

# The Confusing World of Clinical Trials

*A Guide for Patients and Families*



**WRITTEN BY GARY CUTLER, PhD  
AND INMACULADA ABAN, PhD**

*MSAA Monograph #2*

**MSAA**<sup>TM</sup>  
MULTIPLE SCLEROSIS  
ASSOCIATION OF AMERICA

This monograph is dedicated to Jack Burks, MD, by MSAA and the monograph's sponsor, Bayer HealthCare Pharmaceuticals. Dr. Burks currently holds the position of chief medical officer for MSAA. Dr. Burks is an internationally recognized expert in the field of MS and has a long-standing commitment to MSAA and its goal to provide individualized patient-focused care.

This is why dedicating this monograph to Dr. Burks is so appropriate; he has spent a lifetime helping those with MS to better understand the disease and obtain information to better manage it. Providing valuable information is the focus of this monograph.



**Bayer HealthCare**  
Pharmaceuticals

*Funding for this monograph has been generously donated through a grant from Bayer HealthCare Pharmaceuticals. The staff and Board members of MSAA would like to express much appreciation for this kind and purposeful gift.*

Gratitude also goes to the authors of this monograph, Gary Cutter, PhD and Inmaculada Aban, PhD. Other individuals involved with the editing of this monograph include: Dr. Jack Burks, Robert Rapp, Andrea Borkowski, and Susan Wells Courtney.

---

Copyright © Multiple Sclerosis Association of America, 2007. All rights reserved. This booklet is protected by copyright. No part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission from MSAA.

# The Confusing World of Clinical Trials

## *A Guide for Patients and Families*

Written by Gary Cutter, PhD  
and Inmaculada Aban, PhD

**R**arely a day goes by without hearing the results of a new trial that has changed the way we think about a treatment, or confirmed what we already know. We now learn of clinical trial results so frequently, that we often believe only what we want to believe, reacting with skepticism and disbelief. A common response is, “Oh no, don’t tell me something else is no longer true?”

Why are clinical trials so confusing and why don’t they seem to answer our questions? The answer is interesting and is partly because of the hype we put into health issues today. We want the best care, with no risks. We want only successful treatments. We find testimonials and infomercials running 24 hours a day on various TV channels, telling us what works and how good it is.

Some patients looking for a cure may even believe in “new-trial” data and are willing to consider almost anything new. With so much information available to the consumer, how are effective treatments differentiated from those that are ineffective and possibly damaging to our health? The experts must look to clinical trials for answers; how well a trial is planned, conducted, and analyzed, will determine the



true effectiveness and safety of a treatment.

To follow is a detailed description of clinical trials – what they are and what they are not. The purpose is to point out their necessity and the details required to undertake these often multimillion-dollar endeavors.

You might think that clinical trials were invented by the healthcare industry to feed the media for marketing purposes. This is not true! Certainly they may be used for this purpose, but that is not why clinical trials are conducted.

You might be further surprised to realize that these are not just part of our modern hype,

but there is evidence of very early attempts at clinical trials. The following example talks about comparing two groups of people on different diets in Biblical times – and one might notice that the Atkins Diet is not so new!

Consider this story from the Bible, **Daniel 1:8-16 (605 BC)**, where King Nebuchadnezzar II carries out the first clinical trial. Initially, the king orders that a strict diet of meat and wine be followed for three years. However, four children of royal blood convince Nebuchadnezzar to allow them to exchange “pulse” [bread or vegetables] and water for the required meal.

Daniel, one of the four royal children, resolved that he would not defile himself with the king’s rich food, or with the wine which he drank. Then Daniel said to the steward, “Test your servants for ten days; let the four of us be given [bread or vegetables] to eat and water to drink. Then let our appearance, and the appearance of the youths and servants who eat the king’s rich food of meat and wine, be observed by you.”

The steward tested them for ten days, and it was seen that the four children were better in appearance and fatter in flesh than all those who ate the king’s rich food. Upon seeing this, the steward took away their rich food and the wine they were to drink, and gave them [bread or vegetables] and water. Thus, a decision was made about what the children should eat, based on “trial” results comparing two groups of individuals given different diets.

Our requirements and *standards* for clinical trials are higher today, but the concept of comparing two groups, one getting one treatment and the other getting another as a so-called “control,” is clearly not a new concept. Clinical trials are central to our establishing what works better.

