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).

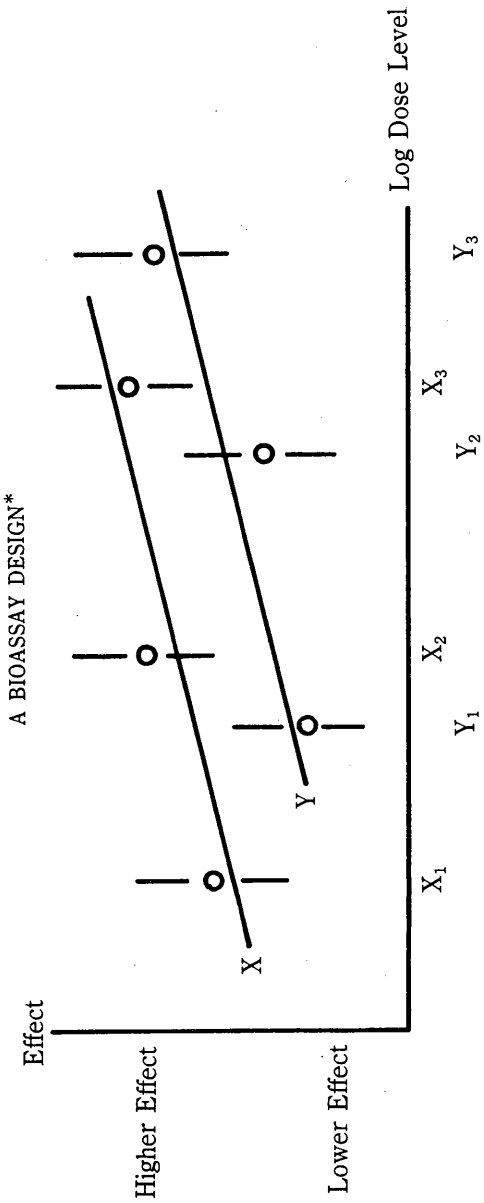
417. Moreover, use of a relative potency permits a clinician to make an assessment of the risk/benefit ratio in using one analgesic as opposed to another. One has to be able to hold effectiveness constant if any comparison of the relative side effect liabilities of the two drugs is to be made. Without 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).

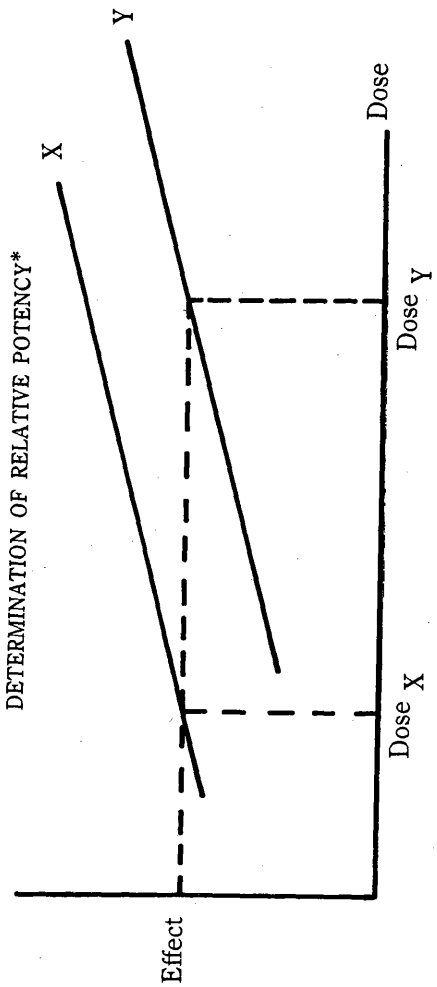


Figure 2.

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

423. In bioassays of analgesics, there is a high degree of variability associated with each average effect level of each dosage of analgesic tested (Beaver, Tr. 5988, 5990; Brown, Tr. 4855). This variation is inherent in the subjective response methodology and particularly where mild analgesics are being investigated (Brown, Tr. 4854–55; Forrest, Tr. 8894; Laska, Tr. 10359–64). This results in part from high patient variability in response to the same dose of a compound and shallow slopes of the obtained dose response curves (Laska, Tr. 10360). The fact that OTC analgesics have shallow-sloped dose response curves means that there will be relatively little increase in effect as the level of dosage increases (Forrest, Tr. 8905–07). Stated another way, in order to produce a small increase in effect, a relatively large increase in log dose is required (*see* CX 514, p. 35364). **[113]**

424. The variation in individual patients' responses is depicted graphically in Figure 1, as the vertical bars crossing each average level of effect (CX 804, 805; Beaver, Tr. 5988). In a bioassay, where it is essential to draw linear dose response curves which "best fit" the data (F. 420, *supra*) and to determine the horizontal distance between them (F. 422, *supra*), it is equally essential that the amount of variation in the data upon which the "best-fit" lines are based be taken into account (Forrest, Tr. 8894; Brown, Tr. 4868). When relative potency is determined, the level of variation in the data is expressed in a confidence interval that permits a reader to know the range in which the relative potency estimate calculated from one bioassay might vary, up or down, upon repeated measurements (Forrest, Tr. 8894; Brown, Tr. 4868–69). Typically, scientists and published articles discussing such biassays do so in terms of a "best estimate of relative potency," with an associated 95% confidence interval, with an upper and lower limit (Brown, Tr. 4868–69; Forrest, Tr. 8894).

425. The qualification of all relative potency ratios as "best estimates" is a scientific necessity reflecting the fact that a bioassay provides only a statistically obtained "best fitting" dose response line for each compound tested (F. 420–23, *supra;* Laska, Tr. 10418–20). The "true" relative potency of one compound relative to another can be obtained only through repeated bioassays, each producing its own "best estimate" with its own level of precision (Brown, Tr. 5146–47). The indicator of each estimate's precision is the "confidence interval" that surrounds it (Brown, Tr. 4868–69). For example, it is possible that a bioassay's "best estimate" of relative potency will be 4.0; but if the 95% confidence interval associated with that "best estimate" is 2.00, on the lower end, to 8.00, on the upper end, it means that on 100 repetitions all that can be said is 95 of those "best estimates" will fall somewhere between 2.0 and 8.0 (Sunshine, Tr. 9687–88; Brown, Tr. 5140–46). Therefore, the wider the confidence interval, the less pre-

cise the relative potency estimate (Brown, Tr. 4869). To take an extreme case, where the confidence interval surrounding a "best estimate" ranges from 0 at the lower end to infinity at the upper end (Brown, Tr. 4869), it would be meaningless and would not permit any conclusions to be drawn about relative potency of the two drugs studied (Brown, Tr. 4869–71).

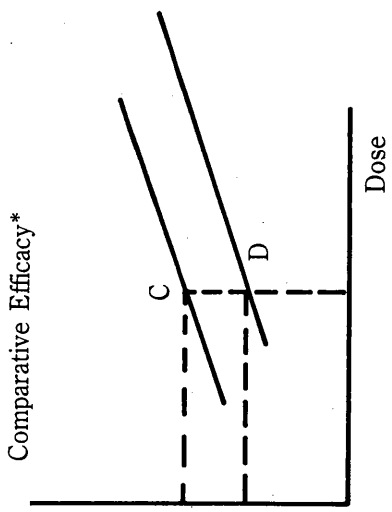
426. To further illustrate, if the relative potency of a new test compound relative to aspirin is estimated to be 4.0, with a 95% confidence interval of 2.0 on the lower end to 8.0 on the upper end, one could state that, according to the best estimate based on the bioassay, about 160 mg. of the test drug may be expected to provide the same effect as a 650 mg. standard dose of aspirin (F. 416, *supra*), and at the same time that, at 95% confidence level, it might take as little as about 80 mg. or as [114] much as about 325 mg. of the test drug to produce the effect equal to 650 mg. dose of aspirin (Forrest, Tr. 8894).

427. A striking illustration of the imprecision inherent in estimated relative potency obtained from a bioassay is provided by a clinical study where the investigators deliberately used morphine as both the standard compound and the test compound. In that case, the investigators were not interested in estimating the relative potency of an unknown test to a standard compound, but wanted to demonstrate the soundness of the bioassay methodology (Brown, Tr. 5005, 5008–09). In that study, the "true" relative potency was, of course, 1 (morphine to morphine). Yet, the bioassay yielded a relative potency of .90, with a 95% confidence interval of .44 on the lower end to 1.8 on the upper end (Brown, Tr. 5008–09).

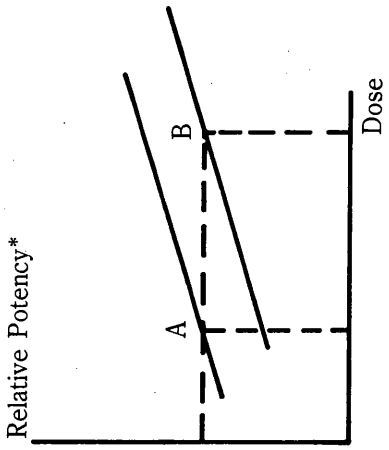
428. The degree of precision of a relative potency estimate obtained from an analgesic bioassay has an important bearing upon the confidence that a scientist can have as he attempts to apply it to clinical situations. The wider the confidence interval surrounding the best estimate, the greater the range of possible equally effective dosages of the test compound relative to the standard. Some clinicians may feel comfortable using the "best estimate" only if the width of its associated confidence interval is no greater than some "reasonable" span, based upon their previous experience (Forrest, Tr. 8913–14). Some may contend that the width of the confidence interval surrounding the "best estimate" that they will accept before they act on it depends upon the purpose for which the drug is to be used or upon the characteristics of the drugs (Laska, Tr. 10206–08). Yet others may take the more liberal view that they will act on the basis of the "best estimate" regardless of the width of its associated confidence interval so long as the interval is not infinite (Sunshine, Tr. 9670, 9689). In any event, it is clear that the relative potency estimate in each of these circumstances provides a convenient and useful device to the clinician

which enables him to make a judgment about the dosage of a new drug that will produce effects about equal to those of a known standard drug.

429. Thus, the function of a relative potency estimate obtained from a bioassay is that of dose-finding. As such, a relative potency estimate is not a statement of the *comparative effectiveness* of the drugs (Forrest, Tr. 8886–8907; Laska, Tr. 10487; Sunshine, Tr. 9693–95). This is not to say that the results of a bioassay cannot be used to arrive at conclusions about the comparative efficacy of the drugs studied (Forrest, Tr. 8885, 8894–8907; Laska, Tr. 10437–38). This point was illustrated by Dr. Forrest and agreed to by Dr. Laska, respondents' expert witness (Forrest, Tr. 8885–8907; CX 834; Laska, Tr. 10487). A graphic depiction of the difference in analysis, one focusing on relative potency and the other on comparative effectiveness, appears in Figure 3. **[115]**



*CX 900 (graph "e"); Laska, Tr. 10437-38



*F. 422, *supra*

Figure 3.

Relative potency, reflecting the distance between A and B, expresses the estimated equianalgesic doses of the two drugs, and is measured on the horizontal axis. On the other hand, comparative efficacy, reflecting the distance between C and D, expresses the difference in analgesic effect produced by an equal dose of the two drugs studied, and is measured on the vertical axis (Forrest, Tr. 8899; Laska, Tr. 10437–38, 10487; CX 900 (graph "e")).

430. When two drugs are equipotent (*i.e.*, where their relative potency is 1.00), their dose response curves lay one atop the other (Laska, Tr. 10426, 10430; CX 900 (graph "d"); Forrest, Tr. 8900). When two parallel dose response curves coincide, the horizontal distance between them is 0, as is the vertical distance (Laska, Tr. 10426; Forrest, Tr. 8900). Thus, when two drugs are equally potent, they are also equally effective (Laska, Tr. 10426–27).

431. Where the issue to be determined is *comparative* efficacy (whether the recommended dose of one drug is more effective than the recommended dose of another), the results of a bioassay need to address the question of whether one can be statistically confident that a difference in their effectiveness exists (Forrest, Tr. 8899–8902; Brown, Tr. 8078). A "head to head" study addresses this question by determining whether the observed difference in effectiveness of the dose of each drug rejects the null hypothesis that there is no difference between the two (F. 410, *supra*).

432. A bioassay can also be used to test a null hypothesis of no difference in effectiveness between the treatments (Laska, Tr. 10426–27, 10519–25; Forrest, Tr. 8899–8902; Brown, Tr. 8078). Graphically, such a test is designed to determine **[116]** whether one can be statistically confident that the two dose response lines do not coincide (Forrest, Tr. 8899–8902; CX 834). Statistically, such a test asks whether one can be statistically confident that the estimated relative potency is above 1.00 (Forrest, Tr. 8899–8901; Brown, Tr. 4934–35, 4939, 5137–38; Sunshine, Tr. 9688–90; Laska, Tr. 10519–25). Unless one can be confident that the dose response curves do not coincide, one cannot reject the possibility that there is no difference in efficacy between the two (Forrest, Tr. 8899–8902; Laska, Tr. 10425–27). Such a test consists of inspecting the 95% confidence interval that surrounds the estimated relative potency. If that confidence interval embraces 1.00, then one cannot reject the possibility (at the 5% level of confidence) that the drugs tested are equally potent and equally effective (Forrest, Tr. 8899–8901; Brown, Tr. 4934–35, 4939, 5137–38; Sunshine, Tr. 9688–90; Laska, Tr. 10426–27, 10519–25). Examining the 95% confidence interval around the "best estimate" of relative potency to see if it includes 1.00 is analogous to testing whether there is a statistically significant difference in the efficacy of the compounds at the 5% level (Laska, Tr.

11358). Unless the 95% confidence interval *excludes* 1.00, it cannot be said that there is a statistically significant difference in their effectiveness at the 5% level (Forrest, Tr. 8899–8902).

433. As Dr. Forrest testified, and as Drs. Laska and Sunshine, both respondents' experts, agreed, knowledge of the estimated relative potency of two compounds does not impart information about the magnitude of difference in their comparative efficacy. That their relative potency is 2.00 does not mean one is twice as effective as the other (Forrest, Tr. 8886–8907; CX 834; Laska, Tr. 10487; Sunshine, Tr. 9690–95). In fact, Dr. Forrest demonstrated that where the parallel "best fit" dose response curves of the drugs are shallow, for any given difference in relative potency one would find little difference in the efficacy of the two compounds, but, when the curves are steep, given the same relative potency one will find a substantial difference in effectiveness (Forrest, Tr. 8905–07; CX 834). Thus, as Figure 4 (CX 834) shows, for a given relative potency (horizontal distance between two dose response lines) one can have either very little, or a large, difference in efficacy (vertical distance between the lines), depending on the steepness of the slope. The parameter that governs the relationship between relative potency and comparative efficacy is the slope of the dose response lines (Forrest, Tr. 8905–07; Laska, Tr. 10487), and in studies of mild analgesics the slopes are shallow or relatively flat (Laska, Tr. 10360, 10414, 10464; CX 514, p. 35364). [117]

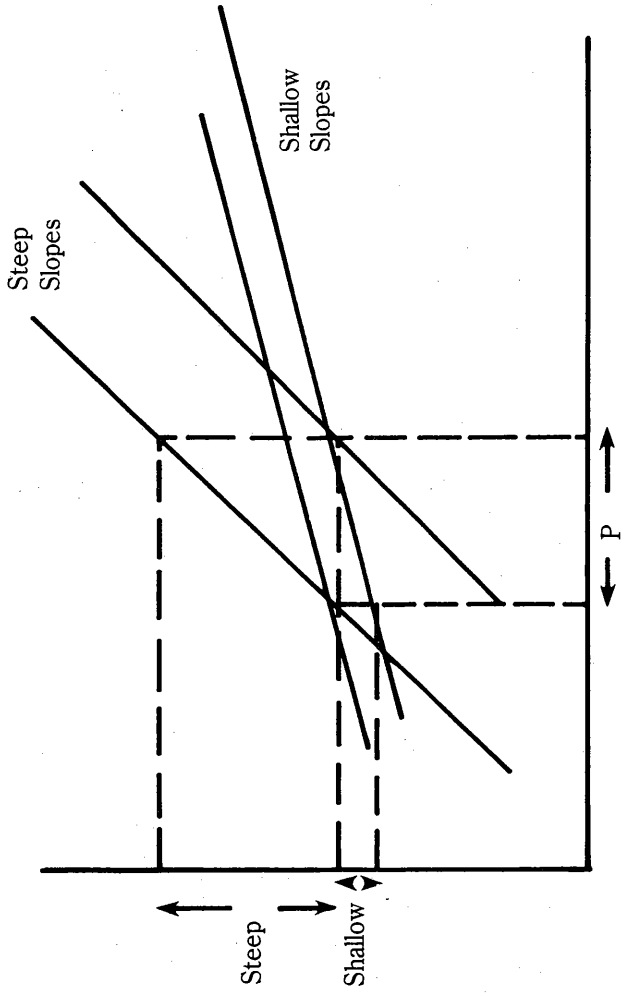


Figure 4.
(CX 834)

434. An estimated relative potency ratio obtained from a bioassay therefore does not by itself provide sufficient information about the precision of that estimate to enable a person to conclude that the drugs studied are, or are not, equally effective, nor does it provide information concerning the magnitude of difference in their effectiveness. Dr. Finney, in his seminal treatise on bioassay, lists as a prerequisite to accurate reporting, the requirement to supply data on the precision of the "best estimate" reported (Laska, Tr. 10506–08). Information on the precision of the estimate supplied by the 95% confidence interval is important to clinicians because, without that information, clinicians cannot make an informed judgment as to what dosage levels of new drugs may be prescribed to obtain effects equal to those of known drugs. As Dr. Laska testified, this dosage-setting application is by far the most prevalent use of bioassays (Laska, Tr. 10405–07, 10428). However, when the results of a bioassay are to be adapted for use in making the wholly separate determination—is there a statistically significant difference in the effectiveness of two analgesics?—then the information supplied by the 95% confidence interval is essential (F. 432, *supra; see* Laska, Tr. 11347–48). If the 95% confidence interval overlaps 1.00, then the study does not reject, at the 5% level, the proposition that the analgesics are equally effective. If, and only if, the 95% confidence interval excludes 1.00 can one conclude that there is a statistically significant difference in the effectiveness of the recommended dosages of the analgesics studied (Forrest, Tr. 8899–8901; Laska, Tr. 10426–27, 10519–25; Brown, Tr. 4934–35, 4939, 5137–38; Sunshine, Tr. 9688–90). [118]

435. Dr. Louis Lasagna published an article entitled "Effect of Naloxone on the Analgesic Activity of Methodone in a 1:10 Oral Combination" in *Clinical Pharmacology and Therapeutics,* Vol. 15, No. 6, 1974 (Tr. 9721). In this article, Dr. Lasagna used the results of a bioassay study design to test the hypothesis that two compounds were equally effective. In the article, Dr. Lasagna concluded that because the 95% confidence interval around the best estimate of relative potency embraced 1.00, his study did not demonstrate a difference in effect (Tr. 10519–22).

