BIOMARKERS, 2003, VOL. 8, NO. 1, 13–32

Taylor & Francis
healthsciences

REVIEW

The randomized controlled trial in studies using biomarkers

PAOLO VINEIS

Dipartimento di Scienze Biomediche e Oncologia Umana, via Santena 7, University of Torino, Torino, Italy. e-mail: paolo.vineis@unito.it

Received 24 October 2001; revised form accepted 5 September 2002

The randomized controlled trial (RCT) is a scientific experiment during which observations on the effects of therapy or a preventive action are conducted by the researcher under rigorous control. The purpose of the experiment is to clear the uncertainties surrounding a clinical/research issue and involves isolating the 'treatment' and 'end result' variables from external influences. RCTs therefore make use of scientific method standards: measuring, which includes the possibility of reproducing observations; controlling factors unconnected to the cause–effect relationship of interest; and the external verification or 'falsification' of the cause–effect relationship. Many RCTs are now including biomarkers to answer scientific questions in a more accurate way. In the present methodological paper, the main aspects involved in the design and conduction of a trial are discussed, with special emphasis on the use of biomarkers. Aspects that are often overlooked by scientists involved in the design of trials include multiple comparisons, subgroup analysis, the duration of the observations, the use of surrogate endpoints, and ethical issues. This review summarizes the main issues that should be addressed in a protocol, and illustrates these with an example.

Keywords: surrogate endpoints, design, power, NNT

Introduction

This review will discuss when it is reasonable to plan a randomized controlled trial (RCT) in the field of biomarkers, how to plan it in order to obtain reliable results, and what ethical issues arise in the process.

An RCT is a scientific experiment during which observations on the effects of therapy or a preventive action are conducted by the researcher under rigorous control. The purpose of the experiment is to clear the uncertainties surrounding a clinical/research issue and involves isolating the 'treatment' and 'end result' variables from external influences. RCTs therefore make use of scientific method standards: measuring, which includes the possibility of reproducing observations; controlling factors unconnected to the cause–effect relationship of interest; and the external verification or 'falsification' of the cause–effect relationship. Typical questions that an RCT might be used to answer are: does a low dosage vitamin complex compared with a placebo prevent cataract? When treating a grade 4 astrocytoma, what works better – the combination of three cytostatics, or carmustine on its own? A glossary of terminology relevant to RCTs is given in appendix 1.

Randomized trials are used in order to gain the highest degree of scientific proof regarding the efficacy of a given treatment. Such treatment may be therapeutic or preventive, and must be open to experimentation and randomiza-

Biomarkers ISSN 1354–750X print/ISSN 1366–5804 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/1354750021000039389

tion. Not all medical research is amenable to experimentation: testing the toxicity of cigarette smoke on humans, for example, would not be ethically acceptable. Equally, randomization of a new therapeutic or preventive treatment is not always appropriate either. For example, many types of psychotherapy are not suitable for randomization because of the fundamental role played by the subjective interaction between a specific therapist and a specific patient, and because of the particular indications for each type of treatment.

Therefore a background question must be asked as to whether the problem posed is suitable for experimentation, and specifically in a randomized trial. Studies using this type of trial mainly concern pharmacological treatments or generally simple interventions with therapeutic or preventive aims.

Why have randomized trials become the 'gold standard' of pharmacological research? As often happens, this did not occur for abstract methodological reasons, but because of mistakes previously made with other study designs. By examining past research it is easy to recognize the main reasons for mistaken conclusions reached by researchers: (1) the lack of a control group, which led to improvements or remissions that had occurred spontaneously being ascribed to the drug; (2) the lack of 'blindness' in the observations, leading to a distorted interpretation of the results due to the observer's 'prejudice'; (3) poor comparability between those treated with the drug under scrutiny and those who were not, for example because the latter were more serious (this is reminiscent of the first observations that suggested that infarct victims treated at home survived longer than those treated in a coronary unit – the study was not randomized, and the latter subjects had suffered more serious infarcts and consequently were not comparable to the former group); and (4) the limited dimensions of the studies, which meant that falsely positive or negative results could be obtained by chance.

The aims of an RCT are (Piantadosi 1997):

(1) To quantify and reduce chance errors (induced by an inadequate dimension of the experiment with regards to the variability of the phenomenon under observation).
(2) To reduce or eliminate systematic errors induced, for instance, by selective and unbalanced recruitment of the groups treated with the drug and the placebo.
(3) To provide a relevant and precise appraisal of the effects of the therapy under scrutiny.
(4) To provide a plan and an analysis of simple and easily understood data.
(5) To provide a high degree of credibility and reproducibility of the results.
(6) To influence future clinical practice.

An example: a randomized trial on preventing DNA damage with diet-related flavonoids

To illustrate the planning of a study we will refer to a trial on dietary prevention currently in progress. Everything that will be said about this trial also applies to pharmacological trials. In what follows we have imagined we need to address a board that will examine our project and decide whether to finance it, and that we have to write a convincing protocol.

The scientific background and the research issue

What is the aim of the research? Do we have a solid scientific basis that justifies its implementation? Is there a valid animal model? Have tests on animals, on culture cells, or other experimental systems been implemented to support the hypothesis? Are there pilot studies on humans? Every trial involves expenditure of energy and resources, and, for the participants, difficulties and discomfort. We must therefore avoid pointless repetitions and must not research subjects that are futile or of little relevance.

In preventive intervention trials such as the one we are discussing, the criteria employed to conduct pilot studies are not particularly stringent, whilst they are for pharmacological trials. Pharmacological trials, when conducted on humans, are usually divided into four phases:

(1) Phase I refers to studies on healthy volunteers, aimed at studying the drug's metabolic destiny, its bioavailability, excretion, etc. (in order to establish the best dosage, i.e. dose-finding studies).
(2) Phase II refers to studies, generally without a control group, aimed at identifying possible toxic effects and a preliminary therapeutic effect in a small number of patients.
(3) Phase III refers to randomized controlled trials.
(4) Phase IV consists of monitoring possible side-effects (too rare to be highlighted in phase II or III) or interactions with other drugs after release into the market.

The aims of our study

(1) To verify within the human population the hypothesis that antioxidant flavonoids contained in fruit and vegetables protect from cancer of the bladder and other organs by inhibiting the formation of DNA damage due to tobacco smoke.
(2) To conduct an RCT amongst heavy smokers comparing the level of (adducted) DNA damage in exfoliated bladder cells between a group of subjects treated with a diet rich in flavonoids and a group treated with a diet poor in flavonoids. Exfoliated bladder cells will be collected before and after the treatment.
(3) To determine the relationship between the carcinogens contained in tobacco smoke, DNA adducts in the exfoliated bladder cells, urinary anti-mutagenic activity and the dietary levels of exposure to flavonoids. These relationships are especially relevant not only to carcinogenesis in the bladder, but also to the induction of cancer in other organs.
(4) To bring about practical consequences, including setting up primary prevention activities through family doctors aimed at increasing the consumption of flavonoids in daily food.