### **HISTORICAL CONTROLS VERSUS CONTEMPORARY CONTROLS**

In 1537, Renaissance surgeon Ambroise Parè was a battlefield surgeon, who in the heat of battle runs out of boiling oil to treat wounds and amputations. He needs to do something immediately, so he mixes a concoction of oil of rose, turpentine, and egg yolk, applying it to the patients he treats for the rest of that day.

One day after this unintentional clinical trial, he notes that the wounds treated with the traditional formula are swollen and extremely painful, while wounds treated with the experimental mixture are not painful. Parè deduces that the new balm is more favorable than the oil usually applied AND VOWS never to use the “standard therapy again.”

Parè has used his “historical perspective” or “historical controls” (using the results of treatment with previous patients) and compared them to his current experience. Such a procedure has problems in that the historical controls are likely to be different.

Consider multiple sclerosis (MS) as the condition of interest and what might result if a series of past patients were used for comparison. Suppose the new treatment was lemonade and the outcome was the number of enhanced areas (showing inflammation) on the patient’s MRI scans. “Enhanced areas” are also known as “contrast enhancing lesions,” or CELs for short.

If I obtained data from a number of patients with relapsing-remitting MS (RRMS), about 40 percent would have CELs on their MRI scans at any point in time. Let’s assume that these patients would be my “historical controls,” never receiving the lemonade treatment. Then, if I selected patients with secondary-progressive

MS (SPMS) for my lemonade trial, I would observe approximately 15 percent of the patients with CELs. By using these historical controls, I could declare my lemonade treatment a success, simply because far more individuals with RRMS have CELs versus individuals with SPMS. This example results in a 62.5 percent reduction in CELs when comparing “historical control” patients versus those treated with lemonade!

Obviously, the real difference is not the lemonade, but rather the historical control group that I used for my comparison. Thus, in clinical trials, we require so-called “contemporary controls.” That is, the most effective com-



***The concept of providing treatments to similar patients, free of bias, is a hallmark of clinical trials.***

parisons involve two or more groups identified in the same way and given their treatments free of biases that could influence the outcomes.

Conceptually, we would like to treat the exact same person with each treatment. To do so, we would need to first give either the experimental or control treatment, then “turn back the clock,” so the patient receives the second treatment at exactly the same point in his or her disease. We would also view the results after the same amount of follow-up time. Of course, comparing two treatments in one person at the same time is impossible, so we take a similar group of patients and split them into two groups. We then follow the two groups forward in a manner that mimics the “turning back of the clock” idea.

**THE CONCEPT OF RANDOMIZATION**

The concept of providing treatments to similar patients, free of bias, is a hallmark of clinical trials. However, none of us can actually talk to a patient and not think about which treatment might be best for them, even when we truly don’t know which treatment is better. This uncertainty we have about two treatments is called “ equipoise,” which is the situation where we really question whether one treatment is better than another. Clinical trials are done when equipoise is present.

Our standards for evidence of treatment effectiveness are much stronger today than they ever have been, requiring us to prove that specific treatments are different with some level of



certainty. A doctor participating in a clinical trial may feel that for certain patients, a certain treatment choice should be given, even though the overall evidence is not complete. For this reason, we are required to assign patients to the different treatment groups of a clinical trial, in such a way that the doctor's belief does not enter into the treatment assignment.

We do this by a process that is similar to tossing a coin. The procedure is called "randomization," and clinical trials utilizing randomization are called Randomized Clinical Trials (RCT). This seems like an unfair way to treat patients and to a degree it is. The assignment does not care which treatment a patient receives! However, it avoids the selection biases

of what the doctor thinks might work better in one patient or another, even if there is no evidence for it. Without such an unbiased assignment process, the same biases as the historical controls (discussed earlier) could result.

While using a random selection process that does not care which treatment an individual patient receives may seem unscientific, please note that great care is taken when defining which patients are initially eligible for a clinical trial. The criteria for who could receive either of the two (or more) treatments are extensive and insure that no patient receives inappropriate treatments. In fact, the extensive consideration of who should be treated (inclusion criteria) and who should not (exclusion criteria) is often far more scientific than any patient would experience in a one-to-one, treatment-decision situation with his or her own physician. Establishing these criteria is often conducted by a group of scientists and is always reviewed by an Ethics Board or Institutional Review Board.

