387. Imbalances or inequalities on study variables at the outset of a study can be an accidental result of the procedure by which subjects are assigned to treatments (Moertel, Tr. 5544; Sunshine, Tr. 9662; Laska, Tr. 10260–64). Use of randomization in that assignment procedure is supposed to guard against such baseline imbalances or inequalities and the attendant problems in interpreting results (Brown, Tr. 5083–85; Forrest, Tr. 8916; Beaver, Tr. 6022–23). In certain cases, statistical techniques may be available to readjust or "correct" for such baseline inequalities and to render results interpretable (Moertel, Tr. 5544; Laska, Tr. 10269; Brown, Tr. 5086–87; Forrest, Tr. 9121). However, the magnitude of the observed imbalance, and the importance of the variable on which the imbalance occurs, are crucial factors in determining whether **[101]** such statistical correction of baseline imbalances restores the study's validity (Brown, Tr. 4911–12, 8052–54, 8146; Forrest, Tr. 9121).

388. An inflexible prerequisite of any well-controlled clinical study, and particularly in the area of mild analgesic drugs and pain relief, is double-blinding. That is, neither the test subject nor the investigator should be able to tell which treatment is being administered (Azarnoff, Tr. 9180; Evans, Tr. 6354, 6357; Moertel, Tr. 5538; Grossman, Tr. 7767–68; Forrest, Tr. 8912; Sunshine, Tr. 9676–77; Laska, Tr. 10166; CX 514, p. 35444). Responses to analgesic drugs can be significantly affected by subjects' pre-existing biases or beliefs and expectations (Beaver, Tr. 6016; Moertel, Tr. 5538; Forrest, Tr. 9052; Evans, Tr. 6357–62; Brock, Tr. 8556–61). The whole point of the double-blind technique is to separate out the effect of expectation from the true pharmacologic effect of the drugs tested (Beaver, Tr. 6014). Moreover, the conscious or unconscious biases of the investigator, nurse observers, the subjects and others involved in the conduct of the study can exert an effect that distorts the action of the actual treatments administered (Evans, Tr. 6341, 6357–62; Moertel, Tr. 5538). Double-blinding effectively controls the expectations and beliefs of subjects and the biases and influences of those conducting the study by assuring that these extraneous effects cannot differentially impact on any particular treatment (Beaver, Tr. 6014–16; Evans, Tr. 6360). Strictly speaking, patient expectations and investigator biases can not be entirely eliminated, but double-blinding at least assures that all treatments in the study will be equally affected (Azarnoff, Tr. 9180; Beaver, Tr. 6015; Forrest, Tr. 8916; Evans, Tr. 6360). To achieve an adequately double-blinded study, it is essential that the treatments look the same, taste the same and appear the same in all respects, so that the subjects in one treatment group will not be prompted to expect something different from subjects in another and investigators will have no clue

as to which treatment they are administering (Azarnoff, Tr. 9180; Beaver, Tr. 6023–24).

389. Whenever possible, a well-controlled study comparing the efficacy of two drugs, particularly mild analgesics, should include a placebo control (Forrest, Tr. 8922; Moertel, Tr. 5539–41; Azarnoff, Tr. 9181; Beaver, Tr. 5979–81; CX 514, pp. 35444–45). The placebo, a pharmacologically inert substance, acts as a separate treatment in the study, and it serves as a built-in measure of the sensitivity of the study and as an analytical tool to aid in the analysis of its results (Forrest, Tr. 8923, 9008–09; Moertel, Tr. 5539–41; Azarnoff, Tr. 9181). Unless the results of a study demonstrate its ability to distinguish a standard analgesic compound—such as aspirin—from placebo, one cannot be certain that the study was sufficiently sensitive to detect differences between the standard and test compounds under study, even if such differences in fact existed (Forrest, Tr. [102] 8923; Moertel, Tr. 5539–41; Beaver, Tr. 5979–80; Lanman, Tr. 12092–93). Similarly, in the absence of a placebo control, the failure to find a difference between the treatments under study may be due to insensitivity in the study methodology rather than to the fact that no difference exists between the treatments (Beaver, Tr. 5979–81; Forrest, Tr. 9008).