What knowledge do we already have?

Before planning the experiment we must be clear about its scientific assumptions and begin with a systematic and reasoned review of the existing literature,

asking ourselves whether others have already had the same idea, how grounded the idea is, and what the possible practical applications of the trial's results would be.

Previous knowledge. Many studies have suggested that a 'Mediterranean' diet, and generally a high consumption of cereals, fruit and vegetables, reduces the risk of cancers at different organ sites, including the colon, breast, bladder and prostate (Miller *et al*. 1994). The Mediterranean diet plays a protective role in cardiovascular disease, and different components of this diet have attracted attention, especially tomatoes and olive oil.

A recent study among American nurses (Giovannucci *et al*. 1995) highlighted a considerably reduced risk of prostate cancer among heavy tomato or tomato juice consumers. In case-control studies conducted in Spain, Greece and Italy (Trichopoulou 1995), a high level of olive oil consumption was significantly associated with a reduced risk of breast cancer, while total fats were not associated with such a risk. It is not known which specific micronutrients are responsible for the protective effect of tomatoes, olive oil and other components of the Mediterranean diet, although it is plausible that it is due to different types of antioxidants (Block *et al*. 1992).

In vitro studies have proven that the Mediterranean diet's polyphenolic compounds, and especially oleuropein (which is responsible for the bitter taste of olives), interfere with biochemical events involved in atherosclerotic pathology (Visioli and Galli 1994). In addition, *in vivo* studies have suggested that polyphenols present in red wine increase the antioxidant capability of plasma and therefore reduce the susceptibility of low density lipoprotein cholesterol to peroxidation (Furman *et al*. 1995).

Flavonoids are a particularly relevant group of polyphenols found especially in fruit, vegetables and some drinks. Their consumption in food in human populations varies from 23 mg to 1 g a day per person (Hertog *et al*. 1993). A high concentration of flavonoids and other polyphenols has been measured in onions, lettuce, red wine and other elements of the Mediterranean diet. It has been demonstrated that these substances are excreted through urine (Lampe *et al*. 1994), and some of them have shown the capability to inhibit the mutagenicity of various heterocyclic aromatic amines (Weisberger, 1994). Heterocyclic amines, including 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhlP), a powerful mutagen and experimental carcinogen, are recognized as a class of carcinogens contained in the human diet (Vineis and McMichael 1996).

It has been suggested that the consumption of flavonoids reduces the level of DNA adducts both in humans and animals. Adducts are the result of the reaction between exogenous 'electrophilic' substances and DNA; if the adducts are not repaired, mutations arise.

Moderate wine consumption (a source of flavonoids) inhibits micronucleated cells induced by peroxides (Fenech *et al*. 1997), and flavonoid consumption inhibits DNA damage related to lipid peroxidation (Cai *et al*. 1997).

In Italy the incidence of cancer of the bowel, bladder, prostate or breast is higher among subjects born and living in the north (where this trial will take place), is intermediate among immigrants, and is lower among those born or living in the south (Rosso *et al*. 1993). A higher prevalence of the Mediterranean diet or Mediterranean dietary habits shared with the Greek and Spanish populations are found in the southern parts of Italy

Pilot studies. If the knowledge basis seems sufficiently solid, we must now show that we are able to run the trial and that we have already conducted pilot studies that allow us to plan the current project better.

We have previously shown that urinary extracts contain substances that powerfully inhibit urinary mutagenicity related to smoking (Malaveille *et al*. 1992). We hypothesized that the source of these antimutagenic substances was diet and could be associated with Mediterranean dietary practices. We then conducted two pilot studies in groups of volunteers. To investigate the biological relevance of the antimutagenic activity in urine, 10 smokers, whose levels of DNA adducts in exfoliated bladder cells were known, were studied. Later, with the purpose of formulating hypotheses on the chemical nature of these antimutagens, we determined the total concentration of polyphenols (including the main flavonoids and/or their metabolites) in the urine of 19 volunteers and correlated these concentrations with antimutagenic activity.

DNA adducts were correlated to cigarette smoke, one of which was adduct *N*-(deoxyguanosine-8-yl)-4-aminobiphenyl (ABP-dG) (Talaska *et al*. 1991). Both ABP-dG and total DNA adducts were found to be inversely correlated to bacterial antimutagenicity, expressed as the decrease in numbers of *Salmonella typhimurium* TA98 revertants (PhlP-induced mutations) per millilitre of equivalent urine. The logarithm of DNA adducts concentration was found to inversely and linearly correlate with antimutagenicity in a statistically significant way ($r = -0.81$, $p < 0.01$ for both ABP-dG and total adducts) (Malaveille *et al*. 1996). To identify the chemical nature of the antimutagens we measured the amount of polyphenols in the urinary extracts of the second group of volunteers ($n = 19$) using a spectrophotometric method. Concentrations were compared with the antimutagenic activity of the urine, again using PhlP as the mutagen in *Salmonella typhimurium* TA98. Every urine extract contained measurable quantities of polyphenols, with concentrations ranging from 3.7–12.5 μg/10 μl of extract solution. A statistically significant linear relationship was found between the antimutagenicity of urinary extracts and the concentration of polyphenolic substances ($r = 0.58$, $p < 0.02$).

In a later pilot study we studied the concentration of adducts with 4-aminobiphenyl DNA in bladder biopsies from 34 patients with cancer of the bladder. We found that the median level of adducts was approximately 4 adducts/10^8 base pairs in the 10 patients who had not consumed fruit and vegetables in the previous 24 h, 2 adducts/10^8 base pairs in the 11 patients who had consumed between one and three portions of fruit and vegetables, 1.0 adducts/10^8 base pairs in the six patients who had consumed more than three portions, and 10.6 adducts/10^8 base pairs in the seven patients who had not consumed fruit or vegetables for some time, having changed their dietary practices as a consequence of the disease's symptoms. Mean values and standard errors in these groups were 5.5 ± 2.7, 4.5 ± 1.9, 1.9 ± 1.0 and 11.2 ± 4.1 adducts/10^8 base pairs, respectively. Although the differences are not statistically significant, these figures suggest a protective effect by fruit and vegetables against the formation of DNA adducts in urothelial cells (unpublished data).