### **THE IMPORTANCE OF USING A PLACEBO**

So far, we have seen that contemporary comparisons are desirable; equipoise (uncertainty about treatment) should be present; and randomization to treatment assignments are key in clinical trials. We now need to think about the treatment being studied. We want to show that the treatment works. The easiest way to show that a treatment is effective, is to show that it works better than if nothing had been done. While doing "nothing" does not sound very ethical, measuring the effectiveness of a treatment typically requires comparing individuals who received the treatment to individuals who

did not receive the treatment.

Many people know of the terms “placebo” or “dummy” treatment. This is a treatment that looks, acts, tastes, or is similar in every way to the comparison treatment, except for the active ingredients. There are several reasons for using placebos. First, doing almost anything in medicine seems to have at least a temporary effect. It has been called the “placebo effect” or “placebo response.” These are real improvements and not just simply patients being fooled. Call it tender loving care or the ability of the body to respond to expected improvements, but they occur in every disease or condition.

When a clinical trial uses a placebo, the results enable us to measure how much improvement or lack of deterioration is due to this placebo. Subtracting the improvement found with the placebo from the improvement found with the purported good or new treatment, enables us to estimate the actual effectiveness of the drug. In using a placebo, we expect that some improvement will occur, allowing the clinical trial to ethically continue because the patient is getting some treatment. (In this case, “treatment” refers to all other care except for the specific active drug or therapy under investigation.)

However, trials with placebos must be carefully considered and must ethically defend the use of an inactive treatment. If patients are denied what is commonly considered standard care, then arguments that “no harm is being done” should be made. Sometimes the argument for using placebos is that such trials often require less time, enabling fewer patients to be exposed to potentially ineffective new drugs. These trials need fewer subjects because it is

easier to see the difference in results of an active drug compared to a placebo, than it would be to compare two active treatments which are already known to have some positive effect.

Furthermore, in using a placebo, we get a better idea of just what side effects and serious adverse consequences are due to the active drug compared to consequences of the disease. In other words, did the patient have a problem with the new treatment or was the problem related to the disease itself?

***Many drugs fail in the clinical trial stages, often because of side effects or serious unexpected results.***

Just because a drug is being tested doesn't mean that it is a good drug. That is the hope, but many drugs fail in the clinical trial stages, often because of side effects or serious unexpected results. Tysabri®, in combination with another drug for MS, had unexpected results. Many doctors and patients expected it to be a successful medication, and it showed exceptionally good treatment effects on MS exacerbations and CELs. But unexpectedly, two patients experienced very serious complications,



## THE CONFUSING WORLD OF CLINICAL TRIALS

with one death, so additional safety measures were taken. Such unexpected findings that occur only in the treatment group change the view of the drug and the complex decisions as to its use.

Returning to placebos and the importance of this class of treatments in clinical trials, we note that there are several forms of placebos. We have stated that they can work to some extent and that using a placebo is better than no treatment at all. But just how do they work? We have noted tender loving care as an explanation, which simply sounds as though patients feel better because someone cares about them. This may be true, but real physiological changes have occurred when participants are given placebos.

In a study of Apomorphin (a drug for Parkinson's disease) published in the journal *Science* (August 2001), the placebo produced changes of 21 percent, compared with 25 percent in the Apomorphin Group (those receiving

the active treatment). These changes were measured on PET (Positron Emission Tomography) brain scans, which are instruments thought to make objective measures of the treatment response. These results suggest that such responses are real and not just psychological or imaginary as commonly thought. However, placebos in and of themselves are not necessarily sufficient drugs. In fact, placebos in a sense were a driving force behind the development of the Food and Drug Administration (FDA).

In 1906, Congress gave the United States' population protection from unknowingly receiving placebos. It was at this time that Congress prohibited labeling medicines with false claims that are intended to defraud the purchaser. Such actions represent a standard that is difficult to prove, but this regulation acts as a stimulus to clinical studies. In the original Food, Drug, and Cosmetic (FD&C) Act of 1906, there is no requirement to disclose ingredients, but the act grew from problems with a largely unregulated industry that was causing numerous public-health problems.

This 1906 act prohibited the sale of "adulterated" and "misbranded" drugs in interstate commerce. However, the act did not prohibit false therapeutic claims, only false claims about what ingredients were included. It disallowed saying that this snake oil would do "x," if indeed the product had no snake oil! In 1912, the Sherley Amendment specifically prohibited false therapeutic claims. One could no longer sell the snake oil to say, for example, that it cured MS. These acts set the stage for the current regulations which include requirements to conduct clinical trials to establish claims of effectiveness (also referred to as "efficacy").



