merits. The Health Board cannot pretend that the report does not exist and it may, therefore, not be left, like so many pieces of research, to accumulate dust on someone's desk. We would like to think that the Health Board must take account of its contents in the planning of maternity services.

FROM : ROBERTS, H (ed.) (1990)

Womens Health Counts

Routledge.

Chapter Seven

WHO'S AFRAID OF THE RANDOMIZED CONTROLLED TRIAL?

Some dilemmas of the scientific method and 'good' research practice

ANN OAKLEY

Randomized controlled trials were originally used in agriculture, and their application to man . . . raises practical difficulties and moral dilemmas. (Dawson 1986:1373)

Human beings are not like grains of wheat. (Faulder 1985:72)

This chapter focuses on the nature and uses of the methodology of the randomized controlled trial (RCT) in the light of recent critiques of science, including the feminist concern with the social structure of science as representing an inherently sexist, racist, classist, and culturally coercive practice and form of knowledge. Using the example of one specific RCT aimed at promoting women's health, the chapter outlines some of the dilemmas thus raised for the pursuit of 'good' research practice. The particular viewpoint from which the chapter is written is that of a feminist sociologist who has been responsible for designing and carrying out a randomized trial in the field of pre-natal health care.[1] While the focus of the chapter is on the use of the methodology of random allocation in health research, it is important to note that it has also been used in other areas of experimental research within the social sciences, for instance in psychology in the evaluation of educational interventions (*see* Rapoport 1985), and in the assessment of professional social work services (Fischer 1973; McCord 1982). Although the study discussed in this chapter and some of the other data drawn on are British, the issues highlighted are of general relevance to all communities where importance is attached to the goal of researching and promoting women's health in the broadest sense.

FROM GUINEA PIGS TO THE CAMEL'S NOSE: ORIGINS AND PROBLEMS OF THE RCT AS A TOOL FOR RESEARCHING WOMEN'S HEALTH

The randomized controlled trial as a research method applicable to human subjects is generally said to have been invented in 1946. In that year, a new drug, streptomycin, was thought to cure tuberculosis in guinea pigs, and initial use on human beings suggested therapeutic effectiveness. Because supplies of the new drug were scarce, the Medical Research Council in Britain decided to administer it in the form of a controlled trial, giving it on the basis of selection according to a table of random numbers to some people with tuberculosis and not to others (Silverman 1980).

Behind this apocryphal story of the origins of the randomized controlled trial lies an intermittent history of previous attempts to carry out unbiased comparisons of the effectiveness of different medical treatments. The RCT is essentially an experimental test ('trial') of a particular treatment/approach (or set of treatments/approaches) comparing two or more groups of subjects who are allocated to these groups at random, i.e. according to the play of chance. Conclusions about the effectiveness of treatments based on an RCT rest upon two issues – an assessment of *significance* and judgement about *causation*. Tests of statistical significance are used to determine whether any observed difference between trial groups is due to sampling variability or is evidence of a 'real' difference. If a difference is significant in this sense, then, as Schwartz and colleagues put it in their classic text *Clinical Trials*,

> a judgement of causation allows us to attribute it to the difference between (the) two treatments. This is only possible if the two groups are strictly comparable in all respects apart from the treatments given. Providing two such comparable groups is another statistical problem the correct solution of which is obtained by randomization.
>
> (Schwartz *et al.* 1980:7)

It is important to note that the prerequisite for any RCT is *uncertainty* about the effects of a particular treatment. If something is known to work (and to be acceptable and without harmful effects) then there is no reason to put it to the test in the form of a trial. It is, however, this very issue of certainty/uncertainty that

constitutes one of the central problems of the contemporary debate about RCTs. People can be certain that something (for example, streptomycin, social workers) *is* effective but have no 'real' basis for their certainty; conversely, unless they are able to admit uncertainty, 'real' knowledge can never be gained.

The RCT has been increasingly promoted over the last twenty years as *the* major evaluative tool within medicine. Over the same period a new critical perspective has emerged towards what counts as 'knowledge' and the methods and techniques appropriate to its accumulation. Sources of this critique include the radical science movement (see for example Rose and Rose 1979), the emergence of 'ethnomethodology' within sociology (for example, Goffman 1968) and the broad consensus located within the women's movement about the 'masculinist' orientation of much scientific activity (Harding 1986). The result of these various critiques has been a heightened awareness of the contribution made by different kinds of research strategies to extending human knowledge in the domain both of the 'natural' and the 'social' sciences.

Over the last twenty years, feminists have increasingly criticized the ways in which the construction of what counts as 'knowledge' omits women's perspectives and experiences and is embedded with masculinist values (see for example, Elshtain 1981; Millman and Kanter 1975; Sherman and Beck 1979). The orbit of feminist concern has included science (Keller 1982; Rose 1983). At the same time as the feminist critique has developed, medical science has expanded its control of life in general and of women's lives in particular. This process has highlighted the need to evaluate all interventions claimed to promote health and has given prominence to the role of the RCT. But, although in recent years feminist researchers have taken to task many methodologies both in the natural and 'unnatural' sciences, there has been virtually no discussion to date of this particular, increasingly advocated approach.

The notion of 'feminist' research, as discussed in this chapter, is taken to mean research that relates to an understanding of women's position as that of an oppressed social group, and which adopts a critical perspective towards intellectual traditions rendering women either invisible and/or subject to a priori categorizations of one kind or another. The research process itself

is subject to the same stipulations: that it should not employ methods oppressive either to researchers or to the researched, and should be oriented towards the production of knowledge in such a form and in such a way as can be used by women themselves (Acker *et al.* 1983; Roberts 1981). These strictures are also a formula for 'good' research practice as applied to human subjects in general. However, the practice of feminist research is often located by its advocates on one side of the divide between 'qualitative' and 'quantitative' research methods. Qualitative methods involving in-depth interviewing are seen to be more suited to the exploration of individual experiences – the representation of subjectivity within academic discourse – and to facilitate (in practice if not in theory) a non-hierarchical organization of the research process (Oakley 1981). Conversely, quantitative methods (large-scale surveys, the use of pre-specified scoring methods, for example in personality tests) are cited as instituting the hegemony of the researcher over the researched, and as reducing personal experience to the anonymity of mere numbers. The feminist/masculinist and qualitative/quantitative divisions are paralleled conceptually by a third, that between the physical and the social sciences. As Hedges (1987: 443) has commented: 'Those of us in the social and behavioural sciences know intuitively that there is something "softer" and less cumulative about our research results than those of the physical sciences.'