390. The statistical techniques for analyzing the results of clinical trials should be set out in advance and should be appropriate to the design and purpose of the study (Azarnoff, Tr. 9180, 9183; Moertel, Tr. 5542). Deciding upon the statistical analysis in advance guards against the investigator peeking at the data and, perhaps, aborting a study before completion when a desired result has been reached or choosing to analyze only those segments of the study that may show favorable results (Moertel, Tr. 5542–43). Failure to adhere to statistical procedures set forth in advance introduces a bias into the analysis (Azarnoff, Tr. 9183). Such "data massaging" destroys the validity of the analysis (Moertel, Tr. 5543).

391. When studies are designed for the purpose of establishing differences between the treatments under study, there must be a method to judge whether any observed differences may be due to chance or simple random variations in the data generated rather than to actual differences in the effects of the treatments (Brown, Tr. 4867–69; Moertel, Tr. 5545). When the observed differences are shown through appropriate statistical analyses to be significant at or beyond the 95% level, scientists will accept those differences as real and not being due to chance (Azarnoff, Tr. 9182; Brown, Tr. 5143; Forrest, Tr. 8912; Moertel, Tr. 5545–46). Scientists are not willing to accept greater than a 5%, or one in 20, likelihood that the differences observed in a study are due to chance (Azarnoff, Tr. 9182; Brown, Tr. 5143; Moertel, Tr. 5545). This maximum 5% chance likelihood as a standard

for statistical significance is generally accepted in the scientific community, including the scientific literature (Brown, Tr. 5138–40, 5142–43; Moertel, Tr. 5545; Forrest, Tr. 8912; Azarnoff, Tr. 9182; Laska, Tr. 10551–53).

392. When an observed difference between two drugs is determined to be statistically significant at or beyond the 95% level, clinicians who evaluate the results of studies on analgesics also address the separate question of whether such statistically significant differences have clinical importance (Beaver, Tr. 5971–72; Moertel, Tr. 5609–13; Forrest, Tr. 8912, 8915; Azarnoff, Tr. 9182–84). As Dr. Beaver stated:

> . . . the difference, to be a difference, must make a difference. What we would normally do is say if the difference is small beyond a certain point, it may, in fact, exist but it doesn't make any [103] difference. It does not serve as a reasonable basis for choosing one product over another [or] making a particular claim about a product. (Beaver, Tr. 5971).

393. Selection of any specific, objective standard of the clinical importance—as opposed to the statistical significance—of differences between drugs is exceedingly difficult (Laska, Tr. 10459). It is clear that unless a difference is statistically significant at or beyond the 95% level, it cannot be clinically important (Moertel, Tr. 5611; Forrest, Tr. 8912; Azarnoff, Tr. 9183–84). On the other hand, by using a large number of patients, it is possible to demonstrate the statistical significance, at the 95% level, of minute differences (Moertel, Tr. 5610). Therefore, a meaningful way to resolve concerns over the magnitude of difference necessary for clinical importance is to require statistically significant differences to be obtained with a reasonable sample size, and no greater (Forrest, Tr. 8914). Generally, past studies comparing the efficacy of analgesics, which have provided results that clinicians have acted upon as clinically important, have had sample sizes in the area of 20–50 subjects per treatment (Forrest, Tr. 8913; Sunshine, Tr. 9772–75). With allowances provided for the additional levels of within-study variation that are inherent in studies of mild OTC analgesics, Dr. Forrest concluded that if a well-controlled study could demonstrate statistically significant differences (at the 95% level) between mild analgesic treatments with no more than 50 to 60 subjects per treatment, he would accept those results as clinically important (Forrest, Tr. 8914–15). If more subjects are required to demonstrate the statistical significance of observed differences, their clinical importance diminishes (Forrest, Tr. 8915).

394. Subjecting a clinical study to peer review, which occurs when a study is submitted for publication in a reputable journal, adds another indication of reliability and allows greater confidence in a study (Moertel, Tr. 5545; Forrest, Tr. 8921). One of the important criteria

used in coming to a conclusion about the validity of a study is whether it is published and whether, thereafter, it meets with the acceptance of other scientists and, ultimately, whether the study is replicated by others (Brown, Tr. 4915).