436. The analgesic bioassay methodology posits that variability in pain relief response among test subjects does not affect the validity of a bioassay, but rather its precision, namely the confidence limits obtained (Tr. 5033–34). For this reason, it is thought to be appropriate to eliminate subjects with mild or slight pain in order to increase the statistical power of a bioassay (Tr. 5432). Dr. Sunshine testified that while there is a big difference between severe and moderate pain for the individual test subject, for evaluative purposes the only difference will be in terms of SPID scores (Tr. 9733–34, 9754).

437. Dr. Sunshine testified that he is unaware of any correlation between initial pain and any other initial variable such as age, sex, type of delivery, type of anesthetic (Tr. 9726) and has been unable to find any consistent difference in pain relief between those in moderate and those in severe pain (Tr. 9655, 9733). He would conclude, therefore, that it is just as easy for a patient in severe pain to move the three points from severe to zero as it is for the patient in moderate pain to move the two points from moderate to zero (Tr. 9733).

438. On the other hand, it is also thought appropriate in bioassays of mild analgesics, such as aspirin and aspirin-order drugs, to exclude subjects with extreme pain. For example, Dr. Moertel and Dr. Smith agree that in studies of post-partum patients the shorter the post-delivery period the less effectively all tested medication performed (Tr. 5649–50, 5433–34). Some post-surgical patients will not respond to aspirin in the first 24 hours because the severity of the injury overwhelms the aspirin order drug (Tr. 11705–06; BMF 952).

439. From the foregoing (F. 436–38), two observations may be made. *First,* a significant imbalance in the baseline pain (or initial pain) among treatment groups can seriously distort the pain relief scores for treatment groups and thereby lead to false or misleading conclusions of relative potency. This observation is equally valid in cases where such baseline imbalance remains after randomization procedure is followed. **[119]**

440. *Secondly,* the applicability of a relative potency estimate obtained from a bioassay of subjects whose baseline pain varied from mild to moderate to severe, to the population with *mild* pain is highly doubtful since two analgesic drugs having relative potencies of above one may in fact be equally effective for the relief of *mild* pain. This observation is valid unless the bioassay studied enough subjects with *mild* pain so that the average pain relief scores of the mild subgroup can be meaningfully compared in order to determine whether the same ratio holds true for *mild* pain (Tr. 5040–44).

441. Statistical significance is an effort to reduce to an acceptable minimal level the likelihood that a particular result is due to chance, but the absence of statistical significance does not necessarily mean that there is no difference (Tr. 9696).

442. Reserved.

443. Clinical pharmacologists generally determine the sensitivity of an analgesic bioassay model by its ability to differentiate the standard drug (usually aspirin) from a placebo (F. 389, *supra*).

444. In cases where a clinical pain study capable of differentiating aspirin from a placebo fails to show statistically significant difference between aspirin and a test drug, two inferences are possible: (1) that there is no statistically significant difference between aspirin and the

test drug, or (2) that whatever difference there may exist is not significant enough to be differentiated by the study model used. In either event, it is reasonable to conclude that the two drugs are about equally effective for all practical purposes.

445. On the other hand, in cases where sound pharmacological reasoning, especially when coupled with a number of clinical studies showing some difference, suggests that there might be differences between the standard and test drugs, some clinical pharmacologists are inclined to attribute the failure to show statistically significant differences to the insufficient sensitivity of the test model. However, that assumption, rational as it may be, remains to be proven by future clinical trials with more sensitive methodology (Tr. 5081, 5979, 7953, 8087, 9006–07; CX 514, p. 35481). In the absence of well-controlled clinical studies showing statistically significant differences, the claim remains unsubstantiated.

446. For example, in discussing the Bufferin studies submitted to the DESI panel Dr. Beaver stated: "So all one could possibly get out of these three studies in relation to the speed of onset claim, is that these medications will be given to a substantial number of people, and they didn't see any difference, but no one on the panel was so naive as to assume that either individually or together these three studies proved there was no difference." The Lasagna & DeKornfeld Study (*see* BMF 661) [120] did not show that there was no difference between Excedrin and aspirin but only that they did not find any difference (Tr. 12006). The Lasagna Naloxone article and the Kruskal Encyclopedia of Statistics confirm that "lack of statistical significance at a conventional level does not mean that no real effect is present. It means only that no real effect is clearly seen from data." (Tr. 10360–61).

447. In a bioassay study, if the lower limit of the relative potency estimate is greater than 1, clinical pharmacologists assume, for dose-finding purposes, that there are significant differences in efficacy along the entire range of doses (Tr. 8903, 8949).

448. "Statistical significance" does not necessarily mean "clinical significance." Generally speaking, clinical pharmacologists determine clinical significance of a statistically significant difference by certain clinical standards, such as the magnitude of the difference shown and side effects. However, there is no clear agreement among clinical pharmacologists regarding specific standards. *See e.g.,* Tr. 8902–03.

449. For example, Dr. Forrest testified that he would like to be able to infer clinical significance from a given statistically significant value but has been unable to get agreement among his peers (those clinical pharmacologists who are knowledgeable in the area of clinical

testing) (Tr. 8943–45, 9147). He would accept as clinically significant, statistically significant differences if obtained by several researchers (Tr. 8945) and he would accept, as clinically meaningful, what consumers agree is clinically significant (Tr. 9144).

450. However, the term "clinical significance" is also used in a nontechnical sense by practicing physicians, on the basis of individual judgment in their clinical practice. What clinicians usually do is to "eyeball" the difference observed in a clinical study and form a pragmatic judgment as to whether the test drug is preferable or worth trying for his patients, namely whether the difference reported is "clinically significant" in his professional opinion as a physician. Some clinical pharmacologists involved in comparative testing of oral analgesics make clinical judgments based upon the best evidence available from clinical trials and use whatever information they have as a basis for clinical judgment (Tr. 10210–24, 10240–47, 10423–28, 10461–62).

451. Reserved.

452. Practicing physicians use relative potency estimates in order to determine what dosage of one medication is needed to obtain the same effect as another. When treating individual patients, they will consider all the available evidence, even where confidence levels embrace one, weigh the possible risks and on that basis [121] reach a decision regarding what medication to prescribe (Tr. 9803, 10244).

453. Among biostatisticians, and clinical pharmacologists trained in analgesic bioassay studies, there is a school of thought that does not insist on statistical significance at the conventional 95% confidence level (P 0.05). With respect to relative potency estimates, they do not insist that the null hypothesis of equipotency be rejected at the 95% level. Generally speaking, the higher the confidence desired, the wider the confidence intervals. The lower the confidence, the narrower the intervals (Tr. 4868).

454. According to the International Encyclopedia of Statistics, by Kruskal and Tanur, an authoritative compendium recognized in the field, "probably the most common significance levels are .05 and .01, . . . but *special circumstances* may dictate tighter or looser levels. In evaluating the safety of a drug to be used on human beings, one might impose a significance level of .001. In *exploratory* work, it might be quite reasonable to use levels of .10 or .15, in order to increase power. What is of central importance is to know what one is doing and in particular to know the properties of the test that is used." (Tr. 10205–06) (emphasis added).

455. According to Finney, " 'by adequate precision' [of an estimate of relative potency] is meant a deviation of the estimate from the true value, almost certainly too small to be of any practical importance in

affecting any action to be based on the assay." (Tr. 10253). This is certainly the case when relative potency estimates are being used for dose-finding purposes. Even for dose-finding, it is agreed that more precision is required when the drug might be ineffective or toxic at either end of the dose range (Tr. 10221).

456. In a bioassay study, if the lower limit of the relative potency estimate is significantly different from 1, then it is proper to assume that there are significant differences in efficacy along the entire range of doses (Forrest, Tr. 8901, 8949). Dr. Brown testified that a lower confidence limit greater than 1 will reject the null hypothesis of equipotency (Tr. 8126).

457. In hypothesis testing, to determine that the relative potency is statistically significantly greater than 1, the lower confidence limit should not embrace 1 at the 95% confidence level (Tr. 8926–27) (RMF 1061).

458. Dr. Brown testified that data reported as having indeterminate or infinite confidence limits cannot be usefully reported to obtain an estimate of relative potency, but that, with finite confidence limits, a meaningful conclusion can be drawn. However, Dr. Brown, along with Dr. Forrest, published **[122]** data and conclusions in his article entitled "Clinical And Statistical Methodology for Cooperative Clinical Assays for Analgesia" that appeared in *Clinical Pharmacology and Therapeutics,* even though the confidence limits obtained were infinite (Tr. 4871, 5009).

459. The upper confidence interval of relative potency estimates becomes especially important in a study of a drug highly toxic at high dose levels in view of the toxicity danger to human subjects (Tr. 10222–24, 10207).

460. Drs. Brown and Forrest in their naproxin article reported their results "with reasonable confidence" even though the confidence limits were indeterminate and embraced 1 (Tr. 5117–21). The Brown/Forrest naproxin study produced infinite ("non") confidence limits for each medication in each hospital and when pooled the limits became finite but still embraced 1 and were "fairly wide," "seven fold" (Tr. 5121). Notwithstanding that even the pooled confidence limits embraced 1, and did not reject the hypothesis of equipotency (Tr. 5123), Drs. Brown and Forrest estimated the relative potency "with reasonable confidence." (Tr. 5127). Dr. Brown explained his use of the term "reasonable confidence" in the naproxin article by saying "the confidence intervals are of the length one ordinarily [finds] those assays." (Tr. 5124–25).

461. Dr. Brown testified that if the lower confidence limit is below 1, as he found in his analysis of the Emich Study, the data cannot reject the null hypothesis of equipotency. However, when his program

error was pointed out and the lower confidence limit rose above 1, he claimed that the difference was of little "practical consequence" and a matter of judgment and opinion (Tr. 8125–26).

462. Dr. Laska testified that the fact that a lower confidence limit falls below 1 is not of a practical consequence in the context of mild oral analgesics since it is the estimate of relative potency that clinical pharmacologists use and accept (Tr. 10240–41).

463. The wide fiducial (confidence) limits found in many oral analgesic studies indicate low statistical power. The reasons for these wide limits include high between-patient within-dose variability, the small sample size and small slopes of the dose response regression. (Tr. 10321).

464. Complaint counsel have agreed that reputable scientific journals on occasion publish studies with P values (confidence limits) greater than 0.05 (Tr. 5460–61), and Dr. Moertel will consider data with a confidence level of P= .065 as borderline (Tr. 5659–60). **[123]**

465. The learned journals in the field of biomedical sciences as a rule adhere to the 95% confidence level of statistical significance. The FDA generally requires, in New Drug Applications, that efficacy and safety be demonstrated at the 95% confidence level. Biostatisticians and clinical pharmacologists generally adhere to the same level of confidence.

466. On the basis of the record as a whole, it is found that, for the purposes of showing that a comparative efficacy or safety claim for an OTC analgesic product is scientifically proven or established, no lesser standard should be accepted.

C. *It Has Not Been Established That Excedrin Is A More Effective Pain Reliever Than Aspirin Or Any Other OTC Analgesic*

1. The Ingredients in Excedrin

467. Each Excedrin tablet contains four ingredients: aspirin (3 grs.), acetaminophen (1.5 grs.), salicylamide (2.0 grs.) and caffeine (1.0 gr.) (F. 2, *supra*). The fact that Excedrin contains four ingredients does not establish its superiority over aspirin or any other nonprescription internal analgesic (F. 400, *supra*).

468. Salicylamide has not been established as an effective analgesic (Beaver, Tr. 6050). The FDA OTC Analgesic Panel confirmed that further well-controlled clinical studies of the compound must be performed to demonstrate that salicylamide alone has adequate and consistent analgesic activity. The Panel concluded that salicylamide is ineffective in currently recommended doses of 300 to 600 and has not been adequately tested for safety. Therefore, it placed the drug in Category III (CX 514, p. 35441). The FDA Panel also stated that "there

are insufficient data to determine that salicylamide is either safe or effective when used in combination as an OTC analgesic in the currently marketed dosage of 97.2 to 400 mg." (CX 514, p. 35439). The Panel classified as Category III combinations of aspirin or acetaminophen with salicylamide because there is "insufficient information to determine the safety and effectiveness of salicylamide as an adjuvant..." (CX 514, p. 35442). Category III was defined as a classification for which the available data are insufficient to permit final classification as either Category I (generally recognized as safe and effective and not misbranded) or Category II (not generally recognized as safe and effective or misbranded) (CX 514, pp. 35347–48).

469. Caffeine is not an effective analgesic. The FDA Analgesics Panel so concluded and placed it in Category II as an analgesic (CX 514, p. 35482; Beaver, Tr. 6050). Moreover, the effect of caffeine as an adjuvant to aspirin or acetaminophen has not been established (Forrest, Tr. 9107). After a careful [**124**] review of the literature and data submitted by drug firms, the Panel concluded that more clinical studies need to be done to show that caffeine contributes to the claimed analgesic adjuvant effect (CX 514, pp. 35483, 35485). Therefore, the FDA OTC Analgesics Panel classified the adjuvant effect of caffeine as Category III (CX 514, p. 35484).

470. Two editions of the *AMA Drug Evaluations* (CX 512 and 518), a reliable and well recognized text on drug therapy (F. 223, *supra*), found no evidence that caffeine in the amounts present in a combination product like Excedrin has any effect on analgesic activity (CX 512I; CX 518G).

471. The *Medical Letter* (CX 510), a reliable and well-recognized publication (F. 227, *supra*), reviewed evidence concerning the addition of caffeine to aspirin, and found that it had never been adequately demonstrated that the addition of caffeine to analgesics produced any difference in analgesic effect (CX 510).

472. Respondent's Medical Director, Dr. Lanman, relied upon a study by Booy *et al.*, published in Holland in 1975 and translated into English, as support for its position on the adjuvant effect of caffeine in analgesic combinations. The Booy study was performed over a two-day period on outpatients with pain from tooth extraction, and it purported to show enhanced analgesia with an acetaminophen/caffeine combination product (Lanman, Tr. 11515–18; 12066). In fact, this purportedly enhanced analgesia was only apparent on the first day of the study (Lanman, Tr. 12080). On the second day, the combination (with caffeine) apparently performed *poorer* than the acetaminophen alone (Lanman, Tr. 12080; CX 514, p. 35484). The authors made no finding of statistically significant results on either day's data (Lanman, Tr. 11524–26). The authors presented their data in a manner

that obscured potential differences in the performance of the compounds studied (Lanman, Tr. 12068). The data as reported by the authors may have resulted from any number of performance results of the compounds with and without caffeine. In fact, the data reported by the authors cannot reject a proposition that there was no difference in the performance of the compounds, or that the compound without caffeine actually performed better than the compound with caffeine (Lanman, Tr. 12069–82; CX 904; CX 905; CX 906). Because of the authors' failure to report any statistically significant results in their study, the reversal on the second day of the first day's favorable trend, and the highly ambiguous way in which the results were reported, permitting the data to be interpreted either as supportive or contradictory to respondent's position, the Booy study cannot be given any weight with respect to the issue of caffeine's adjuvant effect. The Booy study was considered by the FDA Panel on OTC Analgesics as part of its review which led to the conclusion that there are **[125]** insufficient data to support the adjuvant effect of caffeine (F. 469, *supra;* CX 514, p. 35484; Lanman, Tr. 12213).