In terms of this debate about good research practice, the starting point of this chapter is Davies and Esseveld's observation that the problem about the feminist rejection of quantitative methods as necessarily alienating is that it bars discussion both of the ways in which these methods are used, and of those in which they could be used to generate knowledge relevant to the exercise of improving women's situation (Davies and Esseveld 1986: 9). Although feminist research practice requires a critical stance towards existing methodology (the abolition of 'methodolatry' to use Daly's (1973) term), at the same time it has to be recognized that the universe of askable research questions is constrained by the methods allowed. To ban any quantitative (social) science therefore results in a restriction to certain kinds of questions only; this restriction may very well be counter to the same epistemological goal a code of feminist research practice is designed to promote.

According to an Arabic saying deployed by Harris (1974), the problem about letting a camel's nose into your tent is that you are likely then to have to let the whole camel in. The essential question for feminist research posed by the RCT is whether there are benefits of this methodology which can and should be harnessed, without simultaneously dragging into the tent the entire unwieldy superstructure of mixed benefits and hazards (the rest of the camel). Existing published work and the experience drawn on in this chapter suggest that RCTs pose three particular problems for feminist researchers. First and most obviously, there is the principle of *random allocation*, which uses chance – 'the absence of design' (OED) – to determine the treatment received by participants in the research. The extent to which individuals are able to choose the form of their participation in the research is thereby limited. Linked with this is the much debated issue of *informed consent*. What is the meaning of consent, and how much of what kind of information is required by whom? The third problem concerns the epistemology, ownership, and distribution of *certainty*. As already noted, the rationale for undertaking an RCT is uncertainty about the effectiveness/acceptability of a particular procedure. But the professionals may be certain and the lay public not; or the lay public may be convinced about the benefits of a procedure which meets with professional scepticism. It would appear that this issue in particular has provoked a good deal of unclear thinking among those concerned with the promotion of women's health.

Before examining each of these problems in turn, I shall briefly outline the study which highlighted these specific areas of conflict between the practice of *feminist* research on the one hand, and the model of *randomized controlled evaluation*, on the other.

WHO CARES FOR WOMEN? AN RCT OF SOCIAL SUPPORT

The history of the medical care and surveillance of childbearing women is not one of tested and proven effectiveness (Chard and Richards 1977; Wertz and Wertz 1977; Enkin and Chalmers 1982; Oakley 1984; WHO 1986). Studies of how women experience the maternity services have long revealed an iceberg of dissatisfaction, with lack of information, poor communication, long waiting times, and absence of continuity of care coming top of the list (see, for

example, Garcia 1982). The complaints women make about their care resonate with an expanding literature on the importance of social support to the promotion of health (Cohen and Syme 1985). It appears (not surprisingly) that friends are as good or better than the famous apple in keeping the doctor away (Berkman 1984). (This may be one instance of modern scientific knowledge catching up with women's experiental understanding of the world (see Chamberlain 1981).)

For these reasons a study of social support in pregnancy was started at the Thomas Coram Research Unit (TCRU) – part of the University of London – in 1985. The broad aim of the project was to establish whether social support provided as a research intervention has the capacity to make things better for women and their babies. Most previous work on this topic is problematic, because of the repetitive methodological problem that, although better health is generally associated with more support, it is impossible to rule out the explanation that healthier, more supported mothers are different in other ways from less supported, less healthy mothers and babies (Oakley 1985, 1988). Although the better done observational studies make multiple adjustments for confounding variables, still one can only adjust for those variables known to confound; there may be others, equally confounding, of which the researcher is ignorant. For this reason, the TCRU study was planned as an *intervention* study, in which the intervention of providing additional social support would be offered to some women and not to others, and various indices of their experiences, including their health and that of their babies, would be compared at the end of the study. Over a fifteen-month period, a total of 509 women agreed to take part in the study. Random allocation was used to determine who received the intervention, and social support was given by four research midwives who visited women at home during pregnancy, offered a listening ear for individual problems, provided various forms of practical and emotional help when required, and were available 24 hours a day to be contacted in case of need. The Department of Health who funded the study were keen for us to have midwives, rather than any other professional or lay group, giving social support because of the study's possible policy implications. The Department expressed the view that, were the study to be successful in demonstrating the clinical effectiveness of social support, the intervention used

should be one that related to existing maternity care provisions. In order to increase our chances of detecting an effect of social support, we specified that the women needed to have given birth to a small baby in the past and thus constitute a 'high risk' group. The theory behind this was that women with problematic medical histories would be more likely than those without to benefit from extra support. Additionally, use of this criterion would lead to a largely working-class sample (as two-thirds of low birthweight (LBW) babies but only half of all babies in Britain are born in working-class households), and this concentration of social disadvantage might also result in higher benefit. The 'effectiveness' of this social support intervention in terms of a range of outcomes, including women's satisfaction and infant birthweight, was evaluated after delivery, using obstetric case-note information from the four hospitals where the study was done, and by sending all the women a long and detailed postal questionnaire.

Methods used in this study fit more closely within the medical model of controlled evaluation of therapeutic strategies, rather than with the social science model of qualitative research, in which in-depth interviewing is used to build up interpretative accounts of social processes. However, the study began life as a desire to test the idea that in-depth social science interviewing can in itself have a supportive effect for those interviewed (Oakley 1981; Finch 1984). The midwives giving social support also carried out semi-structured interviews with the women in their intervention group; these interviews were partially tape-recorded in order to enable some qualitative analysis of women's experiences.

CHANCE OR CAUSATION? THE ROLE OF RANDOM NUMBERS

The first of the three problems referred to earlier in combining a feminist research consciousness with the technique of an RCT concerns the process of random allocation itself. We had some interesting and some disturbing difficulties with this. Before discussing these, it is worth considering the history of randomization as a research technique. According to Silverman:

> The central question in the study of living things is how to decide whether an observed event is to be attributed to the

meaningless play of chance, on the one hand, or to causation . . . on the other.

