



## ORIGINAL ARTICLE

## A landmark randomized health care trial: the Burlington trial of the nurse practitioner

David L. Sackett\*

*Trout Research and Education Center, Hamilton, Ontario, Canada*

Accepted 10 January 2009

From July, 1971, to July, 1972, in a large suburban Ontario practice of two family physicians, a randomized controlled trial was conducted to assess the effects of substituting nurse practitioners for physicians in primary-care practice. Before and after the trial, the health status of patients who received conventional care from family physicians was compared with the status of those who received care mainly from nurse practitioners. Both groups of patients had a similar mortality experience, and no differences were found in physical functional capacity, social function or emotional function. The quality of care rendered to the two groups seemed similar, as assessed by a quantitative ‘indicator-condition’ approach. Satisfaction was high among both patients and professional personnel. Although cost effective from society’s point of view, the new method of primary care was not financially profitable to doctors because of current restrictions on reimbursement for the nurse-practitioner services [1].

The year was 1970. The Americans invaded Cambodia; Biafra capitulated; the Beatles stopped and Queen started; the US Food and drug Administration (FDA) approved lithium for manic depression and warned that birth control pills might cause blood clots; a damaged Apollo 13 landed safely; Edward Heath became the UK prime minister and Salvador Allende the Chilean president; Charles de Gaulle, Bertrand Russell, Janis Joplin, and Jimi Hendrix died and Queen Latifah, Andre Agassi, Alan Shearer, and Matthew Pinsent were born; Quebec separatists kidnapped and killed a cabinet minister; Baader escaped jail with help from Meinhof; Alexander Solzhenitsyn won a Nobel Prize; Ohio National Guardsmen killed four students at Kent State University; 28 thalidomide victims were compensated; child-proof medicine caps were introduced; a 1st class US postage stamp cost 6 cents; the US median household income (in 2007 dollar value) was \$8,734, and Mick Jagger

was fined 200 pounds (in 1970 sterling value) for possession of cannabis; the personal computer (PC) would not be introduced for another decade; and the editors of this journal were 19 and 26 years old.

That year, although half of Ontario physicians were fee-for-service family doctors (1 for every 1,723 souls), already some of them were swamped and could not accept new patients; however, well-trained and highly experienced nurses were abundant. Indeed, Ian Hay and Pat Sweeney, nonuniversity family physicians in the Hamilton, Ontario suburb of Burlington, thought so highly of their practice nurses, Georgie Lefroy and Isabel Vandervlist, that they asked themselves whether, with a little additional training, their nurses could assume a substantial portion of their responsibilities for primary care. Uncertain about the answer to this question, they asked the fledgling Department of Clinical Epidemiology and Biostatistics at McMaster University for help.

Their request coincided with the arrival of Walter Spitzer after his year of study with Alvan Feinstein at Yale. Nobody in the infant department had ever performed a randomized health care trial (only three had been carried out anywhere), and Walter’s primary care background, recent methodological training, and boundless enthusiasm resulted in a spectacular long shot: he became the Principal Methodological Investigator of the Burlington trial.

Walter immediately led us to confront the conflict between high-order methodology and rational, ethical clinical practice. On the one hand, it would have been scientifically purest to randomize individual patients to receive all their primary care from either a nurse practitioner or a family doctor. On the other hand, for many health problems (progressively fatal disease, emotional illness, and others), the rational, ethical care for the individual patient had to extend to other family members. Furthermore, some health problems would require skills and facilities beyond those that could be provided by family physicians, much less nurse practitioners.

A major intervention of the nurse practitioners in this trial was to be their exercise of clinical judgment. With his typical energy and enthusiasm, Walter collaborated with

\* Corresponding author. Trout Research and Education Center, 303-28, Duke Street, Hamilton, Ontario L8P 1X1, Canada. Tel.: +1-289-396-1502.  
E-mail address: sackett@bmts.com

colleagues in the McMaster School of Nursing in creating a practical (1-year, predominantly part-time) program for their education and training [2].

Cluster randomization was virtually unheard of at the time, and Alan Donner's empirical study of this allocation strategy would not be published for another 12 years [3]. Nonetheless, Walter convinced us that we must make the household family the unit of randomization. Furthermore, anticipating the "adaptive treatment design" strategy that appeared decades later, he brought us to the consensus that the intervention whose outcome we would compare had to be the supremely pragmatic, entire course of care (including specialist and hospital care) initiated by the nurse practitioner or family physician. Finally, nurse practitioner-initiated care had to include using her associated family physician as an intermediate consultant in caring for her patients.