395. The standards for well-controlled clinical trials necessary to establish a claim of absolute or comparative efficacy between drugs are and have been well accepted in the scientific community by experts in the design and analysis of such studies for years (Moertel, Tr. 5545; Forrest, Tr. 8923; Azarnoff, Tr. 9178). The FDA Panel on OTC Analgesics has incorporated these principles and requirements for well-controlled clinical studies into its Final Report (CX 514, pp. 35371, 35444–45), and FDA has codified many of these principles **[104]** into its regulations mandating the need for "substantial evidence" to support effectiveness claims for drugs (21 C.F.R. 314.111(a)(5)(ii)(a) through (c)).

2. Evidence Other Than Well-Controlled Clinical Studies Is Insufficient to Establish Superior Efficacy of One OTC Oral Analgesic Product Over Another

396. Various attempts to measure the absolute or comparative efficacy of analgesics other than by well-controlled clinical trials using appropriate pain models have not been shown sufficiently reliable to establish absolute or comparative efficacy of analgesic agents in man and are not accepted either by experts in the evaluation of analgesic agents or by the FDA (F. 397–404, *infra*).

397. Consumers' perceptions of therapeutic superiority of one product over another product are not reliable evidence for the purpose of establishing the efficacy or comparative efficacy of OTC analgesics because consumers are unable to evaluate for themselves the true pharmacologic efficacy of drugs (Moertel, Tr. 5631, 5749–59; Evans, Tr. 6354–60; Azarnoff, Tr. 9196; Grossman, Tr. 7887–89). Of course, consumers do perceive that they feel better, or that they hurt less after swallowing a pill (Grossman, Tr. 7787–89; Evans, Tr. 6354–55, 6357). The inability to "evaluate" in this context simply refers to consumers' inability to distinguish the true pharmacologic contribution of a drug from a host of factors that have nothing to do with the drug's true pharmacologic effect (Moertel, Tr. 5749–55; Beaver, Tr. 6020; Forrest, Tr. 9052; Evans, Tr. 6355; Azarnoff, Tr. 9196; Grossman, Tr. 7887–89).

398. A consumer's expectations of what a drug will do are an important factor and play a powerful role in influencing his response to the drug (Brock, Tr. 8556–61; Beaver, Tr. 6014, 6016; Evans, Tr. 6355–56). However, such responses do not reflect the true pharmacologic action of the drug and should not be relied on for the purpose of determining

whether a drug is effective or whether one drug is more effective than another. The simple reason is that a consumer's expectations are affected by many extraneous factors, such as his or her disposition, advertising, past experience with the drug, relationship with the physician or nurse administering the pill, and even the size, shape and taste of the pill taken (Evans, Tr. 6355; Moertel, Tr. 5751–52). In fact, in cases where the effect of a drug is somewhat indeterminate or where the consumer has no yardstick or information about its effect, he may well be dependent upon extraneous information or suggestion for making up his mind about what the effect of the drug is (Brock, Tr. 8556–61). **[105]**

399. Thus, consumers on an unblinded basis cannot differentiate between a true pharmacologic response of a drug and a response due to extraneous factors, such as suggestions or expectations, that surround the taking of the drug. The influence of expectations or suggestions are so real that even blinded subjects in a controlled test report pain relief from a placebo (Forrest, Tr. 9050, 9052; Evans, Tr. 6326–30). This phenomenon is known as the "placebo effect" among medical-scientific investigators. The placebo effect is typically reported in the scientific literature to produce subjective pain relief in over 30% of test subjects in controlled analgesic studies (Evans, Tr. 6324, 6328–29; Laska, Tr. 10492). Anyone on any occasion can be a "placebo responder" (Laska, Tr. 10493–94). Expectations and similar factors, and hence the "placebo effect," can never be totally eliminated from any situation where a human suffers pain, but well-controlled testing methodologies can control expectations and other nonspecific factors, and therefore the placebo effect, by ensuring that the treatments under study are equally affected by them (Beaver, Tr. 6015, 6019; Evans, Tr. 6340–43; F. 384, *supra*). Balancing nonspecific factors across the treatments in a study, through techniques of randomization, blinding and the other controls already discussed (F. 384–87, *supra*) is the only accepted way that human tests can be expected to provide reliable information about the true efficacy and comparative efficacy of drugs (Beaver, Tr. 6014–25; Evans, Tr. 6340–48, 6354–63).