(Silverman 1980)

Until the 1920s, scientists were not able to overcome the problem that very long series of observations were needed in order to estimate the frequency of occurrence of chance variations. R. A. Fisher, a statistician working at an Agricultural Research Station in Harpenden, then devised new techniques for reducing the number of observations needed, by dividing the ground into *randomly* ordered blocks to be treated in different ways.

As a research technique, randomization is said to offer three principal advantages. First, each study unit (plot of earth, person, institution, etc.) has an equal chance of being or not being in the experimental group. Estimates of chance variability are consequently much easier to come by. Second, assignment on the basis of a table of random numbers eradicates the potential for *bias*: researchers are unable to influence their results by choosing to load their experimental group with 'favourable' factors – 'good' seeds, middle-class women, well-resourced institutions. Third, the method allows the researchers evenly to distribute both those factors *known* to be associated with different outcomes and those which may be, but are *unknown*. An instructive example of the latter is discussed by Chalmers (1983) in a paper addressing the competing claims of scientific inquiry and authoritarianism in perinatal care. In a trial of a cholesterol-lowering drug versus placebo in the prevention of repeat myocardial infarction in men (Coronary Drug Project Research Group 1980), no overall benefit for the active drug was found. However, 20 per cent of those prescribed the drug had not actually taken it and mortality in this group was significantly higher, which might lead to the conclusion that the drug really did work. Researchers then went on to look at the group given placebo pills: 20 per cent of these had also not taken their pills and *their* mortality was also significantly higher than those who had. In fact the group that fared best of the four (drug prescribed compliers/non-compliers/placebo prescribed compliers/non-compliers) were men who took the placebo as prescribed. The *behavioural* factors of 'non-compliance' had an unanticipated importance greater than that of *physical* risk factors, and use of random allocation distributed the propensity to disobey

doctors' orders equally between treatment groups, thus permitting valid conclusions to be drawn about the 'real' value of the 'active' drug.

These advantages have led to a characterization of RCTs within medicine and health care research more generally as 'the most scientifically valid method' of evaluating different procedures or types of care (Bracken 1987: 1111). According to proponents of the method, the advantage of random allocation is predominantly *scientific*. It improves the *design* of a study, in part by ensuring that the basic premise – of truly random sampling – underlying the use of statistical tests of significance is correct; in part by clearing the field of unknown 'biases', including those of both researchers and the researched. This removal of the human, subjective element is in line with what Reinharz (1981) and others have described as the 'conventional' or 'patriarchal' research model: research design is laid down in advance, research objectives are concerned with testing hypotheses, units of study are pre-defined, the researcher's attitudes to research subjects is detached, data are manipulated using statistical analyses, replicability of the study findings is stressed, research reports are cast in the form of presenting results only in relation to preset hypotheses, and for approval in an academic community where neither researcher nor researched are allowed identities or personal values. It is, however, worth noting that one of the attributed weaknesses of RCTs – their concern with quantity rather than quality, of life measures – is not a weakness of the method itself, but of its application (see Laing *et al.* 1975 for a counter-example).

Having acquired research funds we then needed to discuss use of the method with those we were asking to use it, namely the four research midwives. In our discussions with them, we emphasized the dual facts that: (a) it was by no means clear that social support was of global benefit to pregnant women (too much social support might be too much of a good thing: at least it was a research question as to which sub-groups of women might benefit); and that (b) we wanted to be able to say something definite about the usefulness of giving social support to pregnant women at the end of the study; use of this method was more likely than any other to enable us to do this. Randomization was done by the midwives telephoning us at TCRU with the names of women who had agreed to take part. The study 'secretary' had sheets of allocations derived

from a table of random numbers and she entered each woman in order, then informing the midwife of the result of the allocation.

As the study progressed, we had many discussions about how everyone felt about this procedure. The midwives were sometimes unhappy about both the process and the results of the randomization. They considered it a problem that random allocation was being used to determine which women received additional social support, as this meant that the women themselves could not choose their fates; it also meant that, in agreeing to participate in the study, they were agreeing to a 50 per cent chance of either receiving additional social support or not doing so. Second, the midwives worried because sometimes women they thought were in need of social support were allocated to the control group (standard care) or those they considered had enough of it already were allocated to receive it. One midwife wrote compellingly in a questionnaire we gave them half-way through the study about the conflict between random allocation and the principles of her midwifery training,

> 'It's very strange in that, if this was practice and not research, you would evaluate each woman and decide if she needed the extra care for various reasons . . . . It's hard if you recruit someone who obviously has major problems and is desperate for extra help, and then she becomes control. I can feel guilty at showing them that extra care is available, and then not offering it to her – even more so if she eventually has a poor outcome to her pregnancy. Conversely, if she becomes intervention and has obvious major problems, I may wilt a little at finding the extra time and stamina to help her!
>
> 'It can be a shame if, at first interview, you feel that a woman has no problems, is well-informed and supported, and yet you know you will keep on visiting, when you could spend that time with someone who would benefit more. But it's often not until you visit two or three times that problems become apparent.'

As the study progressed, observations from the midwives about the initial invisibility of women's support needs became more common. The following dialogue occurred during one of our regular meetings:

> Midwife 1: 'I think sometimes after the first interview, I wouldn't mind writing down which group I thought they needed to be in. I mean, you see them the first time, and their history's nothing, but when you talk to them, you know how awful it is.'
>
> Midwife 2: 'And it may not be anything to do with their obstetric history.'
>
> Midwife 1: 'Very often it isn't. I went to a lady[2] the other day. On the first visit, everything seemed fine. We were talking away and I got to the section on major worries. She said well, yes, I suppose I have, and it turned out that her older son and her husband, who was not his father, had never got on, which could have had a bearing on the pregnancy in which she'd had a small baby. He'd been in trouble with the police, writing cheques, and so had her son; she brought out all these problems existing in her family since she'd remarried, and she said she can see such a difference in her life now. But I mean that sort of thing doesn't come out at first does it?'

In other words, random numbers have the edge over human intuition because human beings are not always right in the judgements they make. The professional ideology of midwifery, along with that of other health professionals, has been shown to lead to discriminatory stereotyping of women, based on such characteristics as working-class or ethnic-minority status (Graham and Oakley 1981; Macintyre 1976).