These studies were small-scale pilot studies. In general, such studies should not be too small, otherwise they may produce false positive results (by chance), thus stimulating formally designed investigations with a poor rationale and waste

of resources. Thus even pilot studies require an estimate of the study size necessary to achieve the expected goal.

Plan and methods

Once it had been established that the reason for conducting the trial is sufficiently sound, we needed to choose the most appropriate plan. The pilot studies we have conducted allowed us to clarify some of the plan's details. Since the randomized trial is the gold standard of medical research, we must look at its pertinence and applicability to our problem:

(1) Is it materially possible to randomize dietary behaviour, administering a flavonoid-rich diet to one arm of the study and a normal diet to the other? The answer is yes, although with a few expedients.
(2) Is randomizing diet ethically acceptable? The answer is again yes, although the ethical issues will have to be carefully evaluated.

If randomization had proven impossible, we could have fallen back on an almost-experimental plan (i.e. a 'before and after' scheme, in which a flavonoid-rich diet and then a normal diet, or vice versa, are administered to the same subjects) or some kind of non-experimental plan. The pilot studies described above were in effect observational studies, in which the subjects' exposure to flavonoids was not experimentally modified.

The study will have a randomized plan and will be based on administering a diet rich in flavonoid compounds for a month to a group of healthy, smoking volunteers, whereas the control group will receive a diet poor in flavonoids.

Study population and ethical issues. It is a good practice within all controlled clinical trials to define with great precision the 'target population' a trial is aimed at. This is because therapies and preventive activities are assumed to be specific by type and stage of pathology, an assumption that is not always verified. Furthermore, only people who are able to tolerate the treatment are recruited (in reasonable general health conditions) and who are likely to survive long enough. The negative effect of very rigorous selection criteria is the possibility that the result will be highly artificial, that is applicable only to a very selected patient subpopulation and not generally to patients in routine clinical practice.

In the study we are planning, the volunteers smoke between 10 and 20 cigarettes a day; made from air-cured tobacco this type of tobacco is particularly rich in aromatic amines and, as has been highlighted, increases the risk of bladder tumours more than heat-cured tobacco. All the volunteers are men resident in the province of Torino, aged between 35 and 70 years. They will be recruited through the lists of local blood donors' associations we have worked with over the past 15 years.

Several ethical issues need to be considered:

(1) Is it acceptable to involve smokers in the study, that is people who inflict considerable damage to their health? The answer is yes if we take care to clarify from the start that smoking is highly noxious, that we intend to supply the volunteers with any data that might be useful for them to give up smoking, and that the aim of our trial is not to prevent the damage brought on by smoking through diet, because the best prevention is to abstain from smoking.

(2) Is administering a flavonoid-rich diet to one arm and a diet poor in flavonoids to the other ethically acceptable? In reality the 'placebo' group will not receive a deficient diet, but a diet normally rich in fruit, vegetables and vitamins. Rather we will take ensure that the experimental group receives certain varieties of fruit and vegetables that are especially rich in flavonoids.

    This issue points to a more general ethical problem valid for all RCTs: in what conditions is it right to administer a new drug to a group of patients and a drug already in use or a placebo to another? The main condition is defined as the 'uncertainty principle' or 'equipoise' (Freedman 1997), and consists of ensuring substantial uncertainty regarding the real advantage associated with the new treatment. If it were not so, we would be committing an injustice towards the control group, pointlessly submitting it to a treatment we know to be less effective. It is also obvious that administering a 'placebo' (an inactive compound) would not be ethical if an effective treatment for that disease already exists (Rothman and Michels 1994).

(3) Is it ethically acceptable to involve only men? The scientific basis for this decision is that women's urine contains only modest quantities of exfoliated cells, and in addition it is not easy to find women who are heavy smokers of air-cured tobacco. The restriction to men is acceptable if we clarify its scientific assumptions and if we suppose the results we will find will be equally applicable to women, as there is no reason to assume they would not be.

Diet preparation. A dietician will be involved in the study to identify the most desirable foods from amongst those containing high and low flavonoid levels. Special recipes will be prepared and supplied to the participants of both groups (Pensiero and Oliveria 1998).

Randomization. A total of 120 volunteers will be recruited and will be randomly assigned to four groups corresponding to four different diets (two rich in flavonoids and two normal). No stratification will be applied, as all the subjects are male smokers. Randomization will be performed at the Epidemiology Centre, where a single operator will randomly assign the volunteers to the four groups. To do this, a number will be assigned to each volunteer. Randomization will be based on generating random numbers with a computer: the first 30 numbers generated will receive the first diet, the following 30 the second diet, and so on. To guarantee blindness in the subsequent course of the study, only the project manager will know the corresponding diet for each group. The only other exception will be the cook, who will obviously know which diets are richer in flavonoids.

Participants' involvement and motivation. The future participants will be taught how to prepare the food at a residential course with a professional cook. We have already organized such an event in the context of the pilot study for the international multicentre EPIC study (Pisani *et al*. 1997). The course will take place in a restaurant, will consist of four sessions with 30 participants each, and will be of a practical nature. Different recipes will be taught during the four sessions, but two sessions will be for the treatment group and two for the control group. As far as the

flavonoid content of the recipes is concerned, the whole staff (excluding the cook) and all the participants will operate blindly.

The run-in phase and the qualifying visit. An introductory phase (run-in) will precede the trial; immediately after the residential course there will be a qualifying visit during which the 120 volunteers will be asked to fill in an introductory questionnaire on dietary habits. Inclusion criteria will be verified and an informed consent form will be submitted to the volunteer. The volunteers will be asked to collect their urine and to modify their diet for the following 2 days. At the end of this introductory period they will be interviewed using the 24 h recall method. Only those subjects with a good adherence to the protocol at the end of the run-in will be included in the trial.

Sample collection and storage: laboratory techniques (technical details). Urine samples will be collected once a week for a month and centrifuged to collect exfoliated bladder cells and to measure DNA adducts. A blood sample will be taken to determine the *N*-acetyltransferase Glutathione-S-transferase M1 and catechol-*O*-methyl-transferase (COMT) genotypes in the lymphocytes. We have previously demonstrated that slow acetylators have higher concentrations of DNA adducts in exfoliated bladder cells than fast acetylators (Vineis *et al.* 1994).

N-Acetyltransferase inactivates carcinogenic aromatic amines, especially 4-aminobiphenyl, which has been shown to form the main DNA adduct in smoker's exfoliated bladder cells (Talaska *et al.* 1991). A further urine sample will be collected from each participant 1 year after the end of the trial to measure DNA adducts in exfoliated bladder cells again. At this point a questionnaire will be submitted to evaluate changes in smoking and eating habits and their effects on the formation of DNA adducts.