On the one hand, the processes and outcomes that would tell us whether nurse practitioners were as effective as family physicians were clear to all of us. They comprised clinical judgment, clinical outcomes, physical, social, and emotional function, quality of care, satisfaction, and cost-effectiveness. But on the other hand, in 1970, nobody knew how to measure any of them except death (and we anticipated very few deaths to be informative). If we were to study these processes and outcomes, we had to create measures for them ourselves. Moreover, somebody had to recruit and train a team of research assistants, interviewers, and data coders to capture and analyze them.

Our department had just started weekly "Work in Progress" seminars, where we would brainstorm strategies and tactics for overcoming the methodological challenges of our research projects, and the Burlington trial was a frequent topic presented. Walter made frequent use of them and applied his boundless enthusiasm, creativity, energy, and hard work that caused the success not only of this enterprise, but also of its associated Field Survey Unit, a subsequent randomized trial in 14 family practices [4], and an economic analysis of the financial consequences of employing nurse practitioners [5].

It was our good fortune that our first and second ever graduate students in the Design, Measurement and Evaluation MSc Program at McMaster, Sandy Macpherson (a psychiatrist who later developed the aging programme at McMaster and was the Medical Officer of Health for Toronto) and Larry Chambers (a political scientist who currently heads the Élisabeth-Bruyère Research Institute, a Bruyère Continuing Care and University of Ottawa partnership), were interested in measuring physical, emotional [6], and social function [7], respectively. They took up the challenge of generating the measures of emotional, social, and physical function that were incorporated into the McMaster Health Index Question (MHIQ) [8,9]. The MHIQ now appears in PROQOLID, a Patient-Reported Outcome and Quality of Life Instruments Database (available at <http://www.proqolid.org>). PROQOLID describes over 540 instruments and is visited by an average of 800 users each day.

The Short Form 36 (SF36) would not be developed for another 20 years; hence they had to start from first principles. With key input from Charles Goldsmith, a biostatistician, and Ronald McAuley, a family physician, they identified the prerequisites now applied widely in developing health indices: comprehensiveness, positive orientation, general applicability, simplicity, acceptability, low cost, precision, and amenability to index construction. In pilot studies, the instruments they developed from these prerequisites proved simple to administer, acceptable to patients (refusal rates were 5%), and both biologically and clinically sensible when applied to patients in changing health states. Accordingly, they were applied both before and after the 1-year experimental period in the Burlington trial.

Walter also was the catalyst in the development of the quality of care measures. At his urging, Jack Sibley (a highly regarded local internist who'd hung out with the clinical epidemiologists) tackled the measurement of the quality of care [10]. Three measurements resulted from this work. First were the recorded opinions of consultants to whom patients were referred by the study clinicians—was the consultation request timely and appropriate, and had care before the referral been adequate. Second was an examination of clinical records triggered by the prescription of any of a list of drugs (chloramphenicol, tetracycline, other antibiotics, amphetamines, multivitamins, hematinics, phenylbutazone, antihypertensives, steroids, vitamin B<sub>12</sub>, antidepressants, tranquilizers, and cardiac glycosides); as might be suspected from an examination of this list, prescribing many of its entries (save under special circumstances) constituted low quality of care. Carbonized, personal prescription pads were provided to our study clinicians, and they were monitored throughout the trial.

To develop the third process measure, and again with encouragement, advice, and occasional pestering from Walter, Jack Sibley recruited a group of family physicians and methodologists who nominated clinical presentations that occurred frequently in primary care and whose outcomes were judged likely to be affected by whether clinicians carried out the appropriate diagnostic tests and treatments (or avoided treatments judged to be harmful). Thus, the key attribute of each of these "indicator conditions" was that its outcome had to depend on what the clinician did for it: only specific clinical acts (which we could measure) would result in good outcomes (many of which would occur too rarely or too late to be informative or even measurable).

We scoured the scant literature on effectiveness (the term "evidence-based" would not even be introduced for another 20 years) and debated our combined clinical experience, and eventually nominated 10 "indicator conditions": otitis media, hypertension, prenatal care, care of the newborn, immunization in the first year of life, depression, urinary tract infection, knee injury, pityriasis, and anemia. We specified what ought and ought not to be done by clinicians when they confronted these conditions, and also