprehensive introduction for social scientists. Cambridge, England: Cambridge University Press; 1981.
35. Grilli R, Ramsay CR, Minozzi S. Mass media interventions: Effects on health services utilisation. In: *The Cochrane Library*. Oxford: Update Software, 2002 Issue 1.
36. Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* (in press).
37. Gurwitz JH, Noonan JP, Soumerai SB. Reducing the use of H₂-receptor antagonists in the long-term-care setting. *J Am Geriatr Soc*. 1992;40:359-364.
38. Hammond KW, Snowden M, Risse SC, Adkins TG, O'Brien JJ. An effective computer-based tardive dyskinesia monitoring system. *Am J Med Qual*. 1995;10:133-137.
39. Harris DW, Karandikar DS, Spencer MG, et al. Returned-medicines campaign in Birmingham, 1977. *Lancet*. 1979;1:599-601.
40. Healsmith MF, Graham-Brown RA, Osborne JE, London SP, Fletcher A. Further experience of public education for the early diagnosis of malignant melanoma in Leicestershire. *Clin Exp Dermatol*. 1993;18:396-400.
41. Herd RM, Cooper EJ, Hunter JA, et al. Cutaneous malignant melanoma. Publicity, screening clinics and survival – the Edinburgh experience 1982-90. *Br J Dermatol*. 1995;132:563-570.
42. Joshi UY, Cameron SO, Sommerville JM, Sommerville RG. HIV testing in Glasgow Genito-Urinary Medicine Clinics 1985-1987. *Scott Med J*. 1988;33:294-295.
43. Kong KM, Johnson WF, Anderson PO, Sato LA, Simonian RI. Effect of a pharmacy-based cost-containment program on cimetidine costs and use. *Am J Hosp Pharm*. 1987;44:2068-2073.
44. Legorreta AP, Hasan MM, Peters AL, Pelletier KR, Leung KM. An intervention for enhancing compliance with screening recommendations for diabetic retinopathy. A bicoastal experience. *Diabetes Care* 1997;20:520-523.
45. Lomas J, Anderson GM, Domnick-Pierre K, et al. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med*. 1989;321:1306-1311.

46. Lowe JB, Balanda KP, Del Mar CB, Purdie D, Hilsdon AM. General practitioner and patient response during a public education program to encourage skin examinations. *Med J Aust.* 1994;161:195-198.
47. Macdonald H, Roder D. The planning, implementation and evaluation of an immunization promotion campaign in South Australia. *Hygie.* 1985;4:13-17.
48. Maclure M, Dormuth C, Naumann T, et al. Influences of educational interventions and adverse news about calcium-channel blockers on first-line prescribing of antihypertensive drugs to elderly people in British Columbia. *Lancet.* 1998;352:943-948.
49. Matowe L, Ramsay CR, Grimshaw JM, Gilbert FI, Macleod MJ, Needham G. Effects of mailed dissemination of the Royal College of Radiologists' guidelines on general practitioner referrals for radiography: A time series analysis. *Clin Radiol.* 2002;57:575-578.
50. Morrison JC, Sumrall DD, Chevalier SP, et al. The effect of provider education on blood utilization practices. *Am J Obstet Gynecol* 1993;169:1240-1245.
51. Nattinger AB, Hoffmann RG, Howell-Pelz A, Goodwin JS. Effect of Nancy Reagan's mastectomy on choice of surgery for breast cancer by US women. *JAMA.* 1998;279:762-766.
52. Novich M, Gillis L, Tauber AI. The laboratory test justified. An effective means to reduce routine laboratory testing. *Am J Clin Pathol.* 1985;84:756-759.
53. Paunio M, Virtanen M, Peltola H, et al. Increase of vaccination coverage by mass media and individual approach: Intensified measles, mumps, and rubella prevention program in Finland. *Am J Epidemiol.* 1991;133:1152-1160.
54. Pearce MJ, Begg EJ. Encouraging consensus cost effective drug therapy: Five years experience with a hospital drug utilisation review programme. *N Z Med J.* 1997;110:92-95.