400. The fact that an OTC analgesic contains a combination of ingredients, or more ingredients than another OTC analgesic, is not acceptable evidence that it is more effective (Azarnoff, Tr. 9188; Forrest, Tr. 8977–78). In order to conclude that one analgesic—even with more ingredients—is more effective than another, one needs adequate, well-controlled clinical studies (Forrest, Tr. 8977–78).

401. For many drugs, the relationship between the blood levels and the drug's effect has been determined. However, in the case of aspirin or aspirin products, no direct correlation has yet been scientifically established between the amount of aspirin appearing in the blood-

stream at any time point and the degree of onset, intensity or duration of pain relief afforded by aspirin. Therefore, "blood level" studies, *i.e.,* studies that simply examine the amount of a drug in the bloodstream at various time intervals following ingestion, are not a reliable basis for predicting comparative analgesic performance beyond that the general level of aspirin in blood (serum salicylate concentration, or blood level) associated with pain relief is known. The unique characteristics of aspirin in this regard has been attested to by qualified expert witnesses who testified in this proceeding (Azarnoff, Tr. 9189–90; Beaver, Tr. 5945–46; Forrest, Tr. 8987–90; Moertel, Tr. 5801–05, 5817–18, 5860). This view is shared by the FDA Panel on OTC Analgesics (CX 514, [**106**] pp. 35359, 35361, 35374, 35377–78), by a panel of well-respected experts convened by the National Academy of Sciences/National Research Council to evaluate various claims for analgesics (CX 511F; F. 22–26, *supra*), by the *AMA Drug Evaluations* prepared by a panel of experts to evaluate evidence bearing on the performance and comparative performance of drugs (CX 512H, CX 518G; F. 216–23, *supra*); and by the *Medical Letter,* a recognized publication relied upon by physicians and other scientists for information relating to the performance of medicines (CX 510A, B; F. 225–28, *supra*).

402. Thus, clinical studies which simply show that one analgesic preparation is absorbed more rapidly into the bloodstream than another cannot lead to conclusions with respect to the comparative speed of the analgesics in relieving pain.

403. Studies employing experimental pain, *i.e.,* pain induced in humans in the laboratory by various artificial devices, are not sufficiently reliable for use in establishing the comparative efficacy of OTC analgesics. Experimental pain studies have failed to predict with any consistency the clinical performance of analgesic drugs, particularly those used for OTC medication (CX 514, p. 35444; Evans, Tr. 6353; Elvers, Tr. 11087–88). Pain induced in the laboratory by various artificial means is significantly different from pathological pain or pain in natural state, and for this reason the performance of analgesic drugs in relieving pathological pain must be determined in the clinical setting (Evans, Tr. 6353; CX 425C; F. 544, *infra*).

404. While more advanced forms of experimentally induced pain, such as submaximum tourniquet pain (where the subject's arm is cuffed, and the arm worked until pain is induced), come somewhat closer to imitating pathological pain (Evans, Tr. 6338–39), even these have been found by experienced investigators to be insufficiently reliable predictors of analgesic performance (Evans, Tr. 6375; Elvers, Tr. 12352). The problem of simulating clinical pain in the laboratory is so complex that results obtained with presently employed experimen-

tal pain producers can, in fact, be seriously misleading (Elvers, Tr. 11189–90).