473. Respondents also offered a recent study by Wojcicki *et al.,* published in Poland and translated into English, as support for its position on caffeine. This was in part an outpatient study, and one of the two groups under study suffered from common headache (Lanman, Tr. 11513, 12088). This study purported to confirm the results of the Booy Study. Like the Booy Study, however, the authors of this paper failed to report any test of the statistical significance of their results (Lanman, Tr. 11526). Moreover, from the published report one cannot judge the adequacy of controls employed to assure the blinding in the study (Lanman, Tr. 12083–84). Most important, the authors analyzed and reported the results in terminology different from that used in the study (Lanman, Tr. 12084–91). For example, outpatient subjects were asked to fill in the results of treatment as "pain disappeared," "pain markedly reduced," "pain unchanged" or "pain worse" (Lanman, Tr. 12084). The authors reported the results, without any explanation, as "no more pain," "pain greatly improved," "pain slightly improved" and "pain unchanged" (Lanman, Tr. 12084–85). It is possible, as Dr. Lanman speculated, that subjects' "pain markedly reduced" responses were split into "pain greatly improved" and "pain slightly improved," although, from the questions asked subjects, there was no such gradation employed (Lanman, Tr. 12085–86). The same problem is repeated on data gathered from inpatients, *i.e.,* the data reported do not correspond to what the authors say they asked on the questionnaire (Lanman, Tr. 12087–91). Bristol-Myers obtained from Dr. Wojcicki, and offered in this case, statistical analyses purporting to show statistical significance of his findings. How-

ever, by his own analyses, the study could not differentiate 1,000 mg. of aspirin, an admittedly effective dose, from placebo (Lanman, Tr. 12091–95). The reliability of the Wojcicki Study, therefore, is subject to serious doubt. In another context, Dr. Elvers, Bristol-Myers' Associate Medical Director, took the position that the presence of a significant difference between aspirin and placebo is a "mandatory prerequisite towards the drawing of any meaningful conclusions" from an investigation of clinical analgesia (Lanman, Tr. 12093).

474. For all of these reasons, the Wojcicki study cannot be considered a well-controlled study or a reliable authority and is entitled to little weight on the issue of whether caffeine adds to the analgesia of aspirin and acetaminophen.

475. Respondent also relied on a recent blood level study by Dahanukar, published in an Indian journal as support for its position on caffeine (Lanman, Tr. 11518–19). The study was limited to 12 subjects (Lanman, Tr. 11519). The study did not measure the comparative effectiveness of compounds with and without caffeine (Lanman, Tr. 11519). Blood level studies have **[126]** not been accepted as evidence of degree of analgesia because no relationship between blood levels and degree of analgesia has been established (F. 401, *supra*). This study therefore is entitled to little weight on the issue of the potentiating effect of caffeine.

476. Respondent also relied upon a study by Houde and Wallenstein wherein the authors concluded that "the results with caffeine must be considered equivocal, although it is possible that dosage may be an important factor, and caffeine may simply be ineffective at much below the 60 mg. dose" (Lanman, Tr. 11523). In fact, this study was presented to the FDA Panel on Analgesics, which concluded that it was the only "well-controlled clinical study to determine whether aspirin plus caffeine is more effective than aspirin alone, and the results of this study are equivocal" (Lanman, Tr. 12065; CX 514, p. 35483). Even though the FDA Panel considered this study its equivocal results and the absence of other sound evidence still led the Panel to put caffeine in Category III as an adjuvant (F. 469, *supra*).

477. None of the four studies offered by respondent either alone, or in combination, are adequate support for the proposition that caffeine adds to the analgesia of aspirin and/or acetaminophen. At best, the studies produced ambiguous results (F. 472, *supra*), reported results in a manner inconsistent with the way data were generated (F. 473, *supra*), failed to incorporate tests of statistical significance (F. 472–73, *supra*), were unable to differentiate an effective dose of aspirin from placebo (F. 473, *supra*), produced equivocal results (F. 476, *supra*), or did not even measure pain relief (F. 475, *supra*).

478. The nature and quantity of ingredients in an analgesic product

is not evidence that can establish its superiority to other analgesic products (F. 400, *supra*). In fact, an Excedrin tablet contains only 4.5 grains of ingredients established as Category I analgesics (3.0 grains of aspirin and 1.5 grains of acetaminophen) as compared to the standard 5 grain aspirin tablet. It contains 3 grains of ingredients (2.0 grains of salicylamide and 1.0 grain of caffeine) which the FDA OTC Analgesics Panel has classified as either Category II (ineffective) or Category III (insufficient evidence concerning efficacy or adjuvancy) (F. 468–69, *supra*). In this light, Excedrin can be said to contain a lower amount of proven analgesic ingredients than a plain 5 grain aspirin tablet.

2. Bioassays of Excedrin and Aspirin

479. Respondent has admitted representing that Excedrin is a more effective pain reliever than aspirin (F. 272, *supra*). As primary support for its claim, it relies on the results of [**127**] studies performed on Excedrin and aspirin which, in its expert witnesses' view, adequately support that claim.

480. The Emich Study (CX 425), a bioassay study of post-partum pain conducted in 1968, the Smith Study (CX 453), another post-partum pain study conducted in 1970–1972, and the Sherman Study (CX 439), a pain threshold study of electrical-shock induced dental pain, are in evidence. Three other post-partum pain bioassays offered by Bristol-Myers were rejected, for the reason that Bristol-Myers failed to comply with administrative law judge's long-standing pretrial disclosure directions regarding medical-scientific studies to be offered at trial and that Bristol-Myers failed to show good cause for excepting the studies in question from those requirements (Tr. 9624–41). The three rejected studies are RX 164 for identification (Sunshine Study designated 16H9), RX 165 for identification (Sunshine Study designated 9T1) and RX 148 for identification (Emich Study and data designated W1409). The administrative law judge's modified ruling regarding RX 166 for identification (Sunshine Study designated 10G–12G) would have permitted Bristol-Myers to reoffer it after further interview and cross-examination of Dr. Sunshine regarding that study by complaint counsel, and Bristol-Myers chose not do so (Tr. 11393–400, 11616–18). Summary and analytical tabulations related to the excluded bioassays were likewise rejected. Bristol-Myers was permitted to make an offer of proof regarding all of the excluded material, which are contained in the excluded exhibit binder of the record. Furthermore, Bristol-Myers' expert witnesses were permitted to refer to, but not to discuss the details of, the excluded studies in explaining their opinions, especially opinions regarding the so-called "pooled data" (*See* F. 526–28, *infra*).

481. In this connection, it is noted that Bristol-Myers did not include any of the four bioassay studies (RX 148, 164–166 for identification, in the rejected exhibit binder) to the FDA OTC Analgesics Panel among its submissions in support of its claims of "extra strength" for Excedrin (Lanman, Tr. 12116–17). Dr. Sunshine, who was involved in the conduct of these studies, did not call their results to the attention of the American Medical Association when he was asked in 1971 to comment on a draft of *AMA Drug Evaluations,* which discussed the comparative efficacy of Excedrin and aspirin (Sunshine, Tr. 9702–06). The authors of the Emich Study (CX 425), which included Fred Mueller of Bristol-Myers Statistical Services department (CX 425A), which did not refer to the rejected studies in the introduction to their report, purported to review the available information on Excedrin's efficacy as an analgesic (CX 425G).

a. *The Emich Study (CX 425)*

482. The Emich Study (CX 425) is a bioassay which compares three doses of Excedrin (1, 2 and 4 tablets) to three doses of 5 [**128**] grain aspirin (1, 2 and 4 tablets) and placebo. The study included 269 female patients all suffering from post-partum pain. It began in 1968 at the Philadelphia General Hospital under the general direction of Dr. John Emich (Sunshine, Tr. 9611). Dr. Emich was an obstetrician and gynecologist, but not a clinical pharmacologist (Sunshine, Tr. 9603). Apparently, Dr. Emich had not done any bioassays before 1968, and was initiated into the techniques of analgesic bioassay by Dr. Sunshine (Sunshine, Tr. 9604–06). The authors concluded that the study showed that tablet for tablet, Excedrin is a more potent analgesic than aspirin for post-partum pain (CX 425V).

483. There was no separate protocol specifically designed for the Emich study that set forth, in advance, the design, treatments, sample size, and statistical analysis to be employed. However, Dr. Sunshine provided Dr. Emich with a copy of a protocol (BMRX 161) that had been developed for use by Dr. Sunshine in 1962 for his own studies of Bristol-Myers' analgesic products (BMRX 161, 162; Sunshine, Tr. 9612, 9617, 9620). Assuming that Dr. Emich used the Sunshine protocol, it is evident that he did not follow it. For example, the Sunshine protocol called for use of patients with surgical and fracture pain as well as obstetrical patients (BMRX 161A); the Emich Study was confined to obstetrical patients (CX 425H). The Sunshine protocol called for patients entered onto the study to be free from analgesic medication for the five hours preceding initiation of the study (BMRX 161A); the Emich Study eliminated patients who received analgesics during the previous six hours (CX 425H). The Sunshine protocol calls for a cross-over design, with each patient receiving more than one treat-

ment (BMRX 161A, BMRX 162); the Emich Study was a single dose study, in which no patient received more than one treatment (CX 425I). The Sunshine protocol called for a sample size of 200 subjects (BMRX 161B); the Emich Study tested 269 subjects (CX 425H). The Sunshine protocol called for interviews of patients to extend over a four-hour period after administration, with the first interview at one-half hour after administration (BMRX 162); in the Emich study patients were interviewed over a five-hour period after administration of the treatments, with the first interview at one hour after administration (CX 425K). The Sunshine protocol calls for a statistical analysis on "the summary variable of all the hourly relief scores" (BMRX 161B); the statistical analysis of the Emich Study employed, in part, less than all the hourly relief scores (F. 499, *infra*).

484. Dr. Laska, Bristol-Myers' expert witness, analyzed the data generated by the Emich Study through the use of a bioassay computer analysis program. RX 181A-F comprise the computer printouts of that analyses, according to six different variables: percent SPID at 5 hours, SPID at 5 hours, percent SPID at 4 hours, SPID at 4 hours, TOTAL at 5 hours, and TOTAL at 4 hours. The relative potency estimates (rho) for Excedrin to [**129**] aspirin and the associated confidence intervals at 95% confidence level, based on RX 181, are as follows (Tr. 10174–85):

Variable	5 hrs.		4 hrs.	
	Rel. Pot.	Conf. Int.	Rel. Pot.	Conf. Int.
Percent SPID	2.6	1.1-94.3	4.0	$1.4\text{-}4.8 \times 10^5$
SPID	4.08	$1.3\text{-}3.84 \times 10^{24}$	7.1	0-infinite
TOTAL	2.27	.86-255	2.32	.85-1230

Based on his computer analysis, Dr. Laska expressed an opinion that the Emich study provided "compelling evidence of superiority" of Excedrin to aspirin, in terms of pain relief provided at equidoses (Tr. 10185).

485. It should be noted, however, that, out of the two "standard" or "orthodox" analysis of SPID and TOTAL (Brown, 4908, 5086, 5106), only the SPID analysis shows statistical significance at the 5% level of confidence (or $p < .05$) whose confidence interval does not enclose 1. Thus, only the SPID analysis is able to reject the hypothesis that Excedrin and aspirin produce equal effects at 1, 2 and 4 tablet doses (F. 484; Brown, Tr. 4908, 5105; Sunshine, Tr. 9663).

486. In the Emich Study, the relative potency estimate for Excedrin to aspirin on a tablet for tablet basis is 4.02, with a lower 95% confidence interval of 1.4 (Tr. 9659). Dr. Sunshine testified that the results of the Emich Study as expressed by %SPID4 are "strong scientific

evidence that Excedrin is stronger and more effective than aspirin on a tablet for tablet basis." (Tr. 9660).

487. The relative potency of Excedrin to aspirin on a tablet for tablet basis in the Emich Study range from 2.27 to 7, with 4 of the 5 parameters significant at the 95% level and 2 having confidence intervals above 1 (Tr. 9660).

488. On January 16, 1968, the statistical department of Bristol-Myers prepared a "final report" of the Emich Study (Tr. 10613–14). That report included the data transmitted by Annette Williams' letter of December 3, 1968 on approximately 230 patients (Tr. 10614). On January 30, 1969, Annette Williams sent the data for an additional 44 patients of the study to Bristol-Myers for analysis (Tr. 10614). Despite Bristol-Myers' belief that the Emich Study had concluded with 225 patients, it nonetheless included these final 44 in its final analysis as presented in Atlantic City. It could have discarded those final cases and considered the Emich Study terminated at 225 patients (Tr. 10615). It would have been proper for Bristol-Myers to discard the results of the last 44 patients of the Emich Study, **[130]** thereby increasing the strength of the conclusions one could draw based on the Emich Study (Tr. 10619).

489. The relative potency estimate of Excedrin to aspirin for the variable SPID4 is 7.1, with a 90% confidence interval from 1.97 to 1.44 (Tr. 10183; BMRX 181D). The estimated relative potency of Excedrin to aspirin in the Emich Study using variable total 5 (TOPAR) is 2.27, with a 90% confidence interval from 1.02 to 24.1 (Tr. 10184). The estimated relative potency of Excedrin to aspirin for the variable total 4 (TOPAR) is 2.32, with a 90% confidence interval of 1.01 and 36 (Tr. 10184). The Emich Study results for the response variables SPID 4, total 5 and total 4 have lower confidence limit values above 1 at the 90% level of confidence (Tr. 10184–85).

490. The Emich Study is flawed by a problem that compromises its fundamental validity. Despite the fact that subjects were purportedly assigned to the seven treatments in the study through a randomization technique, more patients in "severe" initial pain were assigned to the Excedrin treatments than to the aspirin treatments (Brown, Tr. 5174; Sunshine, Tr. 9662). This procedure resulted in an imbalance in the baseline pain levels between the Excedrin groups and aspirin groups, before any tablet was ingested, that were large enough to be statistically significant at the .02 level (Brown, Tr. 4903, 4921; Forrest, Tr. 8960; Laska, Tr. 10199). Statistically significant imbalances in initial pain levels among treatment groups at baseline is a serious problem that cannot be ignored (Laska, Tr. 10621; Forrest, Tr. 8960–61, 9090–91; Brown, Tr. 4904–05, 4911, 5083–84, 5093–94, 5100, 8029). Respondent's expert Dr. Laska agreed that he would not have confi-

dence in using data from the conventional SPID analysis of the Emich Study due to this baseline pain imbalance (Laska, Tr. 10440, 10487–88).

491. The level of baseline pain (*i.e.,* pain prior to medication) is the single most important variable influencing the response to analgesics (Beaver, Tr. 5968; Brown, Tr. 8053, 8113, 8118–23, 8128–34). Indeed, the authors of the Emich Study themselves note that in their study "the response of an individual patient to a given medication was closely related to her starting pain level" (CX 425N). Although several experts of Bristol-Myers expressed the view that post-study correction or adjustment of the baseline imbalance problem by the use of percent SPID (as was done in the Smith Study) was not unusual, it is questionable whether such after-the-fact statistical "correction" can reasonably be expected to cure the defect and restore the validity of a flawed analgesic study to that of an unflawed one (Brown, Tr. 8113–14, 8136, 8050–53, 8060–61).

492. It is fair to say that where statistically significant baseline pain imbalance results after randomization, the result is the same as in a nonrandomized study in that the attempted **[131]** control of patient assignment bias failed. In fact, in the Emich study the assignment of larger numbers of patients in "severe pain" to the Excedrin treatments created a bias favoring Excedrin (Brown, Tr. 4094, 4936, 5174; Sunshine, Tr. 9734). The bias results from the fact that Excedrin had the opportunity to relieve more pain in more patients than aspirin did (Brown, Tr. 4904, 5174). Excedrin had more opportunity to reduce pain intensity and to provide pain relief than aspirin, because patients in the Excedrin group on the average started with more pain (Brown, Tr. 4904; Sunshine, Tr. 9734). As the authors of the Emich Study observed: "Patients who had severe pain at the outset proved to receive significantly more relief on the average than those complaining of less discomfort" (CX 425"O").

493. The practical consequence of the statistically significant baseline pain imbalance among the treatments in the Emich study is that it reduces confidence in the study, and all its results, to a point where it cannot be accorded full weight (Forrest, Tr. 8960–62, 9090–91, 9116–17; Brown, Tr. 4905, 4911–14, 4916–17, 4928, 5100, 8149–50, 8154–55). The fact that there was a statistically significant imbalance on baseline pain—perhaps the most important of all variables that influence the results of pain relief studies—raises the specter of bias in patient assignment (Brown, Tr. 4911, 4921; Forrest, Tr. 8960–62, 9091). The record shows that the chance of a true randomization may produce the baseline pain imbalance present in the Emich Study is only two (2) times out of 100 (Brown, Tr. 4903, 4921; Forrest, Tr. 8960). Respondent's expert witness, Dr. Laska, agreed that if subjects were not

assigned to treatments in an unbiased fashion, the entire study would be seriously compromised (Laska, Tr. 10590–94). While after-the-fact numerical transformations of the data may be the only plausible way to address this central problem statistically, no statistical "correction" can address the issue of whether patients were, in fact, assigned to treatments in an unbiased fashion (Brown, Tr. 4911–12, 5092–93, 8143–44; Forrest, Tr. 8960–61). Dr. Forrest, an eminent authority in the field of analgesic bioassays, and Dr. Brown, an expert biostatistician experienced in analgesic bioassays, concluded that the serious baseline pain imbalance present in the Emich Study diminishes the Study's weight to a point where they would not rely on it as credible evidence regarding the issue of whether the superiority of Excedrin over aspirin has been scientifically established (Forrest, Tr. 8960–61, 9121–23; Brown, Tr. 8108, 8149–50, 8154–55).