The 24 h urine samples will be collected in polyethylene bottles, which will then be kept at 4°C in the volunteers' homes. The samples will be collected weekly by the dietician, who will also verify protocol compliance. A part of each sample will be used to prepare the exfoliated bladder cells.

The remainder of the urine samples will be kept at −80°C and analysed within 6 months. Urine extracts will be prepared using Bond Elut cartridges, according to the procedure developed at the IARC laboratories (Malaveille *et al.* 1996). DNA adducts in the exfoliated bladder cells will be measured using the ^{32}P-postlabelling method (Talaska *et al.* 1991). The urine's antimutagenicity will be measured in *Salmonella typhimurium* TA98 in the presence of S9 cells from the liver of a Sprague-Dawley male rat treated with Aroclor 1254.

The inhibiting effect of the urinary extracts on the mutations induced by PhlP will be measured in a solution containing 20 ng of PhlP and up to 10 µl of urinary extract. The amount of flavonoids in a dimethyl sulphoxide (DMSO) solution of urinary extracts will be measured by analysing the concentration of total flavonoids using Singleton and Rossi's spectrophotometric procedure, employing gallic acid for the calibration curve (Malaveille *et al.* 1996).

Analysis of the specific flavonoid quercetine and its metabolites will be performed using high performance liquid chromatography with electrochemical detectors.

Duration. The trial will last for 1 month. Since bladder surface cells last approximately 2 weeks, a total replacement of these cells should occur in 1 month and the adducts in the exfoliated cells will be representative of the events occurring during the trial.

Contamination, monitoring and interviews. Contamination of the control group could occur only because of random variations in dietary habits that could cloud the differences between the two groups. A systematic contamination of the control group by a flavonoid-rich diet is unlikely, as there will be four different diets taught to four different groups. Except for the cook, both the teachers and the participants will be blind with regards to the flavonoid content of each diet. Dietary habits will be monitored every 2 days with a phone interview, using the 24 h recall method. An expert dietician will pay weekly visits to the participants' homes to monitor the actual composition and preparation of food.

No kind of co-intervention has been hypothesized, it being unlikely that the volunteers taking flavonoids would also assume other protective diets that would differentiate them from the control group.

Drop-outs. Our interest lies in the comparison between the groups undergoing the trial; selection operating before recruitment will have no effect on the validity of the comparison. The run-in phase will guarantee the recruitment of a homogeneous smoking population who will show a certain standard of compliance to the protocol for food preparation. During the experiment drop-outs may occur, although these tend to be minimized by the strong motivation shown in the past by blood donors.

A drop-out rate around 10% has been forecast. For this reason, and because of the loss of a few participants during the introductory run-in phase, we will begin with 120 people, with the aim of having at least 100 volunteers who will complete the trial.

Blindness. The participants and the whole staff (except for the cook) will be 'blind'. All the subjects involved in the study will only be informed of the fact that four diets will be compared, without specifying their characteristics.

Sample size. If we expect a modest difference between the two arms of the trial, it is especially important to calculate the study's statistical power, that is the probability of identifying a statistically significant difference if this exists in reality. Power depends on three factors: the size of the study, the frequency of the observed result, and the size of the difference between the arms of the study. Small studies highlight big differences for frequent results, and big studies highlight small differences for rare results. Therefore, when we plan the study we must have an idea of what difference we expect to find between the two arms for a given result, and on this basis establish a level of power (usually 80%, that is the probability of finding an effect of at least eight every 10, if this exists in reality) and the size necessary to achieve it. A common mistake is to conduct small, and therefore uninformative, studies that cannot exclude a modest difference between the arms, which could be clinically relevant.

In the pilot study, when the levels of DNA adducts in exfoliated bladder cells had been divided into tertiles (obtained by dividing the observations in three equal

parts after having sorted by increasing value), the urine's antimutagen activity was of 21.3 revertants/ml of urine in the tertile with the lowest level of DNA adducts, and 12.75 revertants/ml in the tertile with the highest level (a 40% reduction).

In the second pilot study patients with a high consumption of fruit and vegetables showed a median concentration of 4-ABP adducts in bladder biopsies of 1.0 adducts/10^8 base pairs, against a value of 4.0 adducts/10^8 base pairs found in subjects who did not consume fruit or vegetables, with a SD of 3. Very similar results have been reported by Lin *et al.* (1994).

The size of the study was estimated using the following assumptions:

(1) An α error of 0.05, and a β error of 0.20 (power = 80%), where the α error expresses the probability of a chance finding (type I error) due to chance fluctuations related to sampling, and the β error (type II error) is 1-power.
(2) An average expected concentration of 4-ABP DNA adducts of around 40% lower in subjects with higher consumption of flavonoids (with a reduction from 4 to 2.4 adducts/10^8 base pairs).

On this basis it is necessary to recruit at least 50 subjects per arm. The formula used is (Armitage and Berry 1987):

$$n > 2([Z_{2\alpha} + Z_{2\beta}]^* SD/D_0)^2$$

where D_0 is the expected difference between the two arms. To allow for both exclusions and poor compliance to the protocol, we will recruit 120 subjects.

Statistical analyses. Statistical analyses will include:

(1) The relationship between urinary flavonoids, urinary antimutagen activity and the formation of DNA adducts, separately for both groups.
(2) Comparison regarding the quantity and type of DNA adducts found in exfoliated bladder cells between subjects treated with the experimental diet and subjects treated with the control diet; this comparison will be implemented with methods based on confronting averages and also non-parametric methods (odds ratios and relative confidence intervals, number needed to treat; see appendix 2).
(3) The relationships between *N*-acetyltransferase, glutathione S-transferase and COMT genotypes and the frequency and level of DNA adducts in both groups.

Adjustments for age will be made if necessary.

Further analyses will be based on comparing weekly urine samples (testing the trend of DNA adduct levels separately for both groups) with urine samples collected 1 year after the end of the trial, to compare the levels of DNA adducts in both groups with those encountered before the study.

Analyses will be conducted according to two criteria:

(1) According to the 'intention-to-treat' approach, comparing all the subjects originally randomized from whom we have collected urine, including subjects who did not show complete compliance with the treatment protocol.

(2) According to the 'per protocol' approach, limiting the analyses only to subjects who completed the trial in total compliance with the protocol.

The intention-to-treat analysis is considered more correct because it is more conservative and pertinent to real 'daily' working conditions. Limiting ourselves to those who got to the end of the study – thus eliminating those who dropped out because of toxic manifestations of the drug or because they didn't like the treatment (in our case a diet) – can provide highly distorted and even dangerous information for the introduction of a drug or preventive action in practice.