B. *The Design Of In-Patient Clinical Studies To Assess Comparative Analgesic Performance*

405. Studies of analgesic performance in man rely of necessity upon the verbal reports of patients in pain to generate the data which are then analyzed (Forrest, Tr. 8869–70; F. 369, *supra*). Typically, before hospitalized patients are accepted into a clinical analgesic study, they will be interviewed by an observer/investigator to obtain their history, their consent to participate and to ascertain the level of their pain prior to treatment (Brown, Tr. 4976–78, 4981–82, 4985; *see e.g.,* CX 425Z002; Smith, Tr. 5405; CX 454C). This baseline, or initial pain level, is determined by the patient's statement **[107]** that she is in "severe" pain, "moderate" pain, "slight" pain or "none" (Brown, Tr. 4988; CX 425Z002; Smith, Tr. 5404–05; CX 454C). Researchers generally seek patients in "severe" or "moderate" initial pain so that the pain reducing properties of the compounds under study will have fairly good opportunity to perform (Forrest, Tr. 8882–83; Smith, Tr. 5431–32). Indeed, some researchers seek to confine patients to those in "severe" pain to maximize the opportunity for observing any differential performance of the test compounds (Forrest, Tr. 8882–83).

406. Pain relieving performance is typically measured in two ways: (1) reduction in pain intensity; (2) amount of pain relief (Smith, Tr. 5419; Brown, 4880–82). That is, at fixed intervals following the initial interview and the administration of a blinded treatment, patients are asked (1) to describe the amount of their pain as "severe," "moderate," "slight," or "none," and (2) to describe the amount of pain relief they have experienced as "complete," "more than half," "less than half" or "none" (Smith, Tr. 5406–08; CX 454C; Brown, Tr. 4880–82). The difference in pain intensity is quantified by first assigning numerical values to the levels of pain intensity possible. For example, "severe" is frequently given a value of 3; "moderate" a value of 2; "slight" a value of 1; and "none" a value of 0 (Brown, Tr. 4882; Smith, Tr. 5406; CX 454C; CX 425Z007).

407. The pain intensity difference (P.I.D.) between the baseline or pre-treatment pain level and the pain level at the time of the first post-treatment interview is calculated by simply subtracting the pain intensity score at this interview from the initial pain intensity score (Brown, Tr. 4881–82). Thus, if a patient started in pain which she described as "severe" and, after one-half hour (or some other fixed interval) described her pain as "slight," her pain intensity difference (P.I.D.) score would be 2 (*i.e.,* "severe" (a score of 3) minus "slight" (a score of 1) equals 2) (Brown, Tr. 4881–82). The patient's pain relief is

also quantified by assigning an appropriate numerical value to the patient's statements at succeeding interviews, that their pain, for example, has been "completely relieved," "more than half relieved," "less than half relieved," or "no relief" (Smith, Tr. 5406–07; CX 454C).

408. A pain intensity difference (P.I.D.) score can be calculated for each succeeding interval (generally one hour) after treatment by subtracting the patient's pain score for that interval from the baseline, pre-treatment pain score (Brown, Tr. 4881–82; Smith, Tr. 5404–06). A pain relief score can be determined for each interval by assigning the appropriate numerical value to the patient's level of relief reported at each succeeding interval (Brown, Tr. 4881–82; Smith, Tr. 5406–08).

409. If a study is designed to last six hours, and to include six hourly post-treatment interviews, each patient who [**108**] completes the study will have six (6) P.I.D. scores and six (6) pain relief scores (Smith, Tr. 5420–21; Brown, Tr. 4881–82). The standard method of preparing these data for analysis is to add the six P.I.D. scores for each patient, and the six relief scores for each patient, to determine the Sum of Pain Intensity Differences (SPID) for each and the Total Pain Relief Score (TOTAL or TOTPAR), respectively (Brown, Tr. 4882; Smith, Tr. 5420–21). An average score is then calculated for each treatment group on each method of "scoring" analgesic performance, and this is used as a basis for comparing treatments (Beaver, Tr. 5988–89). Obviously, the higher the SPID score, the greater the reduction in pain intensity for a particular treatment. Similarly, the higher the TOTAL score, the greater the pain relief afforded by the treatment.

410. When the investigator wants to determine the question whether a specific dose of a drug (*e.g.,* two tablets of Excedrin) is more effective or faster-acting than a specific dose of a standard (or known) drug (*e.g.,* two tablets of aspirin), it is appropriate to adopt a three treatment study design which compares the performance of each of these two specific dosages and a placebo (Brown, Tr. 8078; Beaver, Tr. 5982, 5987, 6055–56; Forrest, Tr. 8884–85, 8898, 8948–49; Laska, Tr. 10411–12; Moertel, Tr. 5712). Such a "head to head" (or "efficacy") study design enables the investigator to conclude, where a statistically and clinically significant difference is shown, that one treatment was shown to be more effective or faster than the other in that study (Forrest, Tr. 8898, 8948–49; Beaver, Tr. 6055–56; Brown, Tr. 8078; Laska, Tr. 10411–12). In such a study design, one can have confidence in concluding that the observed difference between treatments did not result from chance or insensitivity of the study design if the results show that one treatment was statistically significantly more effective than the other treatment and that the standard treatment was statistically significantly more effective than the placebo (Laska, Tr. 10411–12).