494. The position of Drs. Forrest and Brown regarding the weight to be accorded the Emich Study is corroborated by the fact that, apparently only one published analgesic study has been found where the authors reported statistically significant differences in initial pain levels among the treatment groups (Laska, Tr. 10626–27). Dr. Louis Lasagna, the author of that article, is a respected and well qualified clinical pharmacologist (Beaver, Tr. 5903), whom Bristol-Myers cited in [132] support of its position in its 1968 Comments to the Federal Trade Commission in a Trade Regulation proceeding involving OTC analgesics (Laska, Tr. 10626, 12023–24; Sunshine, Tr. 9721). What Dr. Lasagna concluded regarding that study was that, because of the bias introduced by the statistically significant differences in starting pain levels, he could not come to conclusions about the performance of the tested drugs (Laska, Tr. 10626–27).

495. Reserved.

496. The authors of CX 425 do not report that patients varied in terms of their initial pain to a statistically significant degree (Brown, Tr. 5174; CX 425). However, the authors do outline a technique of analysis, called "Percent SPID," which they say adjusted the "SPID" scores so they were "freed . . . from the influence of starting pain levels" (CX 4250). The authors of CX 425 do not report the estimate of relative potency based either on "SPID" or "Total" (Brown, Tr. 4906–07). The relative potency they reported was based on their "% SPID" analysis. However, the purported "protocol" for the study (BMRX 161B) did not mention "% SPID."

497. Respondent's experts Drs. Sunshine and Laska contend that the use of the % SPID analysis in the Emich Study successfully "corrects" the problem introduced by the existence of statistically significant baseline pain imbalance (Sunshine, Tr. 9659, 9662, 9671; Laska, Tr. 10199–200). However, they did not say that the use of an adjust-

ment for the SPID score (*i.e.*, the use of "% SPID") also corrects what may be the same problem with the other summary variable analyzed in the Emich Study, namely "Pain Relief" (CX 425R; F. 406–07, *supra*). In fact, Dr. Smith, the author of the Smith Study, testified that in his study, even though there was no statistically significant imbalance in starting pain levels, he "tried a variety of correction terms to eliminate any potential bias owing to the fact, that starting pain does, in fact, influence pain relief as it influences pain intensity difference" (Smith, Tr. 5421).

498. "% SPID" is a technique developed by the authors of the Emich Study for purposes of *post hoc* analysis of the data. The normal "SPID" score for each patient is expressed as a proportion of the maximum possible "SPID" score that each patient could have obtained (CX 425K, L). Respondent's expert witness, Dr. Laska, pointed to a general source as support for the type of correction provided by the "% SPID" technique. However, Dr. Laska was unable to cite any published article where the author used % SPID or any other statistical device to correct baseline pain imbalance (Laska, Tr. 10626). Dr. Sunshine, who claimed that baseline imbalances occurred frequently in studies during the 1960's, cited no article that employed an analysis on the % SPID variable or any other **[133]** "correction," and he admitted that he had not used % SPID in any of his published studies (Sunshine, Tr. 9717–20, 9746).

499. Bristol-Myers' experts analyzed the results of the Emich study at four and five hours after administration of the treatments (Sunshine, Tr. 9659; BMRX 181C, D, F). Dr. Laska testified that the four-hour analysis is meaningful because both Excedrin and aspirin recommend a four-hour interval between doses (Laska, Tr. 10548–49). The analysis of "% SPID," "SPID" and "TOTAL" at the four-hour period is a *post hoc* analysis outside the purported protocol's specification that the summary variable analysis cover "all the hourly relief scores" (BMRX 161B; Laska, Tr. 10540–41, 11292–93). When he was asked why he had not, for example, analyzed the summary variables in the Emich Study based on data from three-hour or even two-hour observations, Dr. Sunshine answered that "you can do anything you want . . . It depends what you're looking for" (Sunshine, Tr. 9707). Respondent's expert Dr. Laska admitted that in his published work, and in that of Dr. Sunshine, when the summary variables "SPID" and "TOTAL" were analyzed *all* of the data generated in the studies was included (Laska, Tr. 10548–51). Analysis of a data segment not laid out in advance in the protocol, is "data massaging that destroys the validity of the analysis" (Moertel, Tr. 5543).

500. Even if one were to dismiss the gravity of the baseline pain imbalance problem and accept the % SPID "correction," the Emich

Study is equivocal. Out of the six variables analyzed by Bristol-Myers' experts, only four will give an unbiased estimate of the relative potency because the uncorrected SPID analysis (SPID–5 and SPID–4) is infected with a quantitative bias introduced by the initial pain imbalance (Laska, Tr. 10440, 10487–88; Sunshine, Tr. 9662, 9671). Of that four, two have estimates of relative potency with confidence intervals that embrace 1.00 (RX 181E-F; Tr. 10183–84). Thus, accepting all of respondent's "corrections" and *post hoc* analyses, only two of the four parameters in the Emich Study analyzed by respondents which could give an unbiased estimate of relative potency, reject the hypothesis that Excedrin and aspirin are equally effective.

501. The Emich Study was submitted for publication in the *Journal of Clinical Pharmacology and Therapeutics*. The authors were asked by Dr. Modell, the Journal's editor, to comment on the generalizability of the study results to pain etiologies other than post-partum. They answered that the issue was irrelevant (CX 910). Their study was not published (Lanman, Tr. 12095–97).

b. *The Smith Study (CX 453)*

502. The Smith Study (CX 453) is a bioassay which, like the Emich Study, investigated three doses of Excedrin (1, 2 and 4 [134] tablets), three doses of aspirin (1, 2 and 4 tablets) and placebo (Smith, Tr. 5393). The study was funded by Bristol-Myers and involved 785 female patients (about three times the sample size of the Emich Study) suffering from post-partum pain at the Boston Hospital for Women (Smith, Tr. 5392–93). The study was conducted during the period commencing in the fall of 1970 through January 1972 (Smith, Tr. 5392) under the direction of an experienced, reputable investigator, Dr. Eugene Smith, of the Harvard Medical School and Massachusetts General Hospital (F. 59, *supra*).

503. The protocol for the Smith Study was reviewed and approved by the Research Committee of the Massachusetts General Hospital to ensure that the study followed scientifically appropriate and accepted procedures (Smith, Tr. 5393–94).

504. The primary purpose of the Smith Study (CX 453) was to investigate not only the efficacy of Excedrin but the influential variables that may affect clinical trials generally and to develop a method of investigation and to study the relative potency of Excedrin (Tr. 5445).

505. The Smith Study was well-designed, employed the appropriate controls, and suffered from none of the problems which characterized the Emich Study (Brown, Tr. 8150). All significant variables were satisfactorily balanced across treatment groups (Smith, Tr. 5434, 5506–07). Moreover, all methods of analysis employed in the Study yielded

consistent results: none of the analyses showed statistically significant differences at the tested dose levels between Excedrin and aspirin at the .05 level (Smith, Tr. 5422–24); all of the analyses produced relative potency estimates between 1.1 and 1.3 with lower 95% confidence limits around .50 to .70 (F. 506–08, *infra*).

506. The results of the Smith Study are:

Parameter	Estimate of Relative Potency	Lower Confidence Limit	Upper Confidence Limit
TOTAL4	1.36	.51	7.3
%SPID5	1.25	.69	2.57
SPID5	1.13	.54	2.64
%SPID4	1.33	.7576	2.78
SPID4	1.22	.59	3.04
TOTAL5	1.2	.45	4.35

(Tr. 10294–95) (BMRX 182).

507. Data generated in the Smith Study were analyzed for the five-hour period over which the study ran (Smith, Tr. 5413). The relative potency estimates for Excedrin to aspirin, are 1.13 based on "SPID–5," with 95% confidence limits of .54 to 2.64 **[135]** (BMRX 182B; Laska, Tr. 10294), and 1.2 based on "TOTAL–5," with 95% confidence limits of .45 to 4.35 (BMRX 182E; Laska, Tr. 10294). Neither of the two conventional analyses rejects the null hypothesis that Excedrin and aspirin are equally effective because the lower 95% confidence intervals enclose 1.00 (Smith, Tr. 5423; Laska, Tr. 10426–27; Brown, Tr. 4933–35; Forrest, Tr. 8963–65; Sunshine, Tr. 9751). Thus, neither analysis reflects a statistically significant difference between Excedrin and aspirin at the .05 level at the tested dose levels (Smith, Tr. 5422–24).

508. Dr. Laska, Bristol-Myers' expert, also analyzed the results of the Smith Study using the % SPID method. Since there is no baseline imbalance on initial pain in the Smith Study, and therefore no bias for using % SPID to "correct" it, the results according to %SPID–5, not surprisingly, closely parallel the results of the normal SPID–5 analysis (Brown, Tr. 4936, 8144–45). The relative potency estimate based on %SPID–5 was 1.25, with 95% confidence limits of .69 to 2.57 (BMRX 181A; Laska, Tr. 10294). The four-hour data analyzed by respondent is also consistent with the five-hour data analysis. The "best estimate" and associated 95% confidence intervals for SPID–4, %SPID–4 and TOTAL–4 were, respectively: 1.22 (95% limits of .59 to 3.04) (BMRX 181D); 1.33 (95% limits of .75 to 2.78) (BMRX 181C); and 1.36 (95% limits of .51 to 7.27) (RX 181F). (See Laska, Tr. 10294–95). Each of these four analyses produces a relative potency estimate with a 95% confidence interval well below 1.00. Thus none of them show

a statistically significant difference between Excedrin and aspirin at the .05 level at the tested dose levels. Indeed, the data from the Smith Study, however analyzed, cannot reject the hypothesis that aspirin is more potent than Excedrin at the tested dose levels (Laska, Tr. 10518). The results from the Smith Study are quite consistent with the results that would be obtained in a bioassay where the *true* relative potency of the two compounds was, in fact, 1.00 (Brown, Tr. 5009, 8157–58).

509. The Smith Study showed that for mild pain, the relative potency of aspirin compared to Excedrin is 2.3, with infinite confidence intervals due to the small sample size (Tr. 10301) (BMRX 182).

510. The Smith Study is a more precise and reliable estimate of the relative potency of Excedrin to aspirin than is the Emich Study (Laska, Tr. 10537). It suffered from no methodological flaws that compromised either its reliability or its weight (Brown, Tr. 8150). Moreover, it employed more subjects than the Emich Study (785 vs. 269). 785 is a large sample for bioassay studies of this kind (Forrest, Tr. 8965). Dr. Beaver referred to sample sizes in analgesic studies of 675 to 750 patients as "gigantic" (Beaver, Tr. 6023). Dr. Sunshine indicated that 30 patients per treatment would be a "ballpark" minimum adequate sample size, and having 50 patients per treatment group would be [136] "wonderful" (Sunshine, Tr. 9773). The Smith Study had about 100 patients per treatment (Forrest, Tr. 8964). Generally, the larger the sample, the easier it is to show differences between the two compounds, if there are in fact differences (Forrest, Tr. 8965). The fact that Dr. Smith is a well-known researcher adds to the reliability of his study (Brown, Tr. 8150). For all of these reasons relating to the precision, sample size, and methodological elegance, the results of the Smith Study should be accorded greater weight than the Emich Study regarding the issue of whether Excedrin's claimed superior efficacy over aspirin has been scientifically established (Moertel, Tr. 5597; Brown, Tr. 8150).

511. Respondent has "pooled" the results of the Emich and Smith Studies in order to produce yet another analysis of their results. Essentially, "pooling" is a statistical device that combines the "best estimates" of relative potency, together with other data bearing upon the variability in each study, and produces a "pooled" estimate, with a new set of 95% confidence limits (Laska, Tr. 10319–50; Forrest, Tr. 8965–74). However, "pooling" the Emich and Smith Study data does not create a new, well-controlled study, whose results can be used to establish a claim of superior efficacy. It may be said that pooling reduces the two independent Emich and Smith Studies to one "pooled" study (Forrest, Tr. 8965–68; Brown, Tr. 8159–63). In order to establish a scientific proposition, one needs replication of the statistically significant results of one study by another study (F. 370, *supra*).

What is required is at least two well-controlled clinical studies which demonstrate statistically significant differences between the compounds tested. Pooling does not meet that requirement (Forrest, Tr. 8967–68; Brown, Tr. 8161).

512. "Pooling" combines the results of several studies to arrive at an overall conclusion of relative potency estimates on the basis of available data, across a variety of studies, investigators and locations (Tr. 10186–99, 11312–13).

513. The information pooled includes the relative potency estimates, sample sizes, slopes, sums of the squares and the confidence limits, intervals and values (Tr. 10193).

514. The rationale for pooling is to use all available information in an attempt to obtain an overall estimate of what the true relative potency is (Tr. 10188–89).

515. Finney would restrict pooling of "assays in which different species of animals have been used as subjects or different measurements have been taken as responses or experimental techniques have been fundamentaly different. . . ." (Tr. 10335). In pooling data from more than one hospital, Finney would calculate the relative potencies for each hospital and then pool them using the Bennett method (Tr. 8969). **[137]**

516. Dr. Laska testified that data from different investigators can be pooled so long as it is collected in a reasonably similar way and/or if the several studies are conducted under the same or similar circumstances. For example, subjective response studies would not be pooled with animal or experimental pain studies. Further support for this proposition is seen in the Naloxone Article by Dr. Lasagna (Tr. 8970, 10196, 10324).

517. Dr. Laska testified that he finds support for the pooling of all the Excedrin studies in Bennett ("Combining Estimates of Relative Potency and Bioassay") and Armitage ("Point and Interval Estimation in the Combination of Bioassay Results") (Tr. 10337–40).

518. According to Dr. Laska, pooling is permissible when: (1) the estimate of relative potency for each of the studies is within the confidence intervals of both of them or (2) when one of the estimates of relative potency is within both intervals and the upper limit of one study is below the estimate of the other (Tr. 11306–08).

518a. Dr. Forrest's VA co-op study pooled data from the several hospitals, including data with infinite limits to obtain one relative potency estimate with finite confidence limits (Tr. 5010–11).

519. In cases where the validity of a relative potency estimate is sufficiently demonstrated by a well-controlled bioassay whose findings are then replicated by another well-controlled bioassay by an independent investigator, the rationale for pooling the data from the

two studies with those of others which are flawed and/or fail to show significance at reasonable confidence levels, is difficult to understand to a layman.

520. However, absent two or more well-controlled studies confirming the validity of a relative potency estimate, pooling may be a statistically acceptable device for obtaining a composite estimate on the basis of available information if one must come up with a relative potency estimate. This is akin to the pragmatic approach by which clinicians not well versed in analgesiology assess analgesic bioassay reports (F. 450, *supra*). [138]

521. The pooled results of the Emich and Smith Studies are:

Parameter	Estimate of Relative Potency	Lower Confidence Limit	Upper Confidence Limit
%SPID5	1.58	.99	3.05
%SPID4	1.82	1.12-3	3.88
SPID5	1.67	.95	4.06
SPID4	1.97	1.08	5.92
TOTAL4	1.65	.86	5.03

(Tr. 10311, 10313–14) (BMRX 63).

522. The %SPID5 pooled result of Emich and Smith rejects the null hypothesis at a confidence level of P=.10 or 90% (Tr. 5155).

523. In order for the lower confidence limit of the pooled Emich and Smith studies to rise above 1, the P-value would have to be approximately .06 to .08 (Tr. 10314–15) (BMRX 63).

524. According to Dr. Laska, a reanalysis of the Smith data by baseline pain level shows relative potency estimates (Excedrin to aspirin) of: 2.3 for mild pain, 1.3 to 1.5 for moderate pain, and 1.6 to 1.8 for severe pain. The relative potency estimate for mild pain (2.3) had infinite confidence intervals due to the small size of the subsample. The confidence intervals for relative potency estimates for moderate and severe pain were undetermined (Tr. 10301–02; BMRX 182).

525. Dr. Laska testified, based on his reanalyses of the Emich and Smith data, when combined, show that for moderate pain the relative potency of Excedrin to aspirin is 1.26, with 95% confidence intervals of .54 to 3.51 and for severe pain is 1.82, with 95% confidence intervals from .88 to 10.02 (Tr. 10305–06).