The main point we wish to formally test with our trial is a 40% reduction of adducts in exfoliated cells in the treated group compared with the control group. However, other relationships will also be considered. The chance of finding a statistically significant association grows with the number of studied associations; for example, if $\alpha = 0.05$ we will find five statistically significant associations in every 100 tested associations. There are two ways to counter this problem. In the first, some kind of correction (e.g. Bonferroni or Holm) is applied to the analyses. Bonferroni's approach consists of dividing the nominal significance level by the number of statistical tests implemented (i.e. 0.05 becomes 0.0005 if we implement 100 comparisons). However, such an approach is highly questionable. The second approach, considered more convincing by many, is based on the different *a priori* probabilities assigned to different suppositions. Formally this approach consists of a 'Bayesian' statistical analysis; in practice, it means that finding a statistically significant association in the main comparison being studied by the experiment has a different relevance to the finding of an unexpected association over 100 comparisons.

Subgroup analysis poses an even weightier problem. If our observations are divided into n subgroups (in our case based on age, genotype, diet, etc.), the chance of incurring false positive results (type I error) in multiple comparisons increases. We could (erroneously) invest the differences found with biological significance. For instance, we may observe a statistically significant difference between groups with different diets, but only amongst slow acetylators: although the observation is biologically plausible, it could also be due to chance as a consequence of multiple comparisons. In this case we will have to decide *a priori* which are the associations that have biological significance and that we want to formally test.

This is a particular problem in pharmacological trials. Given the uncertainty about the actual effectiveness of a drug, we are not always able to hypothesize about the subgroups in which the drug could be more effective. In addition, the consequences of a type I error can be particularly serious (e.g. mistakenly including some patient subgroups, such as those at more advanced stages, among those for which the drug is especially indicated). In addition to the Bayesian solution (relating statistical tests to *a priori* suppositions), another obvious solution is to repeat the experiment, maybe increasing the size of the subgroup of special interest.

Finally it is worth mentioning the issue of interactions, where the effect of the treatment is different in different variable strata, such as age or stage of disease. Testing interactions involves especially sophisticated statistical analyses. One of the biggest problems lies in the swift decrease in statistical power when an interaction is studied. If we need n subjects to study the main effect, to observe an interaction we need approximately $4n$ subjects. In the context of multivariate

analysis, the significance of the interactive term suggests the presence of interaction:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \gamma X_1 X_2 + \varepsilon$$

where y = result, X_1 = treatment, X_2 = stage, X_1X_2 = interactive term, γ = coefficient of regression of the interactive term, and ε = residual error.

Confounding potential. Numerous micronutrients contained in food are related to each other and this can cause a potential confounding effect, an error consisting of falsely attributing certain effects to one compound when they are actually due to another compound or the combination of the two. This problem should be overcome by analysing and interpreting the experimental diets through a foods/nutrients matrix. This will make it possible to identify eventual confounding and the kind of statistical analyses that will assess the contribution by flavonoids while adjusting for the effects of other elements of the diet.

Validation of biomarkers. An aspect that is clearly relevant to the discussion of measurement is timing: any inference about the meaning of biomarker measures should be strictly time-specific, since time influences the results in several different ways.

The major components of biomarker variability that affect the design of epidemiological studies are variability between subjects (intersubject) and within subjects (intrasubject), and variability due to measurement errors. The impact of these three categories of variability on the biomarker response can be represented by the linear model (Taioli *et al*. 1994):

$$y_{ijk} = u + a_i + b_j + e_{ijk}$$

where y_{ijk} is the marker response for subject i at time j and replicate measurement k, u is the true population mean response, a_i is the offset in mean response for subject i (assumed to be normally distributed with mean = 0 and variance = s_i^2; this variance represents the extent of intersubject variability), b_j is the offset in response at time j (assumed to be normally distributed with mean = 0 and variance = s_j^2; this variance represents the extent of intrasubject variability), and e_{ijk} is the assay measurement error (normally distributed with mean = 0 and variance = s_{ijk}^2) (Taioli *et al*. 1994). The normality of distribution, assumed in the model, must be verified. In fact, many biomarkers have distributions that are far from being normal; normalization can be achieved through an appropriate transformation, for example log transformation.

The model is based on a linear (additive) assumption, which assumes that measurement errors are independent of average measurements. Such an assumption must be verified case by case, for example by checking whether errors are correlated with the mean.

Intersubject variability in marker response may derive from factors such as ethnic group, gender, diet or other characteristics. For example, vitamin levels vary according to diet-related characteristics such as weight (thinner subjects are expected to have healthier diets and higher vitamin consumption) (Vineis, unpublished data). In this particular example, age was not associated with vitamin levels, although age has always to be considered to be an important source of intersubject variation.

Similarly, the marker response may vary within the same subject over time due to changes in diet, health status, variation in exposure to the compound of interest (for dietary items season is an important variable), and variation in exposure to other compounds that influence the marker response. Intraindividual seasonal variation in levels of micronutrients has been repeatedly observed.

Biological sampling variation is related to the circumstances of sample collection. For example, hyperproliferation of colonic cells is extremely variable in different points of the colon mucosa. Therefore, not only intrasubject variation over time is important, due to variable exposure to agents that induce cell proliferation, but measurements are also strongly influenced by how and where the mucosa is sampled. For example, one study (Lyles *et al*. 1994) estimated that 20% of the variability of the rectal mucosa proliferation index (measured by nuclear antigen immunohistochemistry) is due to the subject, 30% to the biopsy within the subject, and 50% to the crypts within a biopsy. In other words, as much as 80% of the variation is related to sampling.

Variations in laboratory measurements can have many sources. There are two main classes of laboratory errors: those that occur between analytical batches and those that occur within batches. A study designed to assess the different sources of laboratory variation was reported by Taioli *et al*. (1994), using the model described above. In one experiment, they drew blood from five subjects three times in three different weeks ($n = 5, k = 3, j = 3$) in order to measure DNA–protein cross-links. The results indicated that variations between batches were quite important and were larger than variations between subjects. An interaction between intersubject variation and batch variation was also suggested.

Methodological issues differ according to the category of biomarker. Table 1 shows how variations in the observed data can be affected by biomarker type. Intraindividual and sampling variations are listed due to the extent of their

Table 1. Types of veriability by biomarker category

Biomarker category	Intraindividual variation	Biological sampling variation
Internal dose (blood)		
Hormones	Yes (diurnal variation)	No
Water-soluble nutrients	Yes (short half-life)	No
Organochlorines	No (long half-life)	No
Biologically effective dose		
Peripheral white blood cells	Yes (half-life of weeks to months)	No
DNA adducts in exfoliated urothelial cells	Yes (half-life of months)	Yes
Early biological effects		
Lymphocyte metaphase chromosome aberrations	More or less stable	?
Somatic cell mutations (e.g. glycophorin A)	Probably low	No (?)
Intermediate markers		
Cervical dysplasia	Yes	Yes
Colonic hyperproliferation	Yes	Yes
Genetic susceptibility		
Genotype assay	No	No
Non-inducible phenotype	No	No
Inducible phenotype	Yes	No
Tumour markers	Yes	Yes

influence on actual measurements for most markers. For each category a few examples out of many that could be relevant are given.