410a. The dose-response curve ("DRC") is a graphic expression of the anticipated relationship between drug dosage and biologic response and is usually based on tests of graded doses. The classic DRC for most active drugs is positive: a larger dose produces greater biologic response until a plateau is reached, beyond which incremental increase in dose does not produce any increase in response (Tr. 4849–92).

411. The DRC for an analgesic compound is plotted as follows: a bioassay relating graded doses of the active agent to degrees of analgesia generate a series of individual data for each dosage tested (data point); by averaging the results of observations at each data point, a mean value is obtained for each data point; the mean results are then plotted on a graph (usually the horizontal axis showing dosage, and the vertical, pain relief); and a "best-fitting" line is mathematically drawn [**109**] connecting the data points by the use of least squared analysis. The line so drawn is a hypothetical fitted line (Tr. 4849–92, 5015–22, 5041–47).

412. DRCs obtained through bioassays typically form the basis of relative potency estimates of test drugs compared with a standard drug. As such, DRC is generally accepted by clinical pharmacologists and clinicians as a useful statistical tool which offers best estimates of the indicated doses of a new (or test) drug to be used in place of a known standard drug (a dose-finding tool) (Tr. 4850, 4860–67).

413. Clinical pharmacologists engaged in bioassays of aspirin-order drugs agree that there appears to be a DRC for aspirin. However, its precise shape and slope, including its plateau level and the dosage point where reverse response, if any, begins, is not known. In any event, it is generally agreed among clinical pharmacologists that aspirin and aspirin-order drugs are mild analgesics and their DRCs are predictably shallow. Since the relationship of increased analgesia to increased dosage is proportional to the log dose, the relatively flat DRC means that a large increase in dosage is required to obtain a relatively small increase in analgesic response (Tr. 4941–46, 4948–53, 8938–43, 9209; CX 514, p. 35364).

414. When experimental drugs are formulated in anticipation of introducing them into the reservoir of medications available to the public, an obvious and critically important piece of information concerning these new drugs is their recommended dosage range (Forrest, Tr. 8871; Laska, Tr. 10405–07; Sunshine, Tr. 9863–65; Forrest, Tr. 8885). The marketer of a new drug must be able to integrate it into the existing stream of treatments in a fashion that allows physicians to know what effects it will produce at various dosage levels (Laska, Tr. 10405–07).

415. "Relative potency" is defined as the dose of a "test" compound

necessary to produce equal biologic effects to a known "standard" compound. Relative potency ratio is a ratio of dosages that produce equal effects (Forrest, Tr. 8885, 8893; Brown, Tr. 4850, 4852–55, 4860–62; Beaver, Tr. 5987; Laska, Tr. 10405–06; CX 803, 804, 805). For example, if the "relative potency" of Compound X relative to aspirin is 2.00, it will take double the amount of aspirin to produce the effect equal to a given amount of Compound X; or, conversely, it will take half the amount of Compound X to produce the effect equal to a given amount of aspirin (Laska, Tr. 10405–06; Brown, Tr. 4850). Thus in general if one knows that the relative potency of Compound X to aspirin is 2.00, one knows that 325 mg. of Compound X will give roughly the same effect as 650 mgs. of aspirin (Laska, Tr. 10405–06; Brown, Tr. 4850). **[110]**

416. The inclusion of a "standard" compound, with widely acknowledged effects at known dosages in the statement is a prerequisite in communicating the relative potency of a new compound, since the very concept is based upon performance relative to that of the standard (Brown, Tr. 4850). Thus, a clinician who knows the analgesic effect produced by such standard treatments as 650 mg. of aspirin will be able to substitute 325 mg. of a new compound with a "relative potency" of 2.00 as against these standard drugs and expect his patients to obtain the same analgesic effect from this new treatment (Laska, Tr. 10405–06; Brown, Tr. 4850–54; Forrest, Tr. 8885). Or, the clinician would be able to substitute 500 mg. of the new compound for 1,000 mg. of aspirin and expect to obtain the same analgesic effect (Laska, Tr. 10416–17).