The midwives in the TCRU study also tried various ploys to control the randomization process. These included: attempting to spot a pattern in the allocations, so that the order of intervention and control allocations could be predicted, and women entered in accordance with what the midwives thought would suit the women best; and good-humouredly trying to persuade the study secretary to tell them in which order to enter different women (they were quick to realize that the secretary would have the pre-set allocation order in front of her when they telephoned). As well as the factor of women's own needs for social support, the four midwives openly confessed concern about distances they had to travel to carry out the home visits, and about other aspects of their work conditions, such as having to visit possibly dangerous ill-lit housing estates late

in the evening. They understandably hoped their intervention group women would live close to home in places which were comfortable and safe to visit.

My own concern as project 'director' on the SSPO study, on the issue of random allocation was, and remained, confused. In the first place, I was committed to the goal of evaluating the effectiveness of social support in a scientific manner acceptable to the scientific community and to policy-makers;[3] this raises its own problems – for instance about the ethics and relevance to women's situation of targeting research at those in power. It is arguable that the usefulness of research in terms of effecting change is greatest when made accessible to the powerless, rather than the powerful. However, the escalating use of unevaluated technology in the maternity care field is a compelling reason for focusing at least some attention directly on those responsible for formulating policy. Because of this goal of reaching policy-makers, I felt it *was* important to carry out the study according to the rules. This, I think was achieved; indeed, our commitment to open discussion of these difficult issues may even have produced greater rigour and consistency in terms of the orthodox model than is normally obtained.

Professional discussions of RCTs are replete with 'anecdotes'[4] concerning people's natural human attempts to control the randomization process. In a trial described by Silverman, for instance, of the effect of artificial light on the occurrence of retrolental fibroplasia (oxygen-induced blindness) in babies:

> Assignment to 'light' or 'no-light' was made on the basis of blue and white marbles in a box. One day, I noted that our head nurse reached into the box for a marble and then replaced it because it wasn't the colour that corresponded to her belief about the best treatment for her babies. I became convinced that we had to shift to sealed envelopes, as used in the British streptomycin trial. When the first sealed envelope was drawn, the resident physician held it up to the light to see the assignment inside! I took the envelopes home and my wife[5] and I wrapped each assignment-sticker in black paper and resealed the envelopes.
>
> (Silverman 1980: 140)

Such revelations draw attention to the discrepancy between the model of 'scientific' research, on the one hand, and its practice, on the other. This discrepancy has been increasingly discussed in recent years, with reports drawing attention to such aspects as the subjective bias inherent in the peer review process (Peters and Ceci 1982), or the tendency for medical researchers to write up and publish research favouring new, as opposed to standard, therapies (Dickersin *et al.* 1987).

CONSENTING ADULTS?

The issue of randomization is closely bound up with the question of to what extent participants in a research project (either an RCT or any other) consent on the basis of full information to take part in it. The issue of informed consent is the second area of conflict between the principles of feminist research practice and the use of the RCT technique.

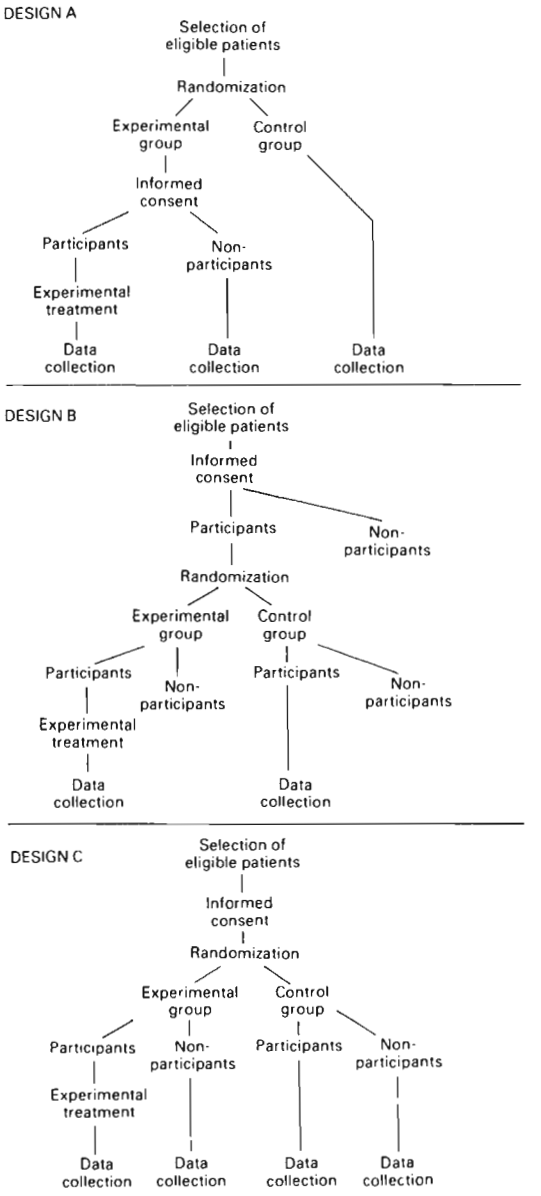
In most, if not all, research studies in the social sciences and the market research/opinion survey domain, the people from whom, or on whom, information is gathered are given the opportunity not to take part. This convention accounts for the citation of what, perhaps revealingly, are called 'refusal' and 'non-response' rates: high refusal or non-response rates are considered to call into question the validity (generalizability) of the research findings, whereas low refusal or non-response rates are generally hailed as an achievement for the researchers.

Consent to participate in research is not, however, the rule in medical research, where, according to Faulder (1985), a considerable proportion of the 10 per cent of British patients she estimates to be included currently in RCTs of one kind or another are likely to be in varying states of ignorance about their status as research subjects. The practice of informing people about their inclusion in RCTs appears to have lagged behind their introduction by many years, and to remain uneven between countries. In the USA, written signed consent came into fashion in the 1960s, although more as protection for the doctor than for the patient, given the different organizational base of the US health care system, and its more obvious domination by the practice and threat of litigation.