In 1937, 107 people died after taking sulfanilamide, a drug in which deadly ethylene glycol was confused with propylene glycol. This prompted a reaction by Congress who passed the FD&C Act in 1938, mandating that products must be safe (or non-toxic). They stated that labeling must be defined and must provide written, printed, or graphic materials to accompany the product.

In 1941, the FDA was required to analyze and attest the potency and purity of insulin. In 1951, the Durham-Humphrey Amendments gave the FDA the responsibility to clarify which drugs were: habit-forming; not safe except under a practitioner's supervision; or limited to prescription sales as part of the approval of a New Drug Application (NDA). The amendment required the label, "Caution: Federal Law Prohibits Dispensing Without a Prescription." In the late 1950s, thalidomide (a drug used for morning sickness in pregnant women), produced horrific birth defects and again Congress reacted and improved the NDA process to enhance the safety of medications.

Continued evolution of the FDA included the drug amendments of 1962 (Kefauver-Harris), which enhanced the pre-marketing requirements for testing new drugs; mandated "Good Manufacturing Practices;" regulated advertising; required informed consent by patients in the clinical testing process; and imposed an effectiveness requirement prior to NDA approval by the FDA. It is the effectiveness requirement specifically that is central in mandating the need for clinical trials. Effectiveness can only be established through clinical trials.



### **DEFINING THE PHASES OF CLINICAL TRIALS**

Various stages of clinical trials are required by the FDA. The different trial phases are: the preclinical phase (or Phase I), Phase II, Phase III, and Phase IV trials.

Preclinical Trials look for changes caused by the drug and involve basic laboratory investigation and small studies in animals. If the results are positive, the next step is to identify the formulation for dosing in humans; the drug maker must also apply to the FDA for an investigational new drug application (IND Application), which requests permission to begin trials in

humans. The FDA examines the preclinical data and makes a determination (based on safety parameters) as to whether or not the drug company may proceed with patient trials. Thus, the primary objectives of a Phase I Clinical Trial are to (1) identify an effective dose and (2) assess toxicity of a new drug in normal (healthy) volunteers.

In Phase II Clinical Trials, the objectives are to (1) insure that the drug provides some degree of effect and (2) insure safety without too much toxicity in the diseased population. Eligibility for entrance into the trial is carefully defined. Often there is more than one Phase II study for a drug being developed: the initial study is to gain one level of knowledge -- possibly about drug dosage, and a second study is to refine assessments of safety or outcome of the treatment. Phase II studies are often called "proof of concept" studies. Phase III studies are required in the final proof of safety and efficacy of a drug being developed. These Phase III definitive or so-called "pivotal" trials are often large and may take years to conduct. The design of these Phase III trials usually involves periodic assessments of the treatment responses, along with assessments of side effects and/or toxicities.

Phase III trials (or "pivotal" trials) are warranted if a new treatment shows some promise (some degree of effectiveness, possibly with fewer side effects than known drugs). The goal is to establish the effectiveness of the treatment as required by the FDA. Phase III trials usually involve large numbers of patients – hundreds or even thousands of people. Obtaining these numbers of study participants often requires using multiple institutions in several countries. With such large samples of patients, Phase III

trials provide more information about side effects and tolerability of treatments, along with the impact on quality of life.

In Phase IV Clinical Trials (occurring after the treatment has been approved), the objectives are to gain additional knowledge regarding treatment and long-term safety data -- as treatments are prescribed in physicians' practices where the rigor of inclusion and exclusion criteria are often not as carefully followed. These "post-marketing studies" can identify uses that were not specified in the pivotal clinical trials. They can also identify any unexpected outcomes that may occur at such low frequencies that they would not likely be seen in the pivotal trials. Identifying other uses of a drug (for other conditions) is referred to as "off-label" uses, since the licensing of the treatment by the FDA is specific and must be included in the labeling of the drug. Generally, the drug treatment usage is limited to the population studied in the pivotal clinical trials.

### **MULTI-CENTER CLINICAL TRIALS**

While many clinical trials are conducted at different sites, all use the same protocol. This allows the results to be combined, so that the greater numbers give increased statistical "power" to demonstrate effectiveness. Multi-center clinical trials began at the end of World War II and incorporated important new mandates. These mandates arose from the 1947 Nuremberg Code, which was created in response to the unethical medical experimentation on concentration camp prisoners. This code establishes a number of key points for the protection of subjects and patients in clinical trials. These tenets require: a voluntary declara-

tion of consent by trial participants; the right of trial participants to comprehensive information on the nature, purpose, and potential risks of the experiment; the right of trial participants to withdraw from the trial at any time; performance of a trial must be based on anticipated beneficial results; and the risk involved must be proportionate to the social and humanitarian significance of the problem being addressed.