Ethical issues

Recruitment of individuals. The ethical implications involve at least three issues. The first is fundamental and consists of the application of so-called equipoise: the trial is ethically acceptable only if there is a reasonable margin of doubt about the effectiveness of the treatment under trial. If we knew the treatment to be more effective than the placebo, the trial would not be acceptable. This is particularly important for clinical treatment, but is less rigorously applied to preventive actions. The second and third issues are both of a substantial and formal nature (unfortunately at times the latter takes over), and involve obtaining informed consent and approval by an ethical committee. These aspects are regulated by the law in most countries.

In addition to providing information on the purposes of the trial, the informed consent form must specify that it is conducted in a condition of doubt (when we don't know if the new drug/preventive action will be useful) and following a probabilistic principle (i.e. even if the new drug/preventive action is more effective, such a result cannot be guaranteed to the individual participants). The form must also detail the risks the patients will be exposed to.

A difficult problem is what rights a subject has as far as knowledge of the results are concerned. A preliminary design issue is whether samples will be anonymous, so that individuals can no longer be recognized at the analysis stage. If samples are not anonymous, the prevailing opinion is that the informed consent form should ask subjects whether they want to know their personal results, just the collective results, or whether they would prefer not to know the results at all. Where the subjects wish to know their personal results, a strategy for positive results (e.g. *BRCA1* mutated) should be defined beforehand (e.g. genetic counselling). The greatest difficulties arise with tests that give continuous results, such as a metabolic phenotype, or results that are not predictive – at the individual level – of the outcome, such as chromosome aberrations. Although policies vary, the lack of individual utility of the result should be made very clear to the study subjects.

A final aspect to take into account is the need for confidentiality, which is regulated by law in most Western countries. The purpose of such laws is to avoid 'sensitive' data regarding the subjects becoming known to third parties (such as insurance companies), who may use it against the subjects' wishes.

During most clinical trials a committee of external experts is formed, known as the data and safety monitoring committee (DSMC), with the task of monitoring both the handling of data (i.e. confidentiality) and the evolution of the trial from the subjects' security (side-effects) point of view. The DSMC has the right to violate blindness to establish whether there is an excess of undesired events in one of the two arms. In addition there are suspension rules (partly based on the statistical significance of differences) aimed at interrupting the trial if one of the two treatments is obviously more effective (see Royall 1991). The DSMC has various options open to it; it can encourage the continuation of the study, it can ask for a modification of the protocol or a more thorough data analysis, or it can ask for suspension of the trial.

In our case the recruitment issues will addressed as follows:

(1) Volunteers will be recruited from a blood donors' association that has previously cooperated in epidemiological studies. These volunteers have already shown a high level of understanding and of compliance to research protocol.
(2) Volunteers will receive a brochure with a description of the purposes of the study and information against smoking. Both diets will be presented as containing unknown factors that may protect the bladder epithelium from the effects of smoking. It will be clearly stated that real protection involves giving up smoking. Volunteers will also be informed that they will not personally benefit from participating in the trial and that they will be able to withdraw from the trial at any time. They will also sign an informed consent form.

Conflicts of interest and sources of sponsorship. When research is financed by industry, various conflicts of interest can arise, including the distortion of results to emphasize the effectiveness of the treatment, interference from national and international agencies in the release of unwelcome data (for example the recent case between Philip Morris Inc. and the International Agency for Research on Cancer), as well as more banal, but not less important, issues such as the tendency to study areas that are economically advantageous but not necessarily relevant to patients. In relation to this last point, a few years ago the British National Health Service assigned a quota of its fund to research and development activities to avoid subjects of great relevance being 'orphaned' (as has happened with international research on malaria).

The study we are examining was financed by the Italian Association for Cancer Research and the World Cancer Research Fund, two non-profit organizations.

Other aspects: the duration of the observation, surrogate results and special plans

There are still a few aspects to consider: the duration of the follow-up stage, the kind of results to take into consideration, and the possibility of improving the planning of the study with specific strategies.

The follow-up to the study we are planning lasts for 1 year, since we intend to collect urine samples and interview the volunteers 1 year after the trial's conclusion. We must ask ourselves whether this is a sufficiently long period to identify steady differences in the results. If the follow-up is too short it may miss the emergence of differences between the two arms for events with a late onset or may highlight temporary changes of little clinical significance.

A further question is whether it is correct to choose an intermediate indicator (DNA adducts) as a result, instead of considering the illness we aim to prevent. Measuring an intermediate indicator has the advantage of increasing the power of the study (because we assume it to be more frequent than the final outcome) and shortening its duration, but has the disadvantage of predicting results that are not necessarily of any real relevance to the patients. Many studies on the 'chemoprevention' of tumours (often based, like the study considered above, on the administration of antioxidants) use the formation of DNA adducts or the frequency of oxidation damage to the DNA as the result, instead of the incidence of the cancer

itself, the result of true interest to the patient. The advantage of using surrogate results is smaller size needed to highlight statistically significant differences, but the disadvantage is the indirect and uncertain relationship between the observed effectiveness and the prevention of tumours.

Different types of results have been used in trials on the effectiveness of cancer treatment, from death, the onset of relapses and metastasis, and a reduction in the size of the tumour (especially in phase II studies), down to surrogate weak results such as the normalization of markers (e.g. carcinoembryonic antigen or prostate specific antigen). Effects demonstrated using surrogate or intermediate results have to be considered to be provisional, because they do not correspond to the ultimate aim (from the patient's point of view), which is prolonged survival and improvement in the quality of life.

With regard to the design, there are many strategies that make it possible to maximize the effectiveness of the trial. One type of design is block randomization. In such studies a block is a group of patients of a predefined number and a predefined proportion of treatment and placebo allocations. The size of each block must be an exact multiple of the number of groups undergoing treatment. For two treatments we can choose a block design of four patients, with treatments randomly assigned within the block. Stratification is often used in association with block design: treatments are allocated to blocks in various strata of a confounding variable to make the control of the confounding more efficient and to increase the power of the study. This is a convenient design, especially when a small number of patients are to be recruited and when only one or few important (able to substantially influence prognosis) confounding variables are present.