Until the late 1950s, there was almost no discussion, either in Europe or North America, of informed consent in medical or health research. The term itself seems to have been created in legal circles in 1957, and this legal base has been a continuing important pressure on doctors and medical researchers to consider the issue of what patients should know. In Britain, the current legal position is that, in seeking informed consent, doctors are not obliged fully to disclose all the risks of any procedure, particularly when disclosure is thought by the doctor likely to cause the patient undue anxiety and/or persuade her/him not to accept treatment medically deemed to be beneficial (Brahams 1983). The essential conflict is between what Fader and Beauchamp (1986: 171) term the 'principle of consent' on the one hand, and a 'methodology of deception', on the other.

There is the issue of *what* patients are told; but there is also the issue of *which* patients are told, *when*, and on *whom* data are collected. From this point of view, it is customary to design RCTs in a number of ways. The differences between the designs centre on two issues: (1) the relationship between randomization and consent, and (2) whether or not the trial analysis is done on 'an intention to treat' basis (so that data are collected on all randomly allocated subjects) or only on those who *were* treated. Figure 7.1 shows three variations on the possible combinations of these practices. In Design A, which has been in common use, informed consent is sought only after randomization from the experimental group. However, data are collected on all randomized subjects whether or not their consent to take part in the research was requested and obtained, so that the sub-group of subjects, randomized to the experimental group, who do not give their consent none the less involuntarily contribute data. In Design B, which is the one we used in the TCRU study, consent is sought before randomization. Only those who agree have information collected on them; those who refuse either initially or subsequently are excluded from data-collection and analysis. Since there is no national register of RCTs, it is hard to know the distribution of different designs. However, it would seem that Design C, in which consent is solicited before randomization, but data-collection and analysis proceed on the 'intention to treat' principle, is probably the most frequently used at the moment in UK medical research (Chalmers 1987).

Figure 7.1 Alternative procedures for random allocation and informed consent



The implications of the different designs for *what* people are told are also quite different (although clearly other factors also influence the extent of disclosure characterizing the consent process). In designs B and C, as opposed to design A for example, what people are asked to consent to is not either to receive a particular intervention or not to receive it. They are asked to agree to putting themselves in a position where they have a fifty-fifty chance of receiving a particular treatment, or not receiving it. In the TCRU study, using Design B, the midwives were asked to follow a text prepared by us, explaining this when talking to the women eligible[6] to enter the study. Women who agreed to take part were randomized and then informed of the result. Twenty-five women out of 534 asked said they would rather not take part in the research. After randomization two women changed their minds – one in the control and one in the intervention group.

Design B is not in common use in RCTs of medical care carried out in Britain. The reasons why we chose to use it can be deduced from the arguments advanced in favour of the other methods, particularly design A. A leading article in the journal *Science* in 1979, 'Informed consent may be hazardous to health' says, for example:

> Before human subjects are enrolled in experimental studies, a variety of preliminary rituals are now required. These include an explanation of the nature of the experimental procedure and a specific elaboration of possible adverse reactions . . . . These rituals are said to increase the subjects' understanding of the procedures . . . .
>
> A considerable body of psychological evidence indicates that humans are highly suggestible . . . . This alone would lead one to suspect that adverse reactions might result from information given during an informed consent discussion . . . possible consequences of suggested symptoms range from minor annoyance to, in extreme cases, death.
>
> If protection of the subject is the reason for obtaining informed consent, the possibility of iatrogenic harm to the subject as a direct result of the consent ritual must be considered.
>
> (Loftus and Fries 1979:11)

This is an extreme statement of a position quite commonly stated in the medical literature. The term 'ritual', used in a derogatory sense, immediately tells one how these investigators regard the task of informing people what is being done to them. In another version of this response, the danger that the stress of receiving information may affect health directly results in the conclusion not that *all* patients in a trial should be equally informed, irrespective of group allocation, but rather that *no* patient should be informed at all. Instead, this job is best left to 'ethical'[7] committees, because:

> Only such committees have the experience, skill, information, and emotional detachment to judge the merits of a research protocol; the individual patient has none. . . .
>
> As a rule, patients will consent to randomization on the basis of skeleton information only, if they know that a committee of professional and other interested people has on their behalf studied and approved the scientific merits and ethical aspects of the trial. Most patients will not need, want, or ask for more.
>
> (Papioannou 1982: 828)

Such a viewpoint invites a number of obvious responses. First of all, reliance on the efficacy or ethical behaviour of ethics committees is dangerously ill-informed. In Britain, no law requires the submission of research proposals to such bodies. There are no statutory or other regulations governing the practices followed in them, which are widely variable, as are the rules controlling membership (British Medical Association 1981).

Second, in the debate about informed consent, the question as to what patients want is rarely constituted as an empirical question, to be settled by appropriate inquiry. Dissatisfaction with the amount and kind of information they are given is the most common complaint patients make about medical care in both the UK and the USA (see King 1986 for a review of this and other areas related to the informed consent debate). A number of studies show that even among people with fatal conditions, over 80 per cent would prefer to know the truth (see, for example, Cassem and Stewart 1975; Aitken-Swan and Easson 1959; McIntosh 1976; for a discussion see Gillon 1987). Studies by Mattson and colleagues (1985) show that among participants in two large trials of drugs to

prevent heart disease, increased knowledge was ranked in first or second place as benefits they personally derived from taking part in research. A major finding of these studies was that altruistic motivation ('to help others') was almost as important as the incentive of gaining better care for themselves. Altruism has been an important finding in other fields, including maternity care (Elbourne 1987) and the non-commercial blood donation system that exists in Britain (Titmuss 1970). In the TCRU study, altruistic motives were referred to by a number of the women. Comments included:

> 'Happy to be of help in getting information to others, and hope to help to improve the system.'
>
> 'Very glad to do anything which might help women to cope better with pregnancy.'
>
> 'Took part in research to help other people.'
>
> 'Quite willing to help with future knowledge for all mothers. Hope it will be helpful.'
>
> 'I feel it is very good that things are being researched about childbirth.'