At about the same time as the introduction of multi-center trials, the random allocation of patients to treatments was initiated. As noted earlier, randomization (assigning patients to treatments essentially by the flip of a coin) is pivotal to protection from biasing trials. The first formal use of randomization in clinical trials is attributed to Sir Austin Bradford Hill in the trial of streptomycin treatment of pulmonary tuberculosis in the late 1940s. In this study, he allocated patients with a formal randomization for the first time: 55 patients to streptomycin and bed rest and 52 patients to bed rest alone. This trial helped establish streptomycin for pulmonary tuberculosis with convincing evidence from the concurrent use of controls (those receiving placebo) and active treatments.

On a larger scale, the Poliomyelitis Vaccine Trials in the 1950s were undertaken to examine the efficacy of Salk's vaccine on preventing the occurrence of polio. This United States' trial, sponsored by The National Foundation for Infantile Paralysis ("March of Dimes"), used counties with populations from 50,000 to 200,000 where high rates of poliomyelitis had occurred. The rates in counties were examined between 1946 and 1950, and counties with sufficiently high rates were included in the trial.

***Time has taught us that the clinical trial is an important tool in demonstrating the effectiveness of new treatments and in preventing the use of worthless or harmful treatments.***

This trial produced major changes in public-health policy and led to the fundamental changes in public-health delivery with efforts to immunize all children. Since then, thousands of multi-center clinical trials have been conducted in virtually all diseases and conditions. Time has taught us that the clinical trial is an important tool in demonstrating the effectiveness of new treatments and in preventing the use of worthless or harmful treatments.

## **CONDUCTING A CLINICAL TRIAL**

### ***The Concept of Masking the Treatment***

The logistics of running trials is extremely complex. Developing a protocol and a manual of procedures are necessary to guide the trial. These documents help insure that all of the sites in the trial are working in the same way, using the same definitions and measurements, while following the same rules in evaluating the patients. This type of organization, implementation, and monitoring is usually accomplished by a coordinating center, consisting of a group of individuals who have the responsibility to insure a common protocol, and in the end, provide the collective analyses.

Up to this point, several concepts about trials have been noted: equipoise (the uncertainty that one treatment is at all better than another); randomization to prevent bias; and interim monitoring to insure patient safety. Another



vital component of a clinical trial is the concept of blinding or masking. The terms “blinding” or “masking” mean that the type of treatment (whether active, and possibly which dose level, or placebo), is not revealed to one or more persons who normally would know which treatment is being taken.

Additionally, trials may be single-blinded, double-blinded, and even triple-blinded. In single-blind trials, the patient or clinician is blinded or masked to the treatment, while in double-blind trials, both patient and clinician are blinded. In triple-blind trials, patient, clinician, and statistician are blinded to the treatment.

Double-blind trials are the most common. As noted, in a double-blind (or masked) trial, neither the patient nor the clinician knows

whether the treatment is the active drug or the control (placebo). As stated earlier, this is done to prevent bias. If the patients knew they were on a placebo, they would likely be disappointed, possibly report no improvement in their disease, and may even be discouraged enough to drop out of the trial. If physicians knew that the patients were on a placebo, they would likely discount any side effects reported because they “know” it could not be from the drug. This would bias the assessment of side effects.

For example, when treating hypertension (high blood pressure) with diuretics, one side effect in males is often impotence (or to use the politically correct term, “erectile dysfunction”). If a placebo versus active-drug trial were to be conducted, a surprising amount of impotence would occur in the placebo group just because of its increasing occurrence with age for males. If clinicians are not blinded, they may discount all cases of impotence in the placebo group as not being due to the drug (since they are not receiving the active treatment). They would, however, count all cases of impotence where the subject is on active treatment. As a result, the amount of impotence associated with the drug would be grossly overestimated. One needs to collect data from both the placebo and treated groups in a blinded manner, and then compare the differences in order to insure an accurate measurement of how much change is due to the active drug.