526. Dr. Sunshine also referred to two other studies of his own, both of which compared the potency of Excedrin and aspirin, using postpartum pain subjects and the Sunshine protocol (RX 166 and 168 for identification). They were both rejected, but Dr. Sunshine was permitted to refer to them in his answers to the ALJ's questions regarding the applicability of bioassays to moderate pain. The first, Hopper Study (16H9) (RX 168 for identification) used 1, 2 and 4 tablets. The

second, Gueria Study (10G1) (RX 166 for identification), used 2/3 of a tablet, 2 and 6 tablets. Both used a modification of the statistical technique used in Emich and Smith, and compared one dose of Excedrin and three dose levels of aspirin (Tr. 9643–45).

527. Dr. Sunshine conducted *post hoc* stratification analyses of the Emich, Smith, Hopper and Gueria Studies in order to determine Excedrin's relative potency for the moderate pain subset of the patient samples and testified that every one of the four studies produced a relative potency estimate of above 1 **[139]** for the moderate pain subgroup. "1.5, 2, 4, depending on the study. There was variability. But in each and every time, it was greater [than 1]. And . . . if you just average it up, it was one-and-a-half times greater." (Tr. 9784–85).

528. Although Dr. Sunshine's above analyses are interesting, they are of little value, for several reasons. First, setting aside several objections to post hoc analyses of subset data, the size of the subset of test patients and the moderate pain group in those studies was clearly inadequate. Dr. Sunshine was emphatic that any subgroup analysis of less than 30 would lead to "distortion" and be incapable of providing any "meaningful data." (Tr. 9769–70). For example, he agreed that the subsample size of less than 17 per treatment in the Emich Study was inadequate for a valid or meaningful post hoc stratification analysis of the uterine pain subgroup and suggested that 30–50 would be reasonable (Tr. 9769–73). In the Emich Study, the size of the moderate pain subsample was less than 15 (Tr. 9719). Dr. Brown also testified that stratification analysis is not valid unless each pain group contained enough subjects and different results showed up (Tr. 5038–40). Further, the Emich Study excluded the mild pain group and no analysis of that study for the mild pain patients is possible. Further questions regarding the applicability of post-partum pain studies to other types of pain have been noted (F. 374–79, *supra*).

529. Dr. Laska introduced a novel analysis of the data generated by the Smith and Emich Studies for the purpose of demonstrating the magnitude of differences in the effectiveness of Excedrin and aspirin (Laska, Tr. 10354–59). Dr. Laska in effect subtracted from the effect level of both Excedrin and aspirin, the effect level of placebo in each study for %SPID–5, and calculated the percentage difference in the *remaining* effect between Excedrin and aspirin (Laska, Tr. 10358, 10444–45, 10475, 10481–82; CX 900 (graph "e"); CX 901). Applying the novel analysis to the Emich Study, Dr. Laska concluded that Excedrin added about 59% to the effectiveness of aspirin over and above what is supplied by placebo (Laska, Tr. 10358; CX 901). Using the same "Laska" formula, Dr. Laska calculated from the Smith Study that Excedrin adds approximately 10% to the pain relieving effectiveness

of aspirin over and above what is supplied by placebo (Laska, Tr. 10358–59; CX 900 (graph "e"); CX 901). The statistical significance of any of these purported differences is not shown.

530. Although complex statistical tests could be performed to test the significance of these percentage differences, Dr. Laska agreed that one could simply use the 95% confidence interval around the best estimate of relative potency on "%SPID–5" to test statistical significance of these percentage differences in effectiveness. If the 95% confidence interval embraced 1.00, then the percentage difference in comparative "%SPID–5" effectiveness of Excedrin and aspirin would not be statistically **[140]** significant (Laska, Tr. 10468). Using this method, since all estimates of relative potency in the Smith Study have 95% confidence intervals that embrace 1.00 (F. 507–08, *supra*), all measures of comparative effectiveness of Excedrin and aspirin would not be statistically different at the .05 level. And of the four relative potency estimates in Emich that Dr. Laska would use for this purpose (Laska, Tr. 10440, 10487–88), only two have 95% confidence intervals that do not embrace 1.00 (%SPID–5 and %SPID–4) (Laska, Tr. 10468–69).

531. Nonparametric analysis takes into account repetitive events which lead to the same general conclusion but none of which independently supports a firm conclusion. A nonparametric analysis of Excedrin's strength compared to aspirin addresses the question of how many repeated trials show Excedrin with a relative potency greater than 1. Pooling addresses the issue of the actual relative potency of one treatment to another (Tr. 10189).

532. In nonparametric analysis, the frequency of results showing relative potency estimate of above 1, is the determining factor. The greater the frequency, the stronger the evidence showing the superiority of one treatment over another (Tr. 10186).

533. BMRX 211 is a graphic representation of the nonparametric analysis of the Emich and Smith Studies showing the estimates of relative potency for each pain condition studied in those tests (indicated by dots), the overall estimate within each study for relative potency (indicated by an X), and the confidence intervals around the estimate of relative potency for each study (indicated by a solid vertical line). BMRX 211B indicates the overall pooled estimate of Excedrin's relative potency (indicated by a circled X) and the 95% confidence interval around that estimate. BMRX 211B shows that Excedrin is superior to aspirin (Tr. 10317).

533a. Nonparametric analysis essentially eyeballs the data generated by a number of studies and attempts to reach an overall observation regarding a general trend either favoring or disfavoring a proposition.

3. The Sherman Study on Experimentally Induced Dental Pain (CX 439)

534. CX 439, entitled *Comparison of the Effectiveness of Two Analgesic Agents by Laboratory Testing,* ("Sherman Study"), is the report of an experimental pain study conducted in 1962 for Bristol-Myers, which purported to compare the relative analgesic effectiveness of 600 mg. of aspirin and two tablets of Excedrin on a double-blind basis (CX 439D; Elvers, Tr. 10771). The Sherman Study presupposes a direct correlation between the clinical effectiveness of analgesics and their ability to raise **[141]** the pain threshold in artificially induced pain, the level of pain intensity at which an experimental pain stimulus is perceived by a subject as first causing pain. In this study, pain was induced by applying electrical shocks to the dental pulp of a selected tooth of test subjects (CX 439N).

535. The Sherman Study was authored by Drs. Harold Sherman, Joseph E. Fiasconaro and Harry Grundfest (CX 439A). Dr. Sherman was a dentist on the faculty of the dental school of Columbia University, and had some experience in clinical testing of anesthetics related to dentistry. Dr. Fiasconaro was a dentist on the same faculty who worked with Dr. Sherman in some of Dr. Sherman's published works in that field. Drs. Sherman and Fiasconaro conducted the experiments. Dr. Grundfest, a respected Professor of neurology at Columbia's College of Physicians and Surgeons, provided neurological assistance to the team (Elvers, Tr. 10761–62). When first approached by Bristol-Myers in 1957, the experience and published work in the area of drug testing of the Sherman-Fiasconaro team was limited to studying the pain-threshold effects of local dental anesthetics by electrical shock method (Elvers, Tr. 10761, 10763).

536. At that time, their methodology using electrical stimulation of dental tooth pulp was incapable of evaluating the performance of mild oral analgesics, which Dr. Elvers admitted was a "far more challenging objective" (Elvers, Tr. 10763–64). After spending several years to adapt their methodology and equipment for use in evaluating mild analgesics, Drs. Sherman and Fiasconaro conducted the study beginning in 1962 without Dr. Grundfest's participation (Elvers, Tr. 10763–64, 10777). Before the testing of subjects began in the study, Dr. Elvers (Bristol-Myers' then Associate Medical Director) informed the investigators that their study might be used to support advertising claims (CX 445A, B).

537. The test subjects in the Sherman Study were dental out-patients and were tested on a single treatment at each test session (Elvers, Tr. 10772–73). At each test session a subject's "baseline" (premedication) pain threshold was determined by measuring the

amount of electrical current necessary to elicit the first detectable sensation of pain (*i.e.*, pain threshold), on the basis of an average of readings taken at five-minute intervals for a period of 20 to 40 minutes before the test drugs were given (Elvers, Tr. 10773).

538. Thereafter the test drug was administered and threshold readings recorded at five-minute intervals for up to 70 minutes (Elvers, Tr. 10775). From these readings a "plateau" period was picked by the investigator as the period of maximum post-treatment elevation of the pain threshold, and an average measure of current flowing at the plateau was recorded (Elvers, Tr. 10818). **[142]**

539. The ability of a test drug to raise the pain threshold was measured in terms of a "percentage elevation of threshold," that is, the percentage increase in electrical current required to reach the post-medication threshold "plateau" over the premedication "baseline" threshold level (Elvers, Tr. 10818).

540. At the conclusion of the study the mean average percentage elevation of pain threshold achieved by a test drug by a subject was calculated (CX 439P), and the average percentage elevation of pain threshold achieved by Excedrin, aspirin and placebo across subjects was determined (CX 439Q).

541. The test drugs used were two tablets of Excedrin, two tablets of 300 mg. aspirin obtained from 4 commercial sources, and placebo (CX 439G). Excedrin and aspirin tablets were left in their commercial form (Elvers, Tr. 10771), except that, after the initial randomization of treatments was completed, unmarked Excedrin (*i.e.*, tablets without the distinctive "E") were substituted for one-third of the scheduled placebo treatments (Elvers, Tr. 10780). Therefore, there were twice as many Excedrin treatments in the study as those for aspirin or placebo (Evans, Tr. 6402; Elvers, Tr. 10814). All treatments were sealed in coded envelopes, and the investigators were instructed to rip open the envelopes and have the subjects swallow the enclosed tablets without anyone looking at the tablets (Elvers, Tr. 10774–76).

542. During a "dry run" of the Sherman Study (without medication), approximately 30% of the initial population was eliminated from further testing because of their reportedly erratic pain threshold readings (CX 439C; Elvers, Tr. 10765–69). The authors of the Sherman Study characterized the dropouts as "placebo reactors" and attributed the absence of placebo effect in their study to the exclusion of placebo reactors (CX 439B-D).

543. The results of the Sherman Study, as reported in CX 439, are as follows: In 65 tests on 14 subjects, Excedrin caused an average elevation of the pain threshold of 15%, with different test subjects' elevations ranging from 2 to 50%. In 48 tests on 15 subjects, the aspirin brands used caused an average elevation of threshold of 2.7%,

with different test subjects' ranging from 0 to 12% (CX 439N). From these results the authors concluded that they were able to "establish clearly a difference in analgesic effectiveness" between Excedrin and aspirin (CX 439D) and that Excedrin is more effective than aspirin "in elevating the threshold to electrical stimulation of the dental tooth pulp" (CX 439L).

544. It is generally agreed among the students of analgesiology that experimental pain studies measuring threshold effects are not reliable for the purpose of determining comparative performance of mild analgesics in the relief of [143] pathological pain or pain in natural state (F. 547–49, *infra*). In the Sherman Study the authors note that "it is widely held (for references see Beecher, 1959, Lasagna, 1964) that laboratory tests are unsuitable for characterizing the relative effectiveness of analgesic agents" (CX 439B). They also noted, in another pain threshold study using the dental pulp electrical shock method published in 1963, that some investigators viewed experimental pain studies as "inaccurate to the point of being hopelessly useless, both as far as offering theoretical insight and as a practical tool for clinical application" (CX 439D; Tr. 10910).

545. In that 1963 article, Drs. Sherman, Fiasconaro and Grundfest compared the threshold effects of codeine and aspirin and concluded that 30 mg. codeine was 3 times more effective than 1800 mg. aspirin (Tr. 10918). This finding is in sharp contrast with the results of clinical pain studies (analgesic bioassays) by Drs. Kantor, Sunshine, Laska, et al., and by Dr. Bloomfield, which suggest that 60 mg. codeine is no more effective than 600 mg. aspirin and possibly too low a dosage to produce reliable analgesia (Elvers, Tr. 10923–24). In CX 439, the authors note that their earlier (1963) study using the same method adopted in CX 439, was contradicted by the available clinical literature (CX 439B). In this connection, the Sherman Study reported the peak effect for aspirin as occurring at 25–30 minutes (CX 439H), in sharp contrast to the generally accepted aspirin peak effect time of one to two hours based on bioassay studies (Beaver, Tr. 5945).

546. Pain induced by electrical shock on tooth pulp is a fast, jabbing type of pain and is unlike most clinical pain, which is described as dull, throbbing, aching and of much longer duration. Electrical stimulation of tooth pulp has proven to be notoriously unreliable even among experimental pain models (Evans, Tr. 6352, 6359, 6373–74). Fast, jabbing pain involves different physiological mechanisms than clinical pain (Evans, Tr. 6349, 6373–74). Other experimental methods which more closely approximate clinical pain have been shown responsive to standard analgesics such as morphine. With all of their shortcomings, they are more appropriate analogs for clinical pain in

the laboratory than the Sherman model (Evans, Tr. 6331, 6338, 6352, 6369, 6373–74).

547. Dr. Beecher, whom Dr. Elvers regards as "the leading man in the field and sort of the father of experimental research and clinical research as well" (Elvers, Tr. 10801), concluded in his treatise, *Measurement of Subjective Responses* (1959), that although some workers believe it satisfactory, "in view of the remarkable inconclusiveness of the method of electrical shocks to teeth in man . . . it is difficult to accept work that depends upon this method and technique" (Elvers, Tr. 11111). As late as 1978, Wolff, whom Dr. Elvers referred to as "definitely a leader in experimental pain research" (Elvers, Tr. 10800), was [144] still attempting to develop a methodology which would achieve reliable results with electrical stimulation of dental tooth pulp (Elvers, Tr. 11084–88).

548. It has been suggested that the type of pain elicited by electrical stimulation of dental tooth pulp might be unique to itself (Elvers, Tr. 11166). Dr. Mumford, a respected researcher, compared subject reactions to real toothache and pain induced by electrical stimulation of tooth pulp, and concluded that "qualitative assessment" of the real toothache "differed considerably" from the pain induced by electrical stimulation (Elvers, Tr. 11163–64).

549. In their 1963 article, Sherman, *et al.*, recognized that experimental models employing transient ("fast") pain, and those employing dull, throbbing prolonged ("slow") pain, produced "qualitative[ly]" different kinds of pain, which involved different "pain reporting pathways" in the body. They therefore cautioned that analgesics found efficacious using their "fast" pain model "may be more or less so for painful sensations elicited by other pathways" (Elvers, Tr. 11156–57).

550. In CX 450, an earlier draft of the Sherman Study (CX 439), the authors stated that "aspirin might conceivably be more effective [than Excedrin] in relieving other types of pain" than that induced by electrical stimulation of dental tooth pulp (CX 450G). Dr. Elvers, then Associate Medical Director of the Bristol-Myers Products Division, instructed the authors to remove this statement from the report as "gratuitous speculation" (CX 449D; Elvers, Tr. 11159). Nevertheless, the authors still state in CX 439 that their results may be limited to pain involving "pain reporting pathways" similar to those involving electrically stimulated tooth pulp pain (CX 439L), clearly indicating that they recognized the doubtful generalizability of results using a transient pain model (Evans, Tr. 6409–10).

551. In any event, it is highly doubtful whether a study based on pain threshold performance of an analgesic agent can provide any meaningful conclusions about pain reduction. Certainly the Sherman

Study did not (Evans, Tr. 6368). The pain threshold is a transient, momentary point in the pain experience and is not a relevant point in the measurement of clinical pain (Evans, Tr. 6472–73).

552. On the other hand, measurement of the suprathreshold point at which pain is intolerable (tolerance level) has been shown to reliably respond to standard test analgesics such as morphine, and more closely correlates with the type of pain patients report in the clinic (Evans, Tr. 6382–6385). Wolff, in a paper co-authored with Dr. Thomas Kantor and Dr. Eugene Laska, noted in 1969 that "[l]ogically, pain tolerance, being **[145]** suprathreshold pain, would seem a better index of analgesic efficacy than pain threshold . . ." (Elvers, Tr. 11127).

553. The Sherman Study also failed to employ the appropriate scientific procedure, the so-called "method of limits," in determining pain threshold. The "method of limits" averages the measurement of the *ascending* threshold (the point where a pain stimulus, increasing from sub-threshold intensity, is first detected as painful) and the *descending* pain threshold (the point where a pain stimulus, decreasing from supra-threshold intensity, is last detected as painful) in order to correct for the tendency of test subjects to under- and over-shoot actual pain threshold (Evans, Tr. 6377; Elvers, Tr. 11140). Wolff, a highly reputable investigator (Elvers, Tr. 10800), measures both ascending and descending pain thresholds and pain tolerance in studies using electrically induced dental tooth pulp pain (Elvers, Tr. 11145).