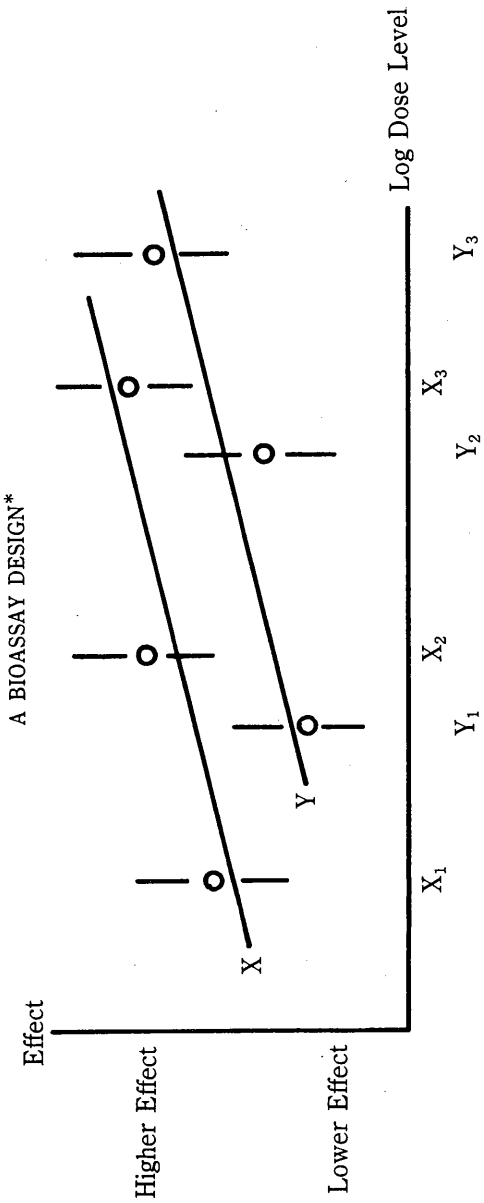
417. Moreover, use of a relative potency permits a clinician to make an assessment of the risk/benefit ratio in using one analgesic as opposed to another. One has to be able to hold effectiveness constant if any comparison of the relative side effect liabilities of the two drugs is to be made. Without such information obtained from a bioassay, one cannot make that judgment (Beaver, Tr. 5998–99).

418. Therefore, the relative potency of two compounds is not the same as their relative efficacy, because the concept of relative *potency* depends upon holding the level of *effectiveness* of the compounds equal (Laska, Tr. 10417; Brown, Tr. 4853–54). Thus, whereas a "head to head" comparison of the effectiveness of a given dose of an analgesic compound to a given dose of another produces a conclusion about the comparative analgesic *efficacy* of the two compounds at the two stated dosages (F. 410, *supra*), "relative potency" produces a conclusion about the relative *dosages* necessary to produce equianalgesia (F. 419–31, *infra*).

419. The determination of the relative potency of a test compound to a standard compound requires a bioassay, a clinical study of more

complex design (using graded doses) than the "head to head" study's single-dose comparison, (Brown, Tr. 4848–49; Forrest, Tr. 8884). A bioassay requires the investigator to compare a *range* of doses of a test compound to a *range* of doses of a standard compound and placebo (Brown, Tr. 4848, 4850, 4852–55; Forrest, Tr. 8884; Laska, Tr. 10417–18). At least two, and frequently three, doses of each compound are generally used, which means that a bioassay may involve five, or seven, or even more treatments (two or three doses of each compound and placebo) (Brown, Tr. 4856, 4872, 8073–76; Beaver, Tr. 5986, 5992–93). [111]

Figure 1.