Masking is not always possible. For example, a trial comparing surgery to medicine would clearly not be able to mask the patient from their scar. Other times, a classic response to treatment un masks the treatment assignment, such as a large change in heart beats on certain

drugs or the absence of hot flashes in women given hormone-replacement therapy. In trials where the subjects must actively participate in the treatment, such as a low-fat diet or exercise trial, masking is impossible. In trials such as these, the common approach is to mask the person making evaluations and/or use an independent observer (who does not know which treatment has been used) to assess the outcome. In such situations, we often try to use outcomes that are totally objective, such as pregnancy (for example) in a treatment trial aimed at increasing fertility.

Another reason for blinding or masking is tied into the concept of equipoise (uncertainty of treatment effect). For a clinician to remain in equipoise, he or she must not know the results of the trial before the formal end of the trial. If he or she is analyzing the results of the trial and knows the outcome results in each group as the trial moves forward, it is unlikely that he or she could remain in equipoise. When trials are conducted at several sites, such unblinded outcome results – should others learn of these results – could alter the behavior of the clinicians evaluating the patients and analyzing the data at these different study locations. Thus, in multi-center trials, the clinical investigators at each site should not know the treatment each patient is receiving, and they should not have any idea of the expected overall results.

### ***The Data and Safety Monitoring Committee***

These results are viewed over time by the coordinating center and a special group advisory to the trial called a Data and Safety Monitoring Committee (DSMC). The DSMC

members are not directly involved in the trial, but instead are responsible for monitoring the safety and efficacy (effectiveness) of the treatment throughout the duration of the study. This committee sees unblinded data with the charge to recommend stopping a trial if clear evidence of benefit or harm is discovered before the trial is scheduled to end. This is not an easy task, because patients do not enter trials on the same day, and thus, the DSMC is always working with partial information.

For example, suppose 500 patients are needed for a trial (250 each per control and active-treatment groups). If the study recruits five patients per month in each of eight centers (40 total patients per month), more than a year (500 divided by 40 equals 12.5 months) would be needed to complete recruitment. Most DSMCs meet at least every six months. At their first meeting, they would be looking at data from the first six months or 240 patients, but



each set of 40 patients would have been followed for one less month than the others who started before them. Making decisions on partial information is biasing for the treating doctors as we noted, but it is equally dangerous for DSMCs. They must use predefined rules for stopping a study for efficacy (effectiveness) and wisdom for stopping a study for safety concerns.

These are difficult decisions. To illustrate how random outcomes can be misleading, the following is a sequence of heads (H) or tails (T) from tossing a coin twenty times:

Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ  
Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ

As would be expected, there are 10 Hs and 10 Ts, a 50/50 chance of tossing “heads.” If H and T represent the successes in the trial for each group respectively, we have the same number of successes in both groups; if these were the results of a clinical trial, the trial would end with no difference. However, consider the job of the DSMC in monitoring. If the first six cases were all the data that were available at the first meeting, the DSMC would review the following results: H H H T H H. They would see five successes in one group and only one success in the other. This would appear to be success for one treatment and might lead the DSMC to consider stopping the trial for benefit in the H group. They would be wrong, but if they stopped the trial early – we would not know the true answer.

There are very carefully crafted statistical rules to prevent the erroneous early termination of a study. In the news media, we often see

researchers accused of either waiting too long to stop a trial, or stopping one too soon. In a recent trial reported in the *New England Journal of Medicine*<sup>1</sup>, a gas was given to premature infants to prevent lung injury when on a ventilator. The trial was stopped for a potential adverse effect (bleeding into the brain). However, when all the data (that were collected at the time of stopping) were sent in to the coordinating center, the results were no longer significantly indicative of this adverse outcome. The threshold for making a decision to stop a trial for safety is clearly lower than the threshold for declaring a treatment successful. This is because of the clear obligation, both scientifically and ethically, to protect patient safety.

Most trials are planned to follow patients for a fixed amount of time. The duration of the trial is usually a year or two for multiple sclerosis trials, but some, including the current Combination Therapy Trial COMBIRx, are scheduled to go three years. Information from these studies accumulates gradually, and most trials are able to continue to their planned termination time. The fact that most trials are not stopped early is in some ways a testament to the planning and prior information used in trials.

Despite the headlines that often “shock” the media when a bad outcome occurs, most trials are safe. Not all trials, however, end with positive results. In the media, this is often lamented or touted as failure, but in the search for effective treatments for diseases – including MS –

---

<sup>1</sup> Van Meurs KP, Wright LL, Ehrenkranz RA, et al. “Inhaled nitric oxide for premature infants with severe respiratory failure.” *N Engl J Med* 2005; 353:13-22.

even trials that end with negative results are actually successes. They are successes because the expectation that a treatment is going to be good is not the same as proving it is good. The history of the FDA requires that we demonstrate effectiveness, and when a trial fails, most often it means that the treatment has not met the standards necessary to show success.