554. Sherman's elimination of 30% of his original subject sample because of reportedly erratic threshold readings (F. 542, *supra*) was a totally unacceptable scientific procedure (Evans, Tr. 6395). Since Sherman never gathered data on these subjects, there is no way of knowing the effect their inclusion might have had on the results of the study (Evans, Tr. 6395), nor the representativeness of the remaining sample. Beecher suggested that elimination of persons with erratic pain thresholds might leave a sample representative only of itself (Elvers, Tr. 11199). Sherman's inference that those subjects eliminated from the study were placebo reactors (CX 439B-C) was an untested assumption (Evans, Tr. 6393–94), and there is no basis for believing it correct (Evans, Tr. 6393–95). Dr. Laska expressed a similar conclusion (Laska, Tr. 10493–94). Sherman also recognized that the attempted elimination of placebo reactors "raises the possibility of 'tampering' with the data" (CX 439C). One researcher, specifically addressing the Sherman Study, suggested in a published article that those eliminated from the Sherman Study as inferred placebo reactors would actually have had *lowered* thresholds with the aspirin, putting the study's methodology in serious question, in light of aspirin's known effectiveness (Elvers, Tr. 11191–92). *Also see* FDA OTC Analgesic Panel Report, CX 154, p. 35444.

555. According to Dr. Evans, the zero response rate for placebo reported in the Sherman Study may indicate a breakdown in double-blinding, since the placebo response rate is known to be always above zero in well-blinded studies (Evans, Tr. 6391). According to the Sherman data, placebo actually began to lower the threshold at precisely the time (25 minutes) when other compounds were shown to elevate it (CX 439K, S). The obvious explanation for *lowering* of pain thresholds after administration of placebo is that the subjects were aware of the identity of the test drugs, and when a placebo is given them, responded more sensitively to pain (Evans, Tr. 6406). The fact that Excedrin **[146]** and aspirin were left in their commercial form also increases the possibility that the subjects may not have been successfully blinded.

556. The raw data for CX 439 is replete with calculation errors (Evans, Tr. 6398, 6402–03). Dr. Elvers agreed that about one-half of Sherman's calculations of percentage elevation of threshold had to be corrected for reanalyses (Elvers, Tr. 11260). Moreover, prior to calculation of baseline thresholds, Dr. Sherman discarded certain readings on the raw data sheets without explanation (Evans, Tr. 6401).

557. Furthermore, the method by which Dr. Sherman selected data points which he believed represented the "plateau" of post-medication elevation was never explained, varied from session to session, and followed no discernible standard or rule (Evans, Tr. 6401–02). Bristol-Myers attempted to address this problem in its reanalyses of underlying data by the so-called "geometric mean peak ratio" technique (Elvers, Tr. 10821; Mueller, Tr. 10092). However, as Dr. Elvers admitted, the "geometric mean peak ratio" technique cannot distinguish between aberrant peak values and true threshold elevations and is not found in the literature (Elvers, Tr. 11237). At any rate, Bristol-Myers' reanalysis of the Sherman data disclosed the inability of the study to differentiate aspirin from placebo at the .05 level of significance (RX 212A, 213A; Elvers, Tr. 11256).

558. The credibility of the Sherman study is placed in further doubt by the extraordinarily high amounts of electric current recorded as flowing through subjects at the point where pain threshold was reached. According to the data, eight (8) of the fifteen (15) test subjects required amounts of electricity as high as 800, 480, 117.5, 111, 82, 78, 57, and 56.5 microamps before reaching threshold pain (CX 886(a)). The pain thresholds for dental tooth pulp in healthy teeth, as reported in the literature, are normally reached at currents of 1.2 to 26 microamps (Elvers, Tr. 11212–13). Dr. Elvers' opinion offered as possible explanations for these abnormally high readings were largely based on speculation (Elvers, Tr. 11217–94).

559. The record shows that Bristol-Myers' subsequent attempt to

replicate the Sherman Study (CX 439) was unsuccessful. Bristol-Myers employed Dr. Ozick, now associated with New York University (Elvers, Tr. 10900–01), for this purpose. According to Dr. Elvers, the study undertaken by Dr. Ozick for Bristol-Myers was "initially comparable" to the Sherman Study (Elvers, Tr. 10897). Dr. Ozick was unsuccessful in reproducing Dr. Sherman's work using Sherman's methodology (Elvers, Tr. 10898–99), and eventually modified Sherman's procedures and equipment "in the hope of replicating the Sherman type of study" (Elvers, Tr. 10899). Even after these modifications by Ozick, the methodology and equipment were not capable of producing "the stability [Bristol-Myers] felt necessary for the study of [147] analgesics," (Elvers, Tr. 12393). The Ozick Study was abandoned. However, Dr. Elvers testified that Ozick "never set out to replicate the study" (Elvers, Tr. 10900), but was merely trying to develop the method, equipment and competence "that would permit him to attempt a replication of the Sherman Study . . . [and] in that attempt he failed" (Elvers, Tr. 10900).

560. From the foregoing discussion of the Sherman Study (CX 439), it is found that CX 439 may have some limited application to dental pain threshold elevation, but it is unreliable for the purpose of comparing the effectiveness of aspirin and Excedrin in any other pathological pain in the natural state.

D. *It Has Not Been Scientifically Established That Speed Of Relief Provided By Bufferin Is Significantly Greater Than That Provided By Plain Aspirin*

1. Claims of Faster Relief and Twice as Fast Relief

561. As Bristol-Myers argued to the Federal Trade Commission in its Comments on a Proposed Trade Regulation Rule on OTC Analgesics filed February 6, 1968, if "one wants to claim that [an] analgesic acts faster on tension headache than some other preparation, *one should be required to prove that it acts faster, i.e., by interviewing people under the proper conditions and finding out how soon the headache goes away"* (F. 375, *supra;* emphasis added).

562. In this proceeding, instead of presenting *studies* done on headache, Bristol-Myers relied on an argument based on analogy by its Medical Director to suggest that Bufferin's onset of analgesic activity occurs sooner than plain aspirin's (Lanman, Tr. 11619–59). Bristol-Myers' argument in this regard is twofold: (1) that Bufferin is absorbed more rapidly than aspirin into the bloodstream, and (2) that, therefore, Bufferin will start to relieve pain sooner than plain aspirin (Lanman, Tr. 11635, 11658–59). In support of this argument, a number of "blood level" studies were offered and received. These studies

report that Bufferin produces somewhat higher blood levels of hydrolized and unhydrolized aspirin than plain aspirin (Lanman, Tr. 11635–58).

563. However, Bristol-Myers' Medical Director, Dr. Lanman, once expressed the same opinion offered by every independent expert who addressed the issue in this proceeding and every expert Panel and publication that has considered it (F. 592–601, *infra*). In an April 1969 memorandum, Dr. Lanman stated:

> It is quite true that aspirin is absorbed more readily from Bufferin than from ordinary aspirin tablets. Unfortunately, it is a much [148] more difficult thing to correlate clinical relief with Bufferin. In fact, we have no such correlation between clinical and laboratory tests and the explanation is a very complex one. (CX 508)

564. In fact, as CX 508 states, no correlation between blood levels of aspirin and onset, degree or duration of analgesia has been demonstrated (F. 401, *supra*). Four of complaint counsel's expert witnesses were examined and cross-examined on this issue, and each of them consistently held to the view that well-controlled clinical investigation is the prerequisite in order to establish that one analgesic compound relieves pain faster than another (Azarnoff, Tr. 9195, 9225; Forrest, Tr. 8980, 8987–90, 9035, 9043–45; Moertel, Tr. 5800–06, 5817–18, 5860; Beaver, Tr. 5947–48, 5951–52, 5957–58, 5961–64). In defense, respondent offered the testimony of not one independent clinical pharmacologist who supported its position. Only Dr. Lanman, an employee of Bristol-Myers for 19 years, was offered to present that position, and Dr. Lanman's opinion testimony concerning Bufferin's superiority in this proceeding is not consistent with his view submitted to the FTC in 1969 (CX 508) (F. 563, *supra*).

565. The proposition that Bufferin provides higher blood (serum concentration) levels of ASA, SA and TSA sooner than plain aspirin is supported by a preponderance of credible evidence. Complaint counsel have admitted that studies and tests submitted by Bristol-Myers to the FTC reported that Bufferin is absorbed into the bloodstream faster than aspirin (BMPF 60–107), and that the blood salicylate level of Bufferin 10 minutes after ingestion and 20 minutes after ingestion is in both instances twice as high as that of aspirin (BMF 114–144).

566. The Stough Study that measured the total salicylate in the blood at 0, 10, 20, 40, 120, 240 and 300 minutes after ingestion of aspirin, Bufferin, Anacin and Bayer aspirin (Tr. 11633–34; CX 506Z405), shows that with incremental doses of aspirin there is an incremental increase in blood level. Bufferin provided more aspirin into the bloodstream at 10 minutes than 10 gr. Bayer aspirin, 13 gr. Anacin or 10 gr. plain aspirin and provided more total salicylate at

20 minutes than 10 gr. Bayer, 13 gr. Anacin and 20 gr. aspirin (CX 506Z413).

567. BMRX 157, a graph of the results of the Stough Study, depicts the wide difference in blood level between the administration of 10 and 20 grains of aspirin (BMRX 157; Tr. 11633).

568. Both the Paul Study (CX 786) and the article published by Dr. Sleight in the *Lancet* (a British medical journal) (CX 787) have shown that the level of aspirin produced in the blood by Bufferin is twice that produced by plain aspirin (Tr. 11635–[149] 36). Paul reported that the Bufferin formula resulted "in at least a two-fold increase in the blood salicylate levels. The ten-minute salicylate level following [Bufferin] exceeds the twenty-minute salicylate level for ordinary aspirin by more than 20%. Furthermore, the salicylate level twenty minutes after ingestion of [Bufferin] is almost 2–1/2 times the twenty minute ordinary aspirin level." (CX 786D).

569. CX 550, the Stanford Research Institute Study entitled "Clinical and Statistical Studies of Blood Salicylate Levels" was a triple crossover design studying St. Joseph aspirin, Bayer aspirin and Bufferin, all purchased on the open market, through analysis of blood samples taken 0, 10, 20, 45, 90, 125 and 150 minutes after ingestion (CX 550B; CX 550J).

570. CX 550 found that after 10, 20 and 45 minutes, the subjects given Bufferin showed significantly higher salicylic acid concentrations in the blood than those given either St. Joseph or Bayer aspirin (CX 550J). For example, after 20 minutes the concentrations of total salicylic acid in the blood of subjects given Bufferin were from 68 to 100% higher than those given Bayer or St. Joseph aspirin (CX 550J).

571. The results of CX 550 corroborate the results of blood level studies conducted in Bristol-Myers' own research and development laboratories (BMF 61–107, 114–133, 134–144) which show that Bufferin is absorbed more quickly (from 50–100% more salicylic acid within the first 10 minutes and approximately twice as much after 20 minutes) than plain aspirin (CX 550K).

572. A second study by Stanford Research Institute entitled, "A Clinical and Statistical Study of Blood Salicylate Levels Following The Ingestion of Two Preparations Containing Aspirin" (Tr. 11640–41; CX 506Z174–Z177, Z405–414) was a double-blind randomized comparison of Bufferin and Bayer aspirin in which blood samples, drawn at 0, 10, 20 and 40 minutes after ingestion, were analyzed for salicylic acid (CX 506Z176). At all time periods Bufferin was found to have statistically significantly higher blood levels than Bayer (59–64% higher on the average), results which were consistent with the earlier (CX 550) Stanford Research Study (CX 506Z176).

573. In 1958 Dr. Paul, and during the period 1959 through 1968 Drs.

Paul and Routh compared serum levels of Bayer with samples of Bufferin and found that all of the Bufferin samples showed numerical superiority of Bufferin to Bayer at 10 and 20 minutes after ingestion. The numerical superiority of Bufferin to Bayer found in the 36 studies were significant to p=.05 or less in 34 of the 36 studies with 31 having a p value of .01 (CX 506N). And in 1958 Dr. Cronk found that the addition of Di-Alminate® caused a similar enhancement in absorption of salicylate (CX 506M). **[150]**

574. Dr. Heimer at Seton Hall College of Medicine and Dentistry, found that Bufferin's higher total salicylate levels (TSA) were statistically significantly higher than those for Bayer up to 120 minutes, Bufferin's free salicylate levels (FSA) were statistically significantly higher than Bayer up to 45 minutes, and the differences between TSA and free salicylate (FSA) was significantly higher for Bufferin at 10, 20 and 30 minutes (CX 506Q; BMRX 136; Tr. 11646–47; BMRX 177D; Tr. 11657).

575. Komoda found in 1965 that Bufferin gave significantly higher TSA levels than Bayer at 10, 20 and 30 minutes (CX 506Q).

576. In 1962, 1963, 1965, 1968, 1969, through measurement of TSA, FSA and ASA, Bufferin was found to have been absorbed faster than Bayer (CX 506Q-R).

577. Truitt and Morgan found the plasma salicylate concentration for Bufferin "approximately twice as high" as for Bayer at 10, 15, 20 and 30 minutes with the differences being highly significant at p=.001 (CX 506R).

578. In 1958, Paul and Ruth reported in a study of 1 and 20 minute blood levels for 1, 2 and 3 tablet (5, 10 and 15 grains) doses of Bufferin, Anacin and Bayer that (1) blood levels increased with increasing dosage and (2) Bufferin levels were far superior at each dosage (CX 506S).

579. At 10 minutes the ASA level of plain aspirin is .5 mg/ml compared to 3 mg/ml for Bufferin. At 20 minutes, the comparison is 1 mg/ml for plain aspirin and 3.5 mg/ml for Bufferin. Both of those differences are statistically significantly in favor of Bufferin (Tr. 11649; BMRX 136A; CX 506R; footnote 67).

580. The ASA blood levels of Bufferin are significantly higher than those for aspirin at 20 minutes (BMRX 136C; CX 506R, footnote 65; Tr. 11651–52; 11652–53).

581. At both 10 and 20 minutes after ingestion, the ASA blood levels of a 10 grain dose of Bufferin are significantly superior to those for a 10 grain dose of aspirin (Tr. 11652–53; BMRX 136D; CX 506R, footnote 66).

582. There is a twofold or larger increase in absorption rate for TSA

comparing Bayer and Bufferin aspirin (Tr. 11654–55; CX 506M; BMRX 177B; CX 523).

583. BMRX 177C, a graph based upon a study by Morgan and Truitt, published in Vol. 54, No. 11 of the *Journal of Pharmaceutical Sciences,* pp. 1640–46 (Nov. 1965) entitled "Evaluation of Acetylsalicylic Acid Esterase in Aspirin Metabolism, Interspecies Comparison" (CX 521A-H; CX 506R; footnote 72) shows that the observed TSA concentrations of [**151**] Bufferin are higher than those for Bayer aspirin (Tr. 11655–56; BMRX 177C): "[T]he aspirin blood levels of [Bufferin] exceed those of a plain aspirin at all of the time periods test[ed] i.e., 10 20, 30, 45, 90 and 240 minutes***. These higher ASA blood levels were comparable with previously reported plasma salicylate levels for [Bufferin]." (CX 521G).

584. Dr. Beaver testified that it is unknown whether the unhydrolized aspirin (ASA) or the salicylate (SA) or some combination of the two is, when in solution in the blood, responsible for analgesic activity (Tr. 5942–53). However, there are some studies (one by Dr. Lasagna and one by Dr. Houde) (Tr. 5977) that indicate that aspirin (ASA) is about 1.5 times as potent an analgesic as an equivalent amount of salicylate (SA) (Tr. 5976–77).

585. Dr. Azarnoff testified that both ASA and SA are active principles, that they have different potencies (Tr. 9108) and that ASA is the more active (Tr. 9193). And Dr. Forrest testified that the state of the art is that the active metabolite in aspirin is ASA (Tr. 9025–27).

586. Dr. Levy, one of the foremost experts in pharmacokinetics wrote, in an article entitled "Aspirin: Absorption Rate and Analgesic Effect," published in *Anesthesia and Analgesia,* November - December 1965:

> There is considerable evidence that aspirin (ASA) is a more effective analgesic than salicylic acid (SA), both in man and in animals. Aspirin in the body is hydrolized rapidly to salicylic acid, and it has been found that oral administration of this drug in rapidly absorbable form (aspirin solution) results in higher and earlier maximum blood levels of unhydrolyzed aspirin than are obtained after administration of aspirin in a more slowly absorbed form (compressed tablets). (Tr. 1161–62).

586a. Dr. Beaver testified that unhydrolized aspirin (ASA) peaks before one-half hour after ingestion and is rapidly eliminated or biotransformed into salicylate or some combination of ASA and SA (Tr. 5946).

587. Total salicylate (TSA) can be measured by measuring either the sum of unhydrolized acetylsalicylic acid (ASA) plus salicylic acid (SA) or by allowing all ASA to hydrolyze and measuring it as SA. The hydrolysis of ASA can be inhibited—to allow measurement of ASA, TSA or SA (Tr. 9236–38).

588. Whether ASA or SA is the active or more active moiety that produces analgesic effect, the studies cited hereinabove [**152**] show that Bufferin produces higher blood levels of them sooner than plain aspirin. "There is clear experimental evidence based upon well designed blood level studies which substantiate the claim that buffered aspirin is more rapidly absorbed than plain aspirin (Refs. 1–3 [citing Bristol-Myers blood level data. *See* BMRX 234; CX 514, p. 35481]). Comparisons of the most commonly used plain and buffered aspirin show that salicylate blood levels are twice as high in the first ten to twenty minutes for the buffered aspirin product compared to regular aspirin. It can be shown that the differences in plasma levels in the first twenty minutes correlate quite well with the amount of drug absorbed (Ref. 4)."