In an interesting account of a 'failed' RCT concerning methods of delivering very low birthweight babies in Australia, Lumley and colleagues (1985) describe how women's reluctance to subject their obstetric care to the play of chance on the basis of informed consent was anticipated by professionals as a problem before the trial started; however, the reason why it proved impossible to carry out in practice turned out instead to be the unwillingness of professionals to subject their behaviour to systematic evaluation. But it is none the less true that research on people's attitudes to information and consent in both health care generally and health research specifically is astonishingly limited. One obstacle is that doctors think patients want to know somewhat less than they do actually want to know (Waitzkin and Streckle 1976). Very little research, for example, has been devoted to the question of how best to give information. One survey of surgical consent forms used in the US found that readability presupposed a university education and four out of five were written at the level of a scientific journal (Grunder 1980). Growing concern about the

informed consent issue has now bred the ultimate in RCTs: an RCT of informed consent procedures itself (Simes *et al.* 1986). This compared 'total disclosure' with written consent to an 'individual approach' with verbal consent in patients with cancer who were candidates for RCTs. 'Total disclosure' resulted in more knowledgeable and informed patients who were also less willing to take part in RCTs and were judged to have more anxiety than the less informed group; since anxiety is assumed to be a negative outcome, it is commonly taken to count against the use of 'total disclosure' procedures.

In a chapter advocating the use of RCTs in perinatal research, Bracken introduces two further favoured medical reasons against informing patients when he says that:

> Patients are blinded in clinical trials for two major reasons. First, if blind, they are unlikely to withdraw from an RCT after being randomized into what they believe to be the less effective treatment. Such selective withdrawal would be extremely serious for the success of a trial. Second, blinding the patient avoids placebo effects.
>
> (Bracken 1984: 406)

Not telling people they are part of a research project thus unsurprisingly results in the advantage (for the researcher) that they cannot refuse to co-operate. But what are 'placebo effects' and what are they doing in this argument?

People's belief in the efficacy of what is being done to them has measurable effects on their health (see Brody 1977). Medical professionals hold interestingly divided views on this point. On the one hand, much of the historical efficacy of medical care must have rested on this basis (Ballantine 1975), since, until the twentieth century, relatively few medical procedures could be shown to have any other mechanism of therapeutic effect. The 'placebo' (literally 'I please') response also shores up the notion, close to the heart of many professionals, that the relationship they have with their patients may in itself be healing. These considerations reveal the tension of the physicalist model of health, since it is evident that mind and body interact and that medicine's very claim to effectiveness depends on this, at the same time as disputing it. From a research point of view, however, the placebo effect is nothing more than inconvenient nonsense. A well

designed trial must assess it, in order to be assured of credibility.

The giving of information is also said to sound the death knell for the double blind RCT itself. Brownell and Strunkard's paper, 'The double-blind in danger: untoward consequences of informed consent', reports a trial of fenfluramine in 'obese women' in which information given in advance about possible side-effects of medication enabled four out of five patients to guess correctly that they were taking the medication. In turn, this correct estimate was associated with greater weight loss. According to Brownell and Strunkard:

> The implications of these findings are grave: the double-blind methodology . . . can be breached and may be endangered by the demands of institutional review boards for increasingly detailed informed consent.
>
> (Brownell and Strunkard 1982: 1488)

What is highlighted are the consequences for the scientific method of allowing people to know what is happening to them, rather than the therapeutic effect of empowering them with this knowledge.

In the TCRU study, we encountered various difficulties with the method and content of the informed consent procedure we used. Much the most important of these was the extent to which our informing women who were subsequently allocated to the control group what the study was about may have resulted in their feeling deprived (see the midwife's comment, p. 176). One example is this experience described by one of the midwives:

> 'Dawn Benn[8] (Control). She was absolutely desperate to be intervention and she was so upset when I phoned her because of her social circumstances that when she asked me if I knew the address of any mother and toddler groups, NCT groups, anything, because she's just moved into the area, I gave her a couple of phone numbers before I'd even got her allocated. She was heartbroken.'

In other instances, control group women got in touch with the midwives for help even though they knew they had been assigned to the no-support group. We asked the midwives to respond minimally in such situations, conscious that responding fully would be to jeopardize the aims of the study. It is interesting to

note that control group women were only able to request this help because the midwives felt it was unethical for them not to give all women in the study a contact address and telephone number. They did so, although this was not part of the study design, and partly in response to our discussions about the informed consent issue.

Evidence of a 'deprivation' response was also gleaned from a few of the women who took part, in the questionnaires they filled in after delivery: for example,

> 'Pleased to help. I would have welcomed home visits from a midwife.'
>
> 'I was very pleased to help, but would have enjoyed the visits . . . '
>
> 'Happy . . . (to take part in the research) if it can help anyone in the future. Would have preferred to be in group visited.'

For other control group women, the desire to feel included took such a form that they felt they had been. One of the sections of the questionnaire asked women to tick a list of possibly helpful people, including the research midwife ('if she visited') and thirteen control group women ticked this, though they had not in fact been visited at all. Other appreciative remarks included:

> 'I wanted to take part and kept a note of how I was feeling at the time of my son's birth.'
>
> 'Pleased to have even a small interest taken in my feelings and opinions only inasmuch as feeling I would have some "right to reply" had I been treated very badly.'
>
> 'I didn't mind, especially if it helps to understand or prevent premature birth. By filling in this form it helped me to understand any problems I have because I had to write them down.'

One woman in the control group even went so far as to say she felt special as a result of taking part in the research, it was 'like belonging to an elite group'.

Because we gave information to all the women the midwives identified as eligible for the study, the women in the control group *were* part of a research project in which they had *chosen* to partici-

pate. In this sense, it is probably true to say that rigorous testing of the hypothesis that social support can improve pregnancy outcome is at odds with the principle of informed consent. A further complication is that the standard scientific model of RCTs presupposes that there is no 'contamination' of the control group by the experimental group; the purpose of the control group is, after all, to act as a 'control' for experiment. (In French, the term for 'control group' is 'le groupe temoin', literally 'witness' group, which carries an interestingly different connotation.) Again, the actual practice, as opposed to theory, of research reveals the chimerical nature of this model. It is assumed, for instance, that people do not talk to one another. In our study, had we not told the control group women about the study, we would have needed to rely on the intervention group women remaining silent about their receipt of social support – at home, with friends and neighbours, in antenatal clinics. But women are not silent. Although this human tendency to communicate may be overcome by randomizing groups (areas, institutions) rather than individuals, the tension between the scientific requirements of research and the humane treatment of individuals remains, and is expressed in the very strategy of designing an experiment so as to restrict people's freedom to discuss with one another the commonality of the process in which they are engaged. A potent example of the scientific disruptiveness of this was described in an RCT of health screening for the elderly in Finland which failed to test the effectiveness of this form of health surveillance, as a local media campaign raised the rate of follow-up examinations in the control group to a figure that was 10 per cent higher than the rate in the experimental group (Antilla and Isokoski 1985).