### ***Showing Success and the Concept of Causation***

How do we show success? We want to demonstrate that the improvement is due to the drug or procedure being studied and not just natural history (i.e., the natural course of a disease, such as when one's MS symptoms remit), time, or other forces. An important factor that is key to the needs of conducting a trial is the concept of causation. That is, the actual treatment is the real reason for any change in the patient.

There are a number of principles of causation. **Cause** is not the same as an association. An association may be found between two characteristics for several reasons. There may be **direct causation**, e.g. smoking causes lung cancer. In contrast, there may be a **common cause**, e.g. ice cream sales and drowning incidents both increase with temperature, but they are not causally related. Sometimes there may be a **confounding factor**, e.g. highway fatalities decreased when the speed limits were reduced to 55 mph, but at the same time, the oil crisis caused supplies to be reduced and people drove fewer miles. Or there may be a **coincidence**, e.g., the population of Canada has increased at the same time as the moon has gotten closer to the earth by a few miles.



***An important factor that is key to the needs of conducting a trial is the concept of causation. That is, the actual treatment is the real reason for any change in the patient.***

When the FDA mandates the proof of effectiveness, it is essentially asking that cause and effect be established. The other associations aside from direct causation (common cause, confounding factor, and coincidence) must not be excluded from the trial results.

How do we establish a cause-and-effect relationship? The following must be considered

before causation can be declared. Seven general categories are used to assess the likelihood of a causative relationship:

- 1) Strength of the association:** The stronger an observed association appears over a number of different studies, the less likely this association is lacking validity because of bias.
- 2) Dose-response effect:** The value of the treatment response changes in a meaningful way with the dose (or level) of the suspected causal agent.
- 3) Lack of temporal ambiguity:** The potential cause precedes the occurrence of the effect. In other words, the improvements in the clinical condition should follow after the initiation of therapy.
- 4) Consistency of the findings:** Another extremely important component for the evaluation of trial results is consistency. Most or all studies concerned with a given

causal hypothesis need to produce similar findings. So when similar patients are treated in other studies, the results should be similar. Identical results are not expected, but the general effects should be the same.

- 5) Biological or theoretical plausibility:** The potential causal relationship is consistent with current biological or theoretical knowledge. Please note, however, that the current state of knowledge may be insufficient to explain certain findings. For example, if a drug is given in a trial to reduce blood pressure, but the treated group experiences reduced fatigue instead, one must consider that the fatigue was influenced by the drug (and then consider what mechanism may have caused this effect).
- 6) Coherence of the evidence:** The findings do not seriously conflict with accepted facts about the outcome variable being studied. In other words, based on a comprehensive understanding of the disease process, what we observe seems to explain the changes we have seen in the trial.
- 7) Specificity of the association:** The observed effect is associated with only the suspected cause (or a few other causes that can be ruled out). This criteria was originally designed by looking at infectious diseases, such as malaria, which is caused by an organism that gets into your blood stream. However, there are many situations where an outcome results from many causes (i.e. heart disease may be caused by smoking, poor diet, lack of exercise, family's predisposition, etc.) and multiple outcomes result from a common cause (i.e. obesity may cause heart disease, diabetes, orthopedic issues, depres-



sion, etc.). Because of these numerous exceptions, “specificity of association” might be considered a weak and potentially unnecessary criteria for causation.

In clinical trials, many of these criteria are met at initiation. Specific background and rationale for trials must be made prior to the initiation of the trial and these criteria are often used to justify the treatments and the expected results of the trials. Endpoints and outcomes are specified in advance (before the trial is started) and the primary hypothesis of interest is stated.

We have previously noted that the study is analyzed when all data are complete and finalized. Following the primary-outcome analysis, we often have additional analyzing called “post-hoc” analyses, to supplement the primary analysis and help further establish that the treatment has indeed caused the outcomes seen. Despite this rigorous analysis, we often need to validate or confirm these findings in another study.

### **PUBLISHING THE RESULTS OF A CLINICAL TRIAL**

Truthfulness in the reporting of trials is critical and expected. A great deal of pressure exists to have public statements of clinical trial endpoints before trials start. This insures the public and the scientific communities that the findings provided are indeed what were expected. In these statements, researchers prospectively define hypotheses, clinical objectives, and planned analysis.