589. Dr. Levy in his article (F. 586, *supra*) stated:

> Differences in gastrointestinal absorption rate have a pronounced effect on the magnitude and time of occurrence of maximum drug levels in the body in the case of drugs (such as aspirin) which are rapidly metabolized and/or excreted. Consequently, absorption rate can affect the onset, intensity, and duration of pharmocologic effects *if* the latter are related to the magnitude of drug levels in the body. Since the absorption rate of drugs administered in tablets can be modified appreciably by the pharmaceutic properties of the tablets, differences in tablet formulation may modify markedly the pharmacologic effect of many drugs. (Tr. 11686–87; emphasis added).

590. It is generally agreed among clinical pharmacologists that the limiting factor governing aspirin's absorption rate is the dissolution rate of the dosage form (tablet) and that the method of formulation can significantly affect a tablet's dissolution rate apart from buffering.

591. The FDA Analgesic Panel corroborates that view and recommended a standard dissolution test procedure for buffered analgesic products.

> From the available data, the Panel finds that simply adding buffering agents to aspirin does not generate an increased dissolution rate over unbuffered aspirin. Important factors appear to be the type of buffering agent used and other undefined factors, *e.g.*, tablet compression during manufacturing, etc. . . . For this reason, actual testing of the dissolution rate of buffered aspirin products is necessary to determine if the buffering agent actually [**153**] does affect the dissolution rate of the aspirin products and to what extent.
>
> Also, the Panel notes that an adequately buffered aspirin product may not have an advantage over a well-formulated unbuffered product. In some studies, unbuffered aspirin performs as well as buffered aspirin products (CX 514, p. 35375; *also see* pp. 35469–70).

592. While Dr. Forrest agreed with the FDA Analgesic Panel that "The basic problem is that there are no well-controlled clinical studies

that unequivocally prove or disprove that these differences in absorption will result in clinically important differences in the onset, intensity or incidence of relief of pain or fever," (Tr. 9024–25; CX 514, p. 35480) he testified that the extrapolation of blood levels to drug's anticipated effect is a "very rational one" and is used in other field where there is "objective measures of what is happening." And "the big problem for us here is the subjective nature of this whole problem of pain and pain relief."

593. Dr. Forrest agreed with the Panel's statement that:

> If the blood level time curves were superimposable, it would be reasonable, based on all known studies, to assume that the formulations would have equal onset, duration and intensity of pharmacological effects. However, if one product were substantially more rapidly absorbed than the other, one cannot conclude that there is necessarily a corresponding difference in onset of effect. The mathematical relationship between changes in blood levels and corresponding changes in onset, or intensity of analgesia response is not presently known for aspirin. (Tr. 9025–27, 9028; CX 514, p. 35373).

594. Dr. Forrest also testified that the blood level curves for aspirin and Bufferin could not be superimposed without moving the baseline onset point (Tr. 9034–35). He further testified that Bufferin's more rapid early absorption could make the onset of pain relief later or earlier, but that the hypothesis of Bufferin's earlier onset is interesting and possibly correct (Tr. 9036).

595. Dr. Beaver did not claim that there was no relationship between Bufferin's higher blood levels and increased clinical pain relief but only that blood level does not correlate "nicely" or that the correlation is not "simple" or "direct" or that blood levels "may not in any tidy way mirror" clinical effect (Tr. 5952). [154]

596. The FDA OTC Analgesic Panel's conclusion with respect to drug blood levels corroborates the expert opinions reviewed above. The Panel states:

> Aspirin is commonly used as a standard analgesic drug for comparison with other drugs in which assays of blood levels are made rather than direct measurements of the analgesic effectiveness of these agents. The Panel has evaluated this technique and concludes that there is inadequate evidence that the amount of drug in the blood correlates directly with clinical analgesia. The Panel emphasizes that this is not to say that a relationship between blood levels and clinical response does not exist, but rather that the relationship is complex and not presently understood. However, the Panel does recognize that an important value of drug blood level determinations is that they do give an indication of comparative dissolution rates. . . .
>
> The Panel recognizes that the drug labeling related to the onset, intensity and duration of pharmacologic effects can influence the consumer's selection of a product that can find no convincing evidence to support labeling claims which suggest a faster onset of effectiveness . . . There is also no direct evidence available to the Panel which suggests a greater intensity of analgesia for comparable products . . .

. . . [S]ome buffered aspirins are somewhat more rapidly absorbed from the gastrointestinal tract than unbuffered aspirin and might also be expected to show earlier higher salicylate blood levels. However, the Panel is unaware of any data that demonstrate that buffered aspirin provides a more rapid onset, a greater peak intensity or a more prolonged duration of analgesic effectiveness than unbuffered aspirin. (CX 514 at 35378).

597. The FDA Panel placed Bufferin's "faster" claims in Category III. The Panel reached these conclusions after reviewing voluminous submissions from Bristol-Myers, which included the same materials and arguments Bristol-Myers has raised in this proceeding (CX 506; Tr. 12115–16, 11443–45, 11469–70, 11630–31, 11640, 11644–47, 11649, 11651–58). **[155]**

598. The AMA Drug Evaluations (2d ed. 1973) also corroborates those views:

. . . It has been suggested that the analgesic effect of aspirin is related to blood levels of acetylsalicylate rather than salicylate; however, it has not been possible to correlate these blood levels with the degree of analgesia in man. (CX 512, p. 261.)

599. Dr. Beaver wrote to AMA's Dr. Lewis in connection with the AMA drug evaluations, "Bufferin does have a somewhat higher dissolution and absorption rate than plain aspirin, but results of controlled studies have not conclusively demonstrated that the use of these mixtures results in fact to onset of greater or longer analgesic effect or less gastric upset than plain aspirin." (Tr. 4239).

600. Dr. Lewis testified that, although there is some correlation between blood levels and analgesia in some situations, studies have not conclusively demonstrated that Bufferin has faster onset, greater or longer action or less stomach upset (Tr. 4254–56).

601. The *Medical Letter's* July 5, 1974 issued entitled "Is All Aspirin Alike?" provides further corroboration of the above views. Regarding buffered aspirin tablets, it states in part:

***It has never been established in patients with painful conditions . . . that there is a difference between buffered and unbuffered aspirin in time of onset of analgesia, duration or degree of relief of pain, or incidence of gastrointestinal distress. (CX 510A-B).

602. The FDA Analgesic Panel seems to be using the word "correlation" in terms of a mathematical, that is statistical, relationship between blood level and analgesic effect (Tr. 9038). The Panel states:

While current studies have failed to show a direct one-to-one correlation between plasma levels of an analgesic drug and pharmacologic response, there is some evidence that a complex nonlinear relationship between these two variables undoubtedly does exist and involves nonlinear complex functions and time lags. . . . There are known

relationships between dose and plasma concentration (also nonlinear). It follows logically and mathe[156]matically that some expression does exist and recent advances in computer assisted pharmacokinetic modeling, analytical methodology and analgesic testing will probably allow elucidation of this function in the future. When an insensitive test does not show clear differences between two products it can only be said that present insensitive methods cannot determine a difference between the two. In the absence of other evidence, no means of validating claims are available. (CX 514, pp. 35480–81).

603. Dr. Beaver testified that due to technical difficulties there have been unsuccessful attempts to correlate actual clinical effect with blood salicylate levels by simultaneous collation of blood samples and measurements of analgesia (Tr. 5957).

604. Blood level studies are quite sensitive and can pick up small differences in blood level (Tr. 9053–54).

605. In order for there to be pharmacological action, the active principles of the drug must reach the site of action and in order to do so, must first get into the blood (Tr. 9038–39).

606. Bufferin puts unhydrolyzed aspirin and salicylate at given levels into the bloodstream faster than aspirin and it is reasonable to suspect and unreasonable to preclude the possibility that the dosage form that got into the bloodstream sooner (Bufferin) will produce clinical effect sooner (Beaver, Tr. 5955), since once into the blood, there is no pharmacological or physiological difference between plain aspirin and the aspirin from Bufferin (Beaver, Tr. 6063).

607. Dr. Azarnoff does not doubt that before pain relief can occur, sufficient quantities of pain reliever (aspirin, the active principle in Bufferin or Excedrin) must reach the receptor site in the sufficient amount to trigger the onset of pain relief (Tr. 9203–04) and that in order for the active principle of pain reliever to teach the receptor site it must get into the blood and reach the receptor site via the blood stream (Tr. 9204).

608. A later appearance of active principles in the blood stream suggests, but does not necessarily prove, later onset of pain relief (Azarnoff, Tr. 9205). Similarly, earlier appearance of active principles in the blood stream suggest, but does not necessarily prove, earlier onset of pain relief.

609. The NAS/NRC Panel agreed that the Bristol-Myers submission and the published literature made a very good case that Bufferin is absorbed to some degree more rapidly than plain [157] aspirin tablets (Tr. 5947–48). Dr. Beaver, a member of the Panel, refused to accept Bufferin's claims of faster relief because of a lack of substantial evidence, by which he meant clinical evidence from controlled analgesic studies (Tr. 6043).

610. Dr. Moertel, a clinical pharmacologist, indicated that absorp-

tion, excretion, metabolism, and various other factors all play a role in the onset of pain relief vis-a-vis blood levels, and that for this reason Bristol-Myers' argument cannot be accepted as a substitute for the ultimate test of clinical trial (Moertel, Tr. 5803–05). Dr. Azarnoff, the only expert in pharmacokinetics who testified in this proceeding, stated that *no* conclusions can be drawn from blood level studies regarding a buffered product's speed in relieving pain (Azarnoff, Tr. 9195). If one desires to show faster pain relief, one would have to conduct a therapeutic trial (*i.e.*, a clinical study) of the drugs in question (Azarnoff, Tr. 9195). Finally, Dr. Beaver, who was a member of the NAS/NRC Panel that evaluated certain faster onset claims for Bufferin, testified that *nothing* that has developed in the literature—over the course of time from his review article in 1965 and the NAS/NRC Panel's review in 1967 up until present—has changed his view that Bufferin's faster onset of relief claims lack substantial evidence (Beaver, Tr. 6042).

611. The FDA's regulations concerning the bioavailability and bioequivalence of prescription drugs (*Bioavailability and Bioequivalence-Requirement,* 42 FR 1624, *codified as* 21 C.F.R. 320), do not support respondent's contention that Bufferin is therapeutically superior to plain aspirin.

612. The purposes of the FDA bioequivalence regulations are (a) to identify pharmaceutically equivalent drugs, or pharmaceutical alternatives, "that are intended to be used interchangeably for the same therapeutic effect and that are not bioequivalent drug products"; and (b) to establish a "bioequivalence requirement for these drug products" (21 C.F.R. 320.50). Thus, pharmaceutically equivalent drugs (*i.e.*, drug products that contain identical amounts of identical active ingredients, *see* 21 C.F.R. 320.1(c)) become a concern under the regulations only if they are *not* "bioequivalent drug products."

613. For purposes of the FDA bioequivalence regulations, Bufferin and any well-formulated aspirin are not only pharmaceutical equivalents, but also bioequivalent drug products. The regulations define "bioequivalent drug products" as pharmaceutical equivalents (or alternatives) "whose rate and extent of absorption [*i.e.*, bioavailability] do not show a significant difference" when administered at prescribed dosages. The regulations further note that: **[158]**

> [s]ome pharmaceutical equivalents or pharmaceutical alternatives may be equivalent in the *extent* of their absorption but not in their rate of absorption and yet may be considered bioequivalent because such differences in the *rate* of absorption . . . are considered medically insignificant for the particular drug studied. 21 C.F.R. 320.1(e) (emphasis added).

614. Differences in *rate* of absorption become "medically signifi-

cant" under the FDA regulations (and are therefore viewed as "bioequivalence problems") *only* if they "would result in therapeutic failure or a hazard to the patient" (42 FR at 1626). Only where such "medically significant bioequivalence problems" exist, will pharmaceutical equivalents (such as Bufferin and aspirin) be found "not bioequivalent," for purposes of the FDA regulations (*see* generally, *Criteria and evidence to establish a bioequivalent requirement,* 21 C.F.R. 320.52). In this proceeding, however, there is no suggestion that because of the difference in rate of absorption of Bufferin and correctly formulated plain aspirin, Bufferin may cause "therapeutic failure or a hazard to the patient."

615. The specific subpart of the FDA bioavailability regulations regarding intra- and inter-batch variation in bioavailability of a single drug product (21 C.F.R. 320.21(f)(2)), cited by Dr. Lanman in his testimony (Lanman, Tr. 11663–71), are irrelevant to the issue of Bufferin's alleged therapeutic superiority to aspirin. Reference to batch variability in both the bioavailability and the bioequivalence regulations is clearly concerned with assuring the adequacy of manufacture and quality control in drug production. For drugs that are *already subject* to a bioequivalence requirement, the bioequivalence regulations state that:

> the ability of a manufacturer to make a satisfactory product consistently in four batches will generally assure FDA that the methods of manufacture and quality control are adequate . . . (21 C.F.R. 320.55).

Under the separate bioavailability section of the regulations, FDA will insist on further reassurance of the adequacy of methods of manufacture and quality control in the form of new bioavailability studies, where

> there are data demonstrating significant intra-batch and batch to batch variability, *e.g.* plus or minus 25 percent, in the bioavailability of *the* drug product (21 C.F.R. 320.2(f)(2)). **[159]**

There is no mention in this subsection that the 25% variability standard is meant to apply to comparisons between *different* products. Clearly, the "plus or minus 25 percent" reference in this subpart of the bioavailability regulations is a guideline for monitoring lapses in manufacturing and quality control of a drug product, not a standard for determining therapeutic equivalence or nonequivalence. Therefore, this clearly allows no inference, as suggested by Dr. Lanman (Tr. 11671), that a drug manufacturer which deliberately "varies" the bioavailability of its product by "plus 25 percent" is in any way superior to other members of its product class.

616. The FDA regulations themselves make clear that the bioavailability of a drug and its effectiveness are separate and distinct issues:

> It is not . . . the intent of a bioavailability study to demonstrate effectiveness. The purpose of a bioavailability study is to determine the rate and extent of absorption. If a drug product is not bioavailable, it cannot be regarded as effective. However, a determination that a drug product is bioavailable is not in itself a determination of effectiveness. The requirement of evidence of bioavailability is intended to supplement, no[t] replace, clinical evidence of effectiveness. 42 FR at 1640.

In fact, the FDA anticipated arguments such as what Bristol-Myers has advocated in this proceeding and specifically warned that:

> The bioequivalence regulations are not an attempt to equate evidence of bioequivalence with evidence of relative therapeutic effectiveness . . . 42 FR at 1625.

617. Dr. Lanman further suggested that the FDA's willingness to accept *in vitro* testing or forms of *in vivo* testing other than clinical testing (in connection with the bioavailability and bioequivalence regulations) in some way indicated FDA's willingness to accept non-clinical, *in vitro* tests where the effectiveness of a class of drugs (*e.g.*, aspirin-based OTC drugs) has been demonstrated (Lanman, Tr. 11672 –76). However, FDA's statement cited by Dr. Lanman relates to determination of bioavailability or bioequivalence, not comparative effectiveness. 42 FR at 1639, 1640). Since clinical tests are not designed to measure the rate and extent of drug absorption, FDA prefers that a more direct "accurate sensitive [and] reproducible" means of measurement be used where the issue relates to bioavailability, and not to the clinical effects of drugs on patients. 42 FR at 1640. The FDA stated in [160] requiring bioavailability data in New Drug Applications (in addition to evidence of effectiveness through clinical trials) that such data is "needed to assure that the dosage formulation intended for marketing has the same characteristics as the dosage formulation used in clinical trials to determine safety and effectiveness and that there is batch to batch consistency." 42 FR at 1639. Thus, clinical tests and bioavailability tests perform different, though complementary functions. Preference for verification of bioavailability, using *in vitro* measures of bioavailability, in no way suggests any relaxation of FDA's clear requirements that issues of safety and efficacy be determined in clinical trials (F. 516, *supra*).

618. It is concluded that the sole purpose of the FDA's bioavailability requirements is to ensure that different batches of an approved drug fabricated by an approved manufacturer or a chemically identical product fabricated by another manufacturer be bioequivalent to the original product which had been approved on the basis of well-

controlled clinical studies. The rationale of the FDA bioavailability requirements is to determine whether product A delivers as much active moiety in the blood as the standard drug and is not applicable to the question in this proceeding of whether an earlier blood level of aspirin proves earlier onset of analgesia in clinical pain.

619. For the same reasons discussed hereinabove (F. 561–618, *supra*), it has not been established that Bufferin relieves pain *twice* as fast as aspirin (Complaint ¶ 7(B)(2)).

620. Bristol-Myers has represented that Bufferin relieves pain faster than aspirin for over 25 years. Throughout that period, it has never subjected that claim to clinical testing despite its realization of the importance of clinical studies to support its superiority claims for Excedrin and despite its public position that faster onset claims must be proved by clinical tests (F. 561, *supra*). In the face of evidence supplied by two respected panels of experts (CX 511 and CX 514), by publications relied upon by scientists in the field (CX 510, 512, 518), and by the testimony of four independent expert witnesses in this proceeding, it is found that it has not been scientifically established that the speed of relief provided by Bufferin is significantly greater than that provided by plain aspirin.