Acknowledgements

This work has been made possible by grants from the World Cancer Research Fund and the Associazione Italiana per le Ricerche sul Cancro to P. Vineis.

References

Research protocol

Armitage, P. and Berry, G. 1987, *Statistical Methods in Medical Research*, 2nd edition (Oxford: Blackwell Scientific).

Block, G., Patterson, B. and Subar, A. 1992, Fruit, vegetables and cancer prevention: a review of the epidemiological evidence. *Nutrition and Cancer*, **18**, 1–30.

Cai, Q., Rahan, R. O. and Zhang, R. 1997, Dietary flavonoids, quercetin, luteolin and genistein, reduce DNA damage and lipid peroxidation and quench free radicals. *Cancer Letters*, **119**, 99–108.

Fenech, M., Stokley, C. and Aitken, C. 1997, Moderate wine consumption protects against hydrogen peroxide-induced DNA damage. *Mutation Research*, **379**, Supplement 1, S173.

Fuhrman, B., Lavy, A. and Aviram, M. 1995, Consumption of red wine with meals reduces the susceptibility of human plasma and low-density lipoprotein to lipid peroxidation. *American Journal of Clinical Nutrition*, **61**, 549–554.

Giovannucci, E., Ascherio, A., Rimm, E. B., Stampfer, M. J., Colditz, G. C. and Willet, W. C. 1995, Intake of carotenoids and retinol in relation to risk of prostate cancer. *Journal of the National Cancer Institute*, **87**, 1767–1776.

Hertog, M. G., Hollman, P. C., Katan, M. B. and Kromhout, D. 1993, Intake of potentially anticarcinogenic flavonoids in adults in the Netherlands. *Nutrition and Cancer*, **20**, 21–29.

Lampe, J. W., Martini, M. C., Kurzer, M. S., Adlercreutz, H. and Slavin, J. L. 1994, Urinary lignan and isoflavonoid excretion in premenopausal women consuming flaxweed powder. *American Journal of Clinical Nutrition*, **60**, 122–128.

Lin, D., Lay, O. J., Bryant, M. S., Malaveille, C., Friesen, M., Bartsch, H., Lang, N. P. and Kadlubar, F. F. 1994, Analysis of 4-aminobiphenyl–DNA adducts in human urinary bladder and lung by alkaline hydrolysis and negative ion gas chromatography-mass spectrometry. *Environmental Health Perspectives*, **102**, 11–16.

Malaveille, C., Hautefeuille, A., Brun, G., Vineis, P. and Bartsch, H. 1992, Substances in human urine that strongly inhibit bacterial mutagenicity of 2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (PhIP) and related heterocyclic amines. *Carcinogenesis*, **13**, 2317–2320.

Malaveille, C., Hautefeuille, A., Pignatelli, B., Talaska, G., Vineis, P. and Bartsch, H. 1996, Dietary phenolics as anti-mutagens and inhibitors of tobacco-related DNA adduction in the urothelium of smokers. *Carcinogenesis*, **17**, 2193–2200.

Miller, A. B., Berrino, F., Hill, M., Pietinen, P., Riboli, E. and Wahrendorf, J. 1994, Diet in the etiology of cancer: a review. *European Journal of Cancer*, **30A**, 207–220.

Pensiero, L. and Oliveria, S. 1998, *The Strang Cookbook for Cancer Prevention: A Complete Nutrition and Lifestyle Plan to Dramatically Lower Your Cancer Risk* (New York: EP Dutton).

Pisani, P., Faggiano, F., Krogh, V., Palli, D., Vineis, P. and Berino, F. 1997, Relative validity and reproducibility of a food frequency questionnaire for use in the Italian EPIC centers. *International Journal of Epidemiology*, **26**, Supplement 1, S152–160.

Rosso, S., Patriarca, S., Vicari, P. and Zanetti, R. 1993, Cancer incidence in Turin: the effect of migration. *Tumori* **79**, 304–310.

Talaska, G., Schamer, M., Skipper, P., Tannenbaum, S., Caporaso, N., Unruh, L., Kadlubar, F. F., Bartsch, H., Malaveille, C. and Vineis, P. 1991, Detection of carcinogen–DNA adducts in exfoliated urothelial cells of cigarette smokers: association with smoking, hemoglobin adducts, and urinary mutagenicity. *Cancer Epidemiology, Biomarkers and Prevention*, **1**, 61–66.

Trichopoulou, A. 1995, Olive oil and breast cancer. *Cancer Causes and Control*, **6**, 475–476.

Vineis, P. and McMichael, A. J. 1996, Interplay between heterocyclic amines in cooked meat and metabolic phenotype in the etiology of colon cancer. *Cancer Causes and Control*, **7**, 479–486.

Vineis, P., Bartsch, H., Caporaso, N., Harrington, A. M., Kadlubar, F. F., Landi, M. T., Malaveille, C., Schields, P. G., Skipper, P., Talaska, G. and Tannenbaum, S. R. 1994, Genetically based *N*-acetyltransferase metabolic polymorphism and low-level environmental exposure to carcinogens. *Nature*, **369**, 154–156.

Visioli, F. and Galli, C. 1994, Oleuropein protects low density lipoprotein from oxidation. *Life Sciences*, **55**(24), 1965–1971.

Weisburger, J. H. 1994, Practical approaches to chemoprevention of cancer. *Drug Metabolism Reviews*, **26**, 253–260.

Main article

Doll, R. 1998, Controlled trials: the 1948 watershed. *British Medical Journal*, **317**, 1217–1220.

Freedman, B. 1987, Equipoise and the ethics of clinical research. *New England Journal of Medicine*, **317**, 141–145.

Lyles, C. M., Sandler, R. S., Keku, T. O., Kupper, L. L., Millikan, R. C., Murray, S. C., Bangdiwala, S. I. and Ulshen, M. H. 1994, Reproducibility and variability of the rectal mucosal proliferation index using proliferating cell nuclear antigen immunohistochemistry. *Cancer Epidemiology, Biomakers and Prevention*, **3**, 597–605.

Noseworthy, J. H., Ebers, G. C., Vandervoort, M. K., Farquhar, R. E., Yetisir, E. and Roberts, R. 1994, The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology*, **44**, 16–20.

Piantadosi, S. 1997, *Clinical Trials. A Methodological Perspective* (New York: John Wiley and Sons).

Rothman, K. J. and Michels, K. B. 1994, The continuing unethical use of placebo controls. *New England Journal of Medicine,* **331**, 394–398.

Royall, R. M. 1991, Ethics and statistics in randomized clinical trials. *Statistical Science*, **6**, 52–88.