55. Pehamberger H, Binder M, Knollmayer S, Wolff K. Immediate effects of a public education campaign on prognostic features of melanoma. *J Am Acad Dermatol.* 1993;29:106-109.
56. Poma PA. Effect of departmental policies on cesarean delivery rates: A community hospital experience. *Obstet Gynecol* 1998;91:1013-1018.
57. Ratnaike S, Hunt D, Eilermann LJ, Hazen R, Deam D. The investigation of chest pain: Audit and intervention. *Med J Aust.* 1993;159:666-671.
58. Royal College of Radiologists Working Party. Influence of Royal College of Radiologists' guidelines on referral from general practice. *BMJ.* 1993;306:110-111.
59. Santerre RE. The effect of the ACOG guideline on vaginal births after cesarean. *Med Care Res Rev.* 1996;53:315-329.
60. Secretaries of State for Social Services WNIaS. *Working for patients.* London: HMSO; 1989.
61. Shelley JM, Irwig LM, Simpson JM, Makaskill P. Evaluation of a mass-media-led campaign to increase Pap smear screening. *Health Educ Res.* 1991;6:267-277.
62. Sherman CR, Potosky AL, Weis KA, Ferguson JH. The Consensus Development Program. Detecting changes in medical practice following a consensus conference on the treatment of prostate cancer. *Int J Technol Assess Health Care.* 1992;8:683-693.
63. Shorr RI, Fought RL, Ray WA. Changes in antipsychotic drug use in nursing homes during implementation of the OBRA-87 regulations. *JAMA.* 1994;271:358-362.
64. Soumerai SB, Ross-Degnan D, Kahn JS. Effects of professional and media warnings about the association between aspirin use in children and Reye's syndrome. *Milbank Q.* 1992;70:155-182.
65. Soumerai SB, Avorn J, Gortmaker S, Hawley S. Effect of government and commercial warnings on reducing prescription misuse: The case of propoxyphene. *Am J Public Health.* 1987;77:1518-1523.
66. Stuart ME, Macuiba J, Heidrich F, et al. Successful implementation of an evidence-based clinical practice guideline: Acute dysuria/urgency in adult women. *HMO Pract.* 1997;11:150-157.
67. Studnicki J, Remmel R, Campbell R, Werner DC. The impact of legislatively imposed practice guidelines on cesarean section rates: The Florida experience. *Am J Med Qual.* 1997;12:62-68.
68. Suwangoon P, Moola-Or P, Waiwatana A, et al. Effect of a selective restriction policy on antibiotic expenditure and use: An institutional model. *J Med Assoc Thai.* 1991;74:272-275.
69. Tesoriero JM, Sorin MD. The effect of 'Magic' Johnson's HIV disclosure on anonymous HIV counselling and testing services in New York State. *AIDS Public Policy J.* 1992;7:216-224.

70. Thamer M, Ray NF, Henderson SC, et al. Influence of the NIH Consensus Conference on *Helicobacter pylori* on physician prescribing among a Medicaid population. *Med Care*. 1998;36:646-660.
71. Turner GC, Mutton KJ. HIV testing: Changing trends. *BMJ*. 1987;295:502.
72. van Walraven C, Goel V, Chan B. Effect of population-based interventions on laboratory utilization: A time-series analysis. *JAMA*. 1998;280:2028-2033.
73. Vincent EC, Hardin PA, Norman LA, Lester EA, Stinton SH. The effects of a computer-assisted reminder system on patient compliance with recommended health maintenance procedures. *Proc Ann Symp Comput Appl Med Care*. 1995;656-660.
74. Wong ET, McCarron MM, Shaw ST Jr. Ordering of laboratory tests in a teaching hospital. Can it be improved? *JAMA*. 1983;249:3076-3080.
75. Zehr KJ, Dawson PB, Yang SC, Heitmiller RF. Standardized clinical care pathways for major thoracic cases reduce hospital costs. *Ann Thorac Surg*. 1998;66:914-919.