*Beaver, Tr. 5990-94; Brown, Tr. 4852-55, 4860-62; CX 803; CX 804; CX 805; CX 425Z011

420. In Figure 1, a "best-fit" dose response line for three graded doses of Compound "X" is drawn through the average effect levels for the three successively higher doses of "X" tested (Beaver, Tr. 5988, 5990–94; Brown, Tr. 4860–62). Similarly, "best-fit" dose response line for the three doses of Compound "Y" is drawn through the mean effect levels of the three successively higher doses of "Y" tested (Beaver, Tr. 5988, 5990–94; Brown, Tr. 4860–62).

421. In order to proceed to determine relative potency in this study, several important assumptions about the nature and validity of the bioassay must be satisfied, namely, assumptions of linearity, significant slope, parallelism and equieffective range (Laska, Tr. 10168–73, 10413–16, 10429; CX 900 (graph "a"); Beaver, Tr. 5987–94). First, one must be able to sustain the assumption that each of the "best fit" dose response lines is, in fact, linear. Second, one must be able to sustain the assumption that the two "best fit" dose response lines for "X" and "Y" are in fact parallel. Indeed, lacking linearity and parallelism, a relative potency study has no meaning (Laska, Tr. 10169). Third, one must be able to sustain the assumption that each "best fit" dose response line has a significant slope; *i.e.,* that the level of effect rises, as the dosages increase, to a statistically significant degree (Laska, Tr. 10415). Finally, one must be able to sustain the assumption that the drugs are performing within an equianalgesic range. Each of these assumptions is tested by appropriate statistical procedure and is sustained only if results are significant at or beyond the 5% level of statistical significance (Laska, Tr. 10413–16). In order for a bioassay to be valid, the "best fit" dose **[112]** response lines must be linear, positively sloped, parallel and must describe performance of the drugs in their equieffective range (Laska, Tr. 10413–16).

422. The importance of verifying the validity of the bioassay before estimating the relative potency of the compounds is apparent from the fact that the relative potency is simply the horizontal distance between the two dose response lines (Figure 2) (Beaver, Tr. 5987, 5994; Laska, Tr. 10417; CX 900 (graph "a," "b," "c"); Forrest, Tr. 8893–94; CX 803, 804, 805). The ratio of Dose $_Y$ to Dose $_X$ necessary to produce the *same level of effect* is the relative potency (Forrest, Tr. 8893; Beaver, Tr. 5987; Brown, Tr. 4853; Laska, Tr. 10416–17). Since it represents the horizontal distance between two parallel lines, the relative potency ratio will be the same, regardless of the level of effect chosen, along the entire range of the two dose response lines (Laska, Tr. 10417; Beaver, Tr. 5991).

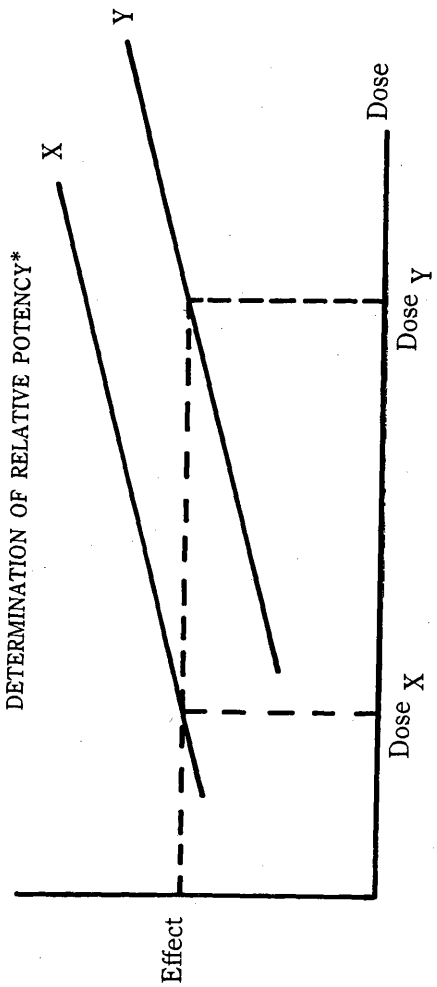


Figure 2.

*CX 803; CX 804; CX 805; CX 900 (graphs "a", "b" and "c")