Taioli, E., Kinney, P., Zhitkovich, A., Fulton, H., Voitkun, V., Cosma, G., Frenkel, K., Toniolo, P., Garte, S. and Costa, M. 1994, Application of reliability models to studies of biomarker validation. *Environmental Health Perspectives*, **102**(3), 306–309.

The Cochrane Library 2000, *The Cochrane Collaboration and Update Software*, Issue 1.

Appendix I: Glossary

Controlled Endowed with a control group

Blindness A procedure that avoids distortions (systematic errors) related to the awareness of what drug each patient is taking and the consequent psychological

	expectations, which can change the accuracy of the results obtained within the different groups under observation. An elegant example of how a lack of blindness can condition results was reported by Noseworthy *et al*. (1994): within a randomized trial on cyclophosphamide and multiple sclerosis both 'blind' and 'not blind' neurologists were used to collect the results; a statistically relevant result was found only in the second group
Chance error	Measuring error due exclusively to chance factors
Co-intervention	Occurs when those receiving the drug on trial are also receiving another concurrent treatment that is not part of the protocol
Confounder	A variable associated with the treatment on trial, and predictive of the result. In the case under scrutiny, we could mistakenly conclude that flavonoids reduce DNA damage while this effect is in reality ascribable to other constituents of flavonoid-rich diets. Confounding can be eliminated with randomization, but only if we can separate the treatment on trial from the potential confounder (i.e. separate the two different types of foods). Another cause of confounding is pure chance, when, because of the small size of the study, subjects of the treatment group differ from those belonging to the control group in some factor predictive of the observed result, for instance the seriousness of the illness (in spite of randomization)
Contamination	Where the treatment on trial is also taken by the control group (e.g. because their doctor has prescribed it); this rarely occurs with new drugs, but can easily happen with preventive actions and screenings
Dose-finding	The purpose of phase I studies is mainly to find the right dose to use in the trial. Techniques are available that allow establishment of the maximum tolerated dose, such as the Fibonacci layout (see Piantadosi, 1997)
Drop-outs	Those who for whatever reasons (side-effects, low motivation, the onset of other pathologies) abandon the trial
Estimate precision	The degree of certainty reached with the estimate. It depends on the dimensions of the study and is measured by the confidence interval
Follow-up	Consists of following the subjects recruited for some time to collect relevant data on the results. It must be long enough to allow for the highlighting of all clinically relevant results and must be conducted

	with the same accuracy and procedures for all the groups being compared
Intention-to-treat	Analysis of results based on the patients actually randomized and not on the treatment received. It is a 'conservative' analysis that does not exclude drop-outs and patients with a low compliance, and so gives realistic estimates of the treatment's efficacy
Placebo	A compound used for a comparison treatment devoid of any pharmacological effectiveness. A placebo is used to reduce systematic errors when the evaluation of the results could be influenced by the doctor or the patient knowing that the control group has not received any treatment
Randomization	Chance assignment of the treatment under scrutiny or the placebo treatment; the randomization process uses algorithms for the extraction of random numbers by computer. Randomization is the simplest and most practical method to avoid interference by unrelated variables with the cause–effect relationship being studied
Reproducibility	The degree of approximation to which the same estimate can be obtained in different experiments
Run-in	Phase in which the recruitment methods are defined and the trial's participants' eligibility and compliance are assessed
Systematic error or bias	Measuring error due to mistaken planning, faulty implementation or wrong analysis of the study. Common systematic errors are due to (1) the subject selection procedure, e.g. if healthier patients are selected for one of the comparison groups; (2) post-recruitment exclusion, which can cause powerful distortions in that it can be influenced by the kind of treatment and by the observed results; (3) a distorted evaluation of the results caused by the lack of blindness in the observation process; (4) retroactive definitions (when treatment and results are redefined at the end of the study on the basis of what has been observed)

Appendix 2: Statistical analyses of the results of a randomized trial

Calculating the odds ratio (OR) is a common way to analyse data. For example, let's suppose we are able to randomize 94 subjects in our trial and that the final distribution is as shown in appendix table 1. The OR is calculated as follows:

$$\begin{aligned} \mathrm{OR} &= \mathrm{A} \times \mathrm{D}/\mathrm{B} \times \mathrm{C} \\ &= 20 \times 21/26 \times 27 = 0.6 \end{aligned}$$

Appendix table 1. Results from 94 randomized subjects

	Treatment	
Level of adducts	Flavonoids	Other
Above the median*	20 (A)	26 (B)
Below the median	27 (C)	21 (D)
Total	47	47

* DNA adducts

Appendix table 2. Fictional results of effect of treatment on prevalence of DNA adducts

	Population	
	Non-smokers	Heavy smokers
Prevalence above the median in the absence of treatment: U	5%	55%
Relative prevalence reduction after treatment: (U − T)/U	40%	40%
Prevalence among the treated patients: T	3%	33%
Absolute reduction in prevalence above the median: U − T	2%	22%
Number needed to treat to prevent high prevalence: 1/(U − T)	1/0.02 = 50	1/0.22 = 4.5

In this case the 'crude' OR (i.e. not standardized by covariates), is equal to 0.6, and shows that subjects who followed the flavonoid-rich diet had a 60% chance of having a high level of adducts (above the median) compared with subjects following other types of diet. In other words, the treatment was associated with a reduction in relative risk (RRR) of 40%. This estimate is not adjusted for any covariate, as it does not take into account the fact that the groups can differ in the distribution of variables having prognostic significance (in spite of randomization). Information regarding correction for covariates can be obtained from more specialized sources. It is important to stress that covariates should be specified *a priori*, and that analysis by covariates is always appropriate because covariates can act as confounding variables.

The OR is a *relative* measure of efficacy. It must be used with care in a clinical context, as it does not give any idea of the *absolute* frequency of the events we want to avoid (relapses, deaths). This is illustrated in the following fictional example given in appendix table 2. The relative reduction of the prevalence of adducts due to treatment with flavonoids is the same in the two groups of patients (40%), but the prevalence is much higher among smokers (55%) than non-smokers (5%). If the two groups are treated with flavonoids, the 40% gain is expressed by an absolute reduction in prevalence of 33% and 3%, respectively. This means a risk reduction of 22% and 2% in the two groups. It is obvious that the efficacy of the treatment, although the same in relative terms, shows up differently in the two groups of subjects. This becomes even clearer if the reciprocal of the measures of absolute efficacy are calculated: $1/0.22 = 4.5$ and $1/0.02 = 50$ for heavy smokers and non-smokers, respectively. These values (4.5 and 50) represent the number of subjects we need to treat to obtain one preventive success (number needed to treat; NNT). It is evident that the NNT carries much more information than the simple relative benefit.