One such source of this information is [www.clinicaltrials.gov](http://www.clinicaltrials.gov), which is a website that lists most of the ongoing trials. All NIH (National Institutes of Health) trials must be listed on this site before patients can be entered

into the trial. This prevents investigators from “data mining,” a term used to describe the act of looking to find extra value of a treatment — value that was not expected or planned before the trial. Data mining results may be questionable in their validity. Mandating that all NIH trials be listed on this site also prevents investigators from obscuring the details of unexpected additional results when given in a publication or presentation.

How can a non-scientist evaluate research? Research publications have a format that is commonly followed:

- Methods (how things were done in a study)
- Protocol (what was to be done)
- Statistics (how were the data analyzed)
- Assignment (whether or not the study was randomized)
- Blinding (if the study was single, double, or triple-blinded)
- Results (what was found)

The paper should also include discussions on how participants were recruited and followed; how one interprets all of these results; and what limitations, sources of bias, and “external validity” (additional facts outside the study) may exist.

Outcomes from data mining are often combined with the pre-planned outcomes, which creates problems with interpretation of the data. Why does this matter? The answer is partly statistical. Just like the heads and tails coin-toss example noted earlier, some results which show differences between groups will occur by chance.

In the design of a trial, we plan for this when defining the primary outcome. We

attempt to insure how often this can occur by our choice of the number of participants. However, we do not plan for all the potential outcome variables that may exist. Thus, when someone is data mining, taking into consideration the many possible outcomes, we do not know if this result is just a chance occurrence or a real difference. With the help of websites such as **www.clinicaltrials.gov**, we are able to learn if the finding presented was specified in advance or if it is just a potentially random finding. If it is indeed just a finding, we should be more cautious in our interpretation of the meaning of such “data dredging.”

What is in a paper about a clinical trial? Generally, the paper’s title tells us about the primary hypotheses or the major question under investigation. The result tables provide the key endpoints and data items.

When reading about trials, strong preference is given to studies that are:

- Prospective (planned in advance)
- Randomized (“flip of a coin” assignment to a treatment group)
- Controlled (meaning that the studies are carefully implemented with standardization amongst all sites and personnel; also, the groups are comparable)
- Analyzed (an analysis has been done of the patients in the groups to which they were randomized)

Some general criticisms of trials are that:

- They do not provide enough new information
- They fail to state their initial reason or hypothesis for the study
- They do not adequately describe what was done
- The trial was conducted on too few patients (This latter concern is a major problem for studies that declare two treatments to be the same or a study that reports no risk associated with the treatments. It may simply be, amongst a small sample of patients, one was just not able to see any rare events.)

### **IN SUMMARY**

Many technical components are involved with clinical trials. Everyone is searching for better treatments and wants to see positive results. When reading about successes, one needs to understand that this is just one study. Scientists are often skeptical to accept the results of a single study, no matter how large or how expensive. We expect that a synthesis of results can lead us to the treatments that work.

---

*“The RCT [Randomized Controlled Trial] is a very beautiful technique, of wide applicability, but as with everything else there are snags. When humans have to make observations there is always the possibility of bias.”*

— Archie Cochrane (1972)

The mission of the Multiple Sclerosis Association of America (MSAA) is to enrich the quality of life for everyone affected by multiple sclerosis. MSAA accomplishes its mission by offering many vital programs and services to members of the MS community.

MSAA's free programs and services include: toll-free telephone Helpline with trained consultants (English and Spanish); MSAA publications; Equipment Distribution Program; Cooling Equipment Distribution Program; MRI Institute and MRI Diagnostic Fund; Barrier-Free Housing Program; regional events and activities; Networking Program; Lending Library; and other programs. Please call the Helpline at (800) 532-7667 or visit MSAA's website at [www.msassociation.org](http://www.msassociation.org) for information and assistance.

Help or support to MSAA in any way is truly appreciated. To inquire about volunteering, fundraising, or making donations, please contact MSAA at (800) 532-7667 or visit MSAA's website at [www.msassociation.org](http://www.msassociation.org) for information and assistance.



**National Headquarters**  
**706 Haddonfield Road, Cherry Hill, New Jersey 08002**  
**Phone: (800) 532-7667 • Fax: (856) 661-9797**  
**Email: [msaa@msassociation.org](mailto:msaa@msassociation.org) • Website: [www.msassociation.org](http://www.msassociation.org)**