A second problem we encountered with our informed consent procedure was in deciding how to present the aims of the study to the women we asked to take part in it. The standard procedure for an RCT involves the researcher setting out to test something affectionately known as 'the null hypothesis'. Adoption of the scientific method requires that one begins from the standpoint that there is no difference between the treatment and the lack of treatment, or between the treatments that are being compared. But since we had designed our study to test the hypothesis that social support might affect the baby's weight, as well as to investigate other factors such as the type of delivery mothers had, and

how they felt about their experiences, the dilemma was whether we should say so; or if we said 'we want to see if social support can make babies grow better' (for example) were we somehow biasing the study from the start? Might we turn the study into a self-fulfilling prophecy the results of which would never be believed except as such? This formulation of the question is interesting, for it leads to the next question, which is, if the hypothesis is that a *social* process may be therapeutic in a *clinical/physical* sense, what sensible arguments are there for concealing the purpose of the exercise from those taking part in it? The reason why researchers are wary about the placebo response is, after all, because of the possible beneficial impact on someone of feeling they are being cared for: it is the very presence and effect of the social process that is counted as disturbing. But for us the social process was of central concern. It was for this very reason that we decided to enlist the midwives' confidence in the aims of the study from the start: not to have done so would, we felt, not only have been unethical, but also intuitively counter to its aims. For the same reason, we were conscious of the tension with the principle that women allocated to the control group should not feel deprived. Thus, we also stated what we also believed, that we did not *know* whether social support could improve the health of all women and their children in this sense.

THE IMPORTANCE OF BEING (UN)CERTAIN

Third, we come to the last of the issues raised at the beginning of this chapter as especially problematic in this type of research – the question of uncertainty.

Much of the literature on informed consent refers to an uncomfortable prerequisite for the seeking of consent, which is that of researcher or practitioner uncertainty (our not knowing whether social support works). Certainty can be a consequence of very different political and ideological positions. Medical certainty, for example, lies behind one of the most commonly used informed consent/randomization procedures (design A, p. 181). According to Zelen, whose advocacy of this design has been influential, the main advantage of withholding information from the control group is that clinicians do not need to admit to these patients 'that they do not know which treatment is best'. Zelen explains that

> . . . many investigators and patients are reluctant to participate in randomized clinical trials. . . . One of the principal reasons why clinical investigators decline to participate in randomized studies is that they believe that the 'patient-physician relation' is compromised . . .
>
> (Zelen 1979: 1242)

For this reason, use of the RCT is sometimes said to be 'unethical'; the question of ethics enters the medical debate at the point at which the traditional paternalism of the doctor–patient relationship is threatened. From a vantage point outside medicine, however, such professional understandings are precisely the reason why RCTs have an ethical advantage over routine medical practice. They subject to external assessment the medical claim to therapeutic effectiveness (which rests partly on the ambiguous notion of trust). As Freidson has observed, technological/scientific autonomy is *the* premise on which the professionalization of medicine rests; medicine is 'free to develop its special area of knowledge and to determine what are "scientifically acceptable" practices' (Freidson 1970: 83). The aphorism 'doctor knows best' is both a manifestation of this control, and its translation into the politics of the doctor–patient relationship. As Freidson also observes, the *claim* to sole ownership of scientific wisdom says nothing about the *actual* scientific status of the knowledge thus claimed.

However, certainty is not the prerogative of medical professionals. It is also possessed by lay people and by women. The women's health movement has been guilty of a fair amount of misguided certainty over the years, as, for example, in the recently fashionable demand that cervical screening programmes be made more readily available to all women. As Robinson (1987: 51) has noted: 'Neither the ethics, the efficacy, nor the adverse effects of screening have been adequately discussed by women's organizations.' In the childbirth field, many attempts systematically to evaluate different modes of care have been shipwrecked on the rock of women's certainty about the effectiveness of apparently natural and innocuous methods such as childbirth education, vitamin supplements, or raspberry leaf tea. (While such methods may have this effect, there is, as yet, no scientific evidence, and even some to the contrary (see Chalmers 1983).) Perhaps best-

publicized of recent examples in Britain has been the response of maternity service user groups to the Medical Research Council's RCT on vitamin supplementation in pregnancy. The trial was designed to test the hypothesis that such supplements, taken around the time of conception, can reduce the chances of a baby having a neural tube malformation. User groups such as the National Childbirth Trust and the Association for Improvements in the Maternity Services contended that the need for a good diet in pregnancy was well established, making the trial unethical (Micklethwaite *et al.* 1982). The point is that, whatever form it takes and whoever professes it, certainty blocks progress towards greater understanding of the role of chance versus causation in the patterning and human experience of events and processes, including those responsible for health or its absence.

Genuine scepticism about something is probably rare. It appears that researchers must merely possess sufficient *un*certainty about something in order to want to find out about it. The issue of certainty was complicated in our study. None of us was prepared to say that we did not know whether social support was a good thing (in the same way as we would have been prepared to say, for example, that we did not know whether it is helpful for women prone to premature labour to be admitted to hospital during pregnancy). But while it may seem almost axiomatic that social support, like love, is something we all want, what is at issue is the range and type of event/process social support is capable of affecting, and the mechanism by which it does so (see Madge and Marmot 1987 for a discussion). Assumptions about the inevitably therapeutic effects of social support may prove unfounded when subject to systematic evaluation. This was the case, for example, in the Cambridge-Somerville Youth Study, an early attempt initiated in 1935 to evaluate the long-term effectiveness of social work help in preventing delinquency and other 'undesirable' outcomes, in which it was found that the intervention of social workers was associated with more problems rather than fewer later on (McCord 1981, 1982).

CONCLUSION

The RCT is a method of *experimental* research, and the term 'experiment' has been linked with what Chalmers (1983) has

called the 'Auschwitz' view of scientific inquiry, according to which all experimental research is inherently suspect. The view of experimental research as inherently unethical is central to the feminist critique (Spallone and Steinberg 1987; Birke 1986) but also comes from other quarters (see Silverman 1985). Much of it misses the absolutely crucial point that the condemnation of experimentation under the heading of 'research' allows a great deal of experimentation to pass unnoticed under the heading of standard practice. The frequency with which doctors impose on patients experiments of an uncontrolled nature has been one of the strongest objections to professionalized medicine made by the women's health movement over the last twenty years in Europe and North America (Ruzek 1978). The fact that very large numbers of women have been treated with medical and surgical procedures of unknown or suspect effectiveness and potentially or actually harmful consequences has been taken to signal both women's status as a minority group, and medicine's essentially unscientific standing. For this reason, women have been, and continue to be, important beneficiaries of the advocacy of randomized controlled evaluation within medicine. One significant example concerns the treatment of breast cancer, a disease which affects one in twelve women in the United Kingdom at some point in their lives. Analysis of the results of trials of breast cancer treatments has been responsible for the production of persuasive evidence that 'conservative' treatments are superior to 'radical' treatments both in prolonging life and in assuring a better quality of life. Overviews of RCTs concerned with systematic treatment (chemotherapy or endocrine therapy) of the disease show important differences between the effectiveness of such treatments in older and younger women. They also provide evidence that short courses of treatment are as effective as longer courses – an important consideration, given their sometimes unpleasant side-effects (Consensus Development Conference 1986). It is, however, of interest and important to note that the benefits that accrue to women as a result of their willingness to participate in such studies cannot be held out as a carrot to those who do except by trading on altruism. It is the *future* health of *other* women that stands to benefit by the willingness of *some* to be experimented on *now*.

What our experience with an RCT of social support in pregnancy has shown is the need to subject every precept of the

traditional scientific method to scrutiny. Is it necessary? Do its benefits outweigh its hazards? It is as important to ask these questions of a trial of something as apparently harmless as social support as it is of trials of other more obviously ambiguous therapies. The argument against 'methodolatry' is then transformed into the case for an *appropriate* methodology which, like its namesake, appropriate technology, requires that individuals involved in it be treated with sensitivity and respect, and *that there be no division between this ethical requirement and other requirements of the method.* This is not, of course, to say that the procedure of randomized controlled evaluation is the *only* means to reliable knowledge, is *sufficient* in itself, or is *always* the right approach; for the pursuit of truth in human affairs is, as we all know, ultimately an illusion, and reliable knowledge definitely not a good in itself. The point is that what Rowbotham (1985: 51) has called 'the attraction of spring cleaning' should be seen as a means to an end, not an end in itself. The frenetic housewife is unable to enjoy the product of her labours. In Evelyn Keller's words:

> The intellectual danger resides in viewing science as pure social product; science then dissolves into ideology and objectivity loses all intrinsic meaning. In the resulting cultural relativism, any emancipatory function of modern science is negated, and the arbitration of truth recedes into the political domain.
>
> (Keller 1982: 593)

ACKNOWLEDGEMENTS

The original version of this chapter was published in *Women and Health* (Oakley 1990). We are grateful to the journal and its editors for permission to reprint.

NOTES

1. The hierarchy of academia insists on the notion of a project 'director'. However, the Social Support and Pregnancy Outcome study has been a collective venture which would not have been possible without the energy and commitment of: Carol Galen-Bamfield, Sandra Buckle, Rosemary Marsden, Rosie Smith, Lyn Rajan, and Sandra Stone. I should also like to acknowledge the help of Penrose Robertson as the

token man facilitating our use of appropriate technology. I am indebted to members of the National Perinatal Epidemiology Unit in Oxford, and especially to Iain Chalmers and Adrian Grant, for social support, intellectual inspiration, and technical help from (and before) the beginning of the study. Thanks are also due to the Department of Health who funded the study.

2. The midwives referred to the women in the study as their 'ladies'.
3. This raises the question, which is not discussed here, of to what extent the scientific community does indeed practise science.
4. Anecdote means 'things unpublished' but has degenerated to equate with 'an item of gossip' (OED). (And 'gossip' itself is a term which, from describing the important witnessing and authenticating of childbirth by women in the community, is now used in a derogatory sense to refer to 'talk amongst women'.)
5. The role of wives in research, as in other spheres, is revealing.
6. Aside from having had a LBW baby in the past, the women had to be booking into a hospital before 24 weeks of pregnancy and they needed to be able to speak English fluently, as we unfortunately had no money to pay interpreters.
7. An incorrect term. I am grateful to Iain Chalmers for the observation that ethics committees may be ethical – or not.
8. A pseudonym.

Chapter Eight

BEHAVING WELL:

Women's health behaviour in context

HILARY GRAHAM

The election of a Conservative government in 1979 is often presented as marking a radical break in welfare policy. In particular, it has been associated with a well co-ordinated challenge to a consensus about the role of the state in health care, publicly adhered to by the major political parties since 1945. However, in emphasizing the changes that the last decade has brought to health policy, we should not lose sight of continuities. We should be aware of the themes which span the years before and after 1979. One such theme emphasizes the responsibility that individuals have for their own health. Here, it is health behaviour – behaviour affecting health – that is stressed.

The theme was succinctly expressed in the title of the 1976 Consultative Document, *Prevention and Health: Everybody's Business.* Subtitled *A Re-assessment of Public and Personal Health*, it argued that, unlike a century ago, the major health issues of the late twentieth century related not to public health measures but to personal health behaviour. Increasingly, the document suggests, it is the habits of individuals (defined in male terms) which determine the health of the nation. It noted that:

> We as a society are becoming increasingly aware of how much depends on the attitudes and actions of the individual about his health. Prevention today is everybody's business .... Many of the major problems of prevention are less related to man's outside environment than to his own personal behaviour; what may be termed his lifestyle .... To a large extent, it is clear that the weight of responsibility for his own health lies on the shoulders of the individual himself. The