2. Claims that Tests Prove Bufferin Is Twice As Fast

621. Bristol-Myers represented that tests prove Bufferin acts twice as fast (F. 262–64, *supra;* Complaint ¶ 14A). These "tests" referred to in respondent's advertisements are "blood level" studies (CX 519C, D, 521, 522, 523–26, 527, 530, 788; Azarnoff, Tr. 9190–91; CX 536I (Spec 13a); CX 519A; Tr. 3861–71). Dr. Azarnoff specifically addressed these blood level [**161**] studies, and the issue of whether they prove that Bufferin relieves pain twice as fast as aspirin. He stated that they are blood level studies that show only that buffered aspirin is somewhat more rapidly absorbed than unbuffered aspirin and that no conclusions regarding buffered aspirin's pain relieving speed can be drawn from them (Azarnoff, Tr. 9192–95). Therefore, it is found that Bristol-Myers' speed claim challenged in Paragraph 14A is false.

3. The Substantial Question

622. Because Bufferin's superior speed of action claims challenged in Complaint Paragraph 7(A)(1)–(3) have not been scientifically established according to the criteria set forth and adhered to by experts in the relevant scientific community, these claims were made in the face of a substantial question recognized by such experts as to their validity as alleged in Complaint Paragraph 9(A)(1)–(3) and 10.

E. *It Has Not Been Scientifically Established That Bufferin Causes Significantly Less Stomach Upset Than Plain Aspirin*

623. Bufferin contains 5 grains of aspirin, 97.2 mgs. of basic magnesium carbonate and 49 mgs. of dihydroxy aluminum aminoacetate (aluminum glycinate) (CX 925C, R; CX 927I).

624. Magnesium carbonate and dihydroxyaluminum aminoacetate (aluminum glycinate) are recognized antacids (CX 514, p. 35469). An antacid may be defined as "[a]n agent that reacts with acid, such as the hydrochloric acid of the stomach (gastric acid), to neutralize it (decrease its amount)" (CX 514, p. 35373).

625. While it has been suggested by some that the presence of antacids of the type and in the amount found in Bufferin may lessen gastric irritation by speeding the dissolution of the aspirin tablet, and thereby increasing the rate at which aspirin leaves the stomach and is absorbed into the system, this theory is open to serious doubt (Grossman, Tr. 7772; CX 518G; CX 512H). To the extent the antacids in Bufferin increase aspirin dissolution, the increase is quite small (Grossman, Tr. 7772). The best that can arguably be claimed on the basis of biomedical evidence is that, for the relatively small population subset who experience occasional gastric discomfort from aspirin ingestion, Bufferin may reasonably be expected to provide somewhat less gastric discomfort for some of them some of the time.

626. The FDA OTC Internal Analgesics Panel has noted that "there is little meaningful difference between the rates of absorption of sodium salicylate, aspirin and the numerous buffered aspirin preparations of salicylates" (CX 514, p. 35378). **[162]**

627. It is generally agreed that disintegration rate is the limiting factor for absorption. While it is known that buffers can speed up disintegration of an aspirin tablet, the disintegration rate of an aspirin product such as Bufferin depends on many other factors. The disintegration and dissolution rate of aspirin is probably as dependent on the way it is made as the addition of buffers (Grossman, Tr. 7772), as well as the amount of food in the stomach (CX 514, p. 35378).

628. However, even if the increased rate of dissolution, disintegration and absorption of aspirin is appreciably increased by the addition of antacids, there is not direct evidence to date linking this phenomenon with a decrease in aspirin side effects such as stomach distress (Grossman, Tr. 7772).

629. Nor could antacids in the amount present in Bufferin be expected to neutralize the acidity of the stomach's contents and thereby lower the incidence of stomach distress associated with aspirin (Grossman, Tr. 7772, 7786–89, 7800). The amount of antacid in Bufferin is barely sufficient to neutralize the acid present in the aspirin portion

of Bufferin and could not significantly decrease, much less neutralize, the acidity of the stomach's contents as a whole (Grossman, Tr. 7771–72). Therefore, Bufferin could not decrease the damaging effects of aspirin on the stomach because it cannot neutralize the acid in it (Grossman, Tr. 7800). While respondent suggested that Bufferin was formulated as a substitute for the simultaneous administration of antacid with aspirin (Lanman, Tr. 11472–73), an effective dose of antacid employed for this purpose has over 75 times as much neutralizing capacity as Bufferin (Grossman, Tr. 7774).

630. Furthermore, since the addition of antacids to aspirin would only have effects prior to the absorption of aspirin into the system, it could in no event decrease the *systemic* effects of aspirin, which may contribute to aspirin-related stomach distress (Grossman, Tr. 7772–73; F. 651, *infra*).

631. Bristol-Myers did not present the testimony of a single expert witness in the field of gastroenterology. Again, its Medical Director, Dr. Lanman offered the only testimony supporting its position. Dr. Lanman has no experience of any kind in this field of science (BMRX 1). On the other hand, complaint counsel offered Dr. Morton Grossman, a renowned gastroenterologist (F. 44–47, *supra*), well qualified to render expert testimony on Bufferin's claims relating to side effects and on the issue of the medically significant side effects of aspirin.

632. The FDA OTC Analgesics Panel placed the claim that buffered aspirin "may prevent the stomach distress that plain **[163]** aspirin occasionally causes . . ." in Category III, finding available data insufficient to support the claim (CX 514, p. 35480). The Panel further noted that, even if buffered aspirin does reduce the incidence of aspirin-associated stomach distress, it would do so in "some but not all patients who exhibit gastric intolerance with plain aspirin tablets," and that the number of persons who might benefit from buffered aspirin over plain aspirin "is probably small," (CX 514, p. 35470). The Panel urged individual evaluation of label claims for buffered aspirin's lower incidence of gastric intolerance out of concern that such claims not "imply . . . decreased incidence of gastric distress is significant for most people" (CX 514, p. 35470). Moreover, the Panel stated: "Based upon the total evidence available to the Panel, it concludes that the evidence is insufficient to substantiate the claims that buffered aspirin or highly buffered aspirin solution is safe for use in patients who should not take regular, unbuffered (plain) aspirin" (CX 514, p. 35471). Dr. Grossman testified he would never prescribe Bufferin to a patient who experiences gastric intolerance with aspirin but would instead prescribe a non-aspirin, *e.g.*, acetaminophen, product; if, as in rheumatoid arthritis, the patient were required to take aspirin, Dr.

Grossman would place the patient on a full antacid regimen to be simultaneously administered with aspirin (Grossman, Tr. 7773).

633. There are no well-controlled clinical studies demonstrating that buffered aspirin, such as Bufferin, causes stomach distress less frequently than plain aspirin (Grossman, Tr. 7769–70; F. 634–42, *infra*). The existing evidence is equivocal at best (Grossman, Tr. 7770). The NAS/NRC Panel (CX 511) reviewed the claim that Bufferin helps prevent stomach upset often caused by aspirin, and concluded that most of the published studies with which it was familiar indicated little difference in the incidence or intensity of side effects from Bufferin or plain aspirin (CX 511F). The *Medical Letter* (CX 510) concluded that it has never been established that there is a difference between buffered and nonbuffered aspirin, *inter alia,* as regards incidence of gastrointestinal distress (CX 510B). Two editions of the *AMA Drug Evaluation* (CX 512, CX 518) similarly concluded that controlled clinical studies have not conclusively demonstrated that buffered aspirin will result, *inter alia,* in less gastric upset than plain aspirin (CX 518G; CX 512H). The FDA OTC Analgesics Panel placed the claim that buffered aspirin "may" cause less incidence of gastric intolerance in Category III, available evidence being insufficient to support the claim (F. 632, *supra*).

634. Respondent cited four (4) clinical studies which purported to compare the incidence of side effects with plain aspirin and Bufferin and reported a lower incidence of stomach upset with Bufferin. However, none of these studies were "well-controlled." (F. 635–40, *infra*). **[164]**

635. In the first study, by Tebrock, subjects who reported to a number of industrial clinics with ailments for which aspirin was normally prescribed were given Bufferin, and they were later interrogated regarding side effects (Lanman, Tr. 11478, 11486). The subjects were asked to compare the side effects they experienced in the study with their past experience with plain aspirin (Lanman, Tr. 11486). Such a "clinical" study is entitled to little weight in this proceeding. The subjects in this study were not tested with aspirin on a blinded or any other basis (Lanman, Tr. 12047). The study called for no administration of an aspirin treatment. The subjects reported the incidence of side effects with 12 tablets of Bufferin (2 tablets every 3 hours) while in the study, and then were asked to compare this side effects experience with Bufferin with what they remembered about past stomach distress which they thought was associated with plain aspirin (Lanman, Tr. 11486). This is called an "historical control" (Lanman, Tr. 12047). There is no way to determine whether the test subjects here accurately remembered and recounted their past experience with aspirin side effects, or, more importantly, whether they were able to

distinguish side effects attributable to aspirin from gastric discomfort occasioned by any one of a number of other possible causes (Lanman, Tr. 12043–44).

636. The design used in the Tebrock Study—employing a "historical control"—was described by Dr. Sunshine as "so far afield from reality" that he could not comment on its validity (Sunshine, Tr. 9686). The FDA regulations allow "historical controls" only where the nature and course of the disease being studied, if left untreated or treated by means other than the test treatment, is so well known and unacceptable that historical control is the only alternative to clinical trial, for example, "the high and predictable mortality" of childhood leukemia. *See* 21 C.F.R. 314.111(a)(4).

637. For similar reasons, in the second study, by Paul (CX 786), cited by respondent (Lanman, Tr. 11486–88), does not even approximate a well-controlled clinical trial. Only Bufferin was tested in the trial, using "historical control." For these reasons, the Paul Study is not reliable.

638. The third study, by Fremont-Smith, was published in the *Journal of the American Medical Association* in 1955. The study employed subjects suffering from arthritis and was divided in two parts: one involving short-term and the other long-term crossover administration of Bufferin and aspirin (Lanman, Tr. 11489). The long-term portion of the study was an "open trial," *i.e.,* both subject and investigator knew which drug was being administered, and thus cannot qualify for consideration as a well-controlled clinical investigation (Grossman, Tr. 7962). Apparently, only the short-term part of the study was double-blind (Lanman, Tr. 11489). However, patients were not randomly **[165]** assigned to treatments (Grossman, Tr. 7961): aspirin was given first to *most* subjects, then Bufferin (Lanman, Tr. 11489). The problem of drug administration "order effects" was thus built into the design of the study. The order in which test drugs are administered can significantly affect the results of a study unless controlled for, and a clinical study is seriously flawed if drugs are given in the same sequence to all subjects, and there is no way to examine the effects of such order bias (Sunshine, Tr. 9682, 9829). The physiological and psychological "carry over" problems which result where only one test drug is given during a particular period of the study (*e.g.,* here, where only aspirin was given for the first treatment), and/or the test drugs are given in the same order to all patients, can lead to "very, very misleading results" (Laska, Tr. 10433). A more proper way to conduct a study comparing two drugs is to randomize the patients to the treatments, or simply to give half the subjects one drug (*e.g.,* Bufferin) and half the other (*e.g.,* plain aspirin) during each period of treatment in the study (Laska, Tr. 10435).

639. Another flaw in the Fremont-Smith study lies in the fact that while the nurse administering the drugs and recording subject reactions to them was blinded, there is no statement that the nurse was unaware when the changeover was made from aspirin to Bufferin (Lanman, Tr. 12052), seriously compromising the blinding design of the study. The study itself notes that arthritic patients were subject to a variety of gastrointestinal abnormalities (Lanman, Tr. 12050). Therefore, even if it were otherwise well-controlled, the study might be generalizable only to persons suffering similar gastrointestinal abnormalities (Lanman, Tr. 12050). Dr. Grossman noted that the report of the study, as published, did not provide sufficient information to allow full evaluation (Grossman, Tr. 7961). Dr. Grossman also noted that there have been no published studies since the 1955 Fremont-Smith Study which purport to be well-controlled and double blind, addressing the same question. If such studies had been done, they would be of great interest to the medical community and would have been published (Grossman, Tr. 8011).

640. The third study (1958), by a Dr. Sher (CX 506Z572–Z580), reported the results of a clinical trial conducted at a prison hospital in Michigan, comparing the incidence of gastric intolerance with Bufferin, four (4) unnamed brands of aspirin, and three (3) unnamed APC products (Lanman, Tr. 11491–98). The Sher Study is entitled to little weight in this proceeding. The study was never published (Lanman, Tr. 12061). Dr. Lanman, the only witness who testified about the study, and was not even employed by Bristol-Myers at the time the study was undertaken (Lanman, Tr. 12054). While it is known that Dr. Sher was a prison doctor (Lanman, Tr. 12054), there is no evidence that he had ever conducted clinical research before this study (Lanman, Tr. 12054). Nor is there any evidence of the identity, qualifications, experience and training of others who administered **[166]** the study. For all we know, they may have been prison "trusties" or other untrained personnel. Dr. Lanman admitted it was highly likely that Sher's study was submitted by Bristol-Myers to the NAS/NRC Panel (Lanman, Tr. 12061), which considered Bufferin's claim of lower incidence of stomach upset and concluded that the claim lacked support (F. 633, *supra*).

641. Finally, respondent cited Dr. Calabro, a doctor who conducted some studies for Bristol-Myers in the mid–1960's (Lanman, Tr. 12040–41), for a statement regarding lessened abdominal complaints with buffered aspirin than with plain aspirin (Lanman, Tr. 11501). The only basis for Dr. Calabro's statement is a reference to an article by Brewer (Lanman, Tr. 12035), which in turn cited no support other than personal experience (Lanman, Tr. 12036). This is nothing more than anecdotal evidence.

642. In sum, of the four clinical studies cited by respondent, two did not directly compare Bufferin and aspirin; one was not randomized, failed to correct for order effects, and therefore is seriously flawed; and one was unpublished, and there is no record evidence to attest to its reliability or to accord it any weight. Not one of the studies cited by Bristol-Myers meets the criteria of well-controlled clinical studies necessary to establish that there is a lower incidence of gastric distress with Bufferin than with plain aspirin. It is not surprising that Bristol-Myers had clinical studies, other than those it chose to rely upon, which failed to show any superiority of Bufferin over aspirin with regard to gastric discomfort (Lanman, Tr. 11499).

643. Therefore, the advertising representations challenged in Paragraphs 7(A)(4) and 7(A)(5) of the Complaint were made in the face of a substantial question recognized by experts as to their validity, as alleged in Complaint Paragraphs 9(A)(4) and (5) and 10, and therefore are false.

F. *The Fact that Bufferin, Excedrin, and Excedrin P.M. Contain Aspirin is Not Known to a Substantial Number of Consumers and is a Material Fact Which Should be Disclosed in Advertising*

1. Gastrointestinal Effects of Aspirin

644. Bufferin, Excedrin and Excedrin P.M. contain aspirin (F. 2, *supra*).

645. Aspirin has well recognized adverse effects on the gastrointestinal tract. These side effects include dyspepsia and massive gastrointestinal bleeding (Grossman, Tr. 7724–28, 7741–43, 7821, 7985). Aspirin can also exacerbate and may even cause gastric ulcers in substantial numbers of people (Grossman, Tr. 7727, 7744–45). It is well known among experts that **[167]** initiation or exacerbation of stomach ulcers, stomach irritation and intestinal inflammation occurs in a significant number of individuals who take aspirin (CX 514, p. 35390).

646. Aspirin-induced dyspepsia includes general gastric discomfort, pain, nausea, and what is commonly called heartburn, occurring in the upper abdominal region (Grossman, Tr. 7724–25; CX 514, p. 35387).

647. Dyspepsia due to aspirin is a common occurrence (Grossman, Tr. 7725). The estimated incidence of dyspepsia in persons taking smaller doses of aspirin (*e.g.*, single dosages) is up to 10% (Grossman, Tr. 7725; CX 514, 35387). However, the estimated incidence increases to between 20 and 30% among those taking larger doses over an extended period of time, such as arthritics (Grossman, Tr. 7725–26).

648. Everyone experiences some occult blood loss (*i.e.*, imperceptible

loss of blood) from the gastrointestinal tract upon ingestion of aspirin (Grossman, Tr. 7757). However, such occult bleeding has no clinical significance, except in those few individuals with higher than normal blood loss and a tendency toward anemia, where bleeding induced anemia may occur (Grossman, Tr. 7757). No association has been established between occult bleeding and clinically important side effects of aspirin, such as dyspepsia and massive gastrointestinal bleeding (Grossman, Tr. 7758). While highly buffered aspirin (*e.g.,* aspirin preparations, such as Alka-Seltzer, containing larger quantities of antacids and in which the aspirin is put in soluble form) has been shown to reduce the magnitude of occult blood loss due to aspirin, it has not been shown that this decrease is associated with a decrease, for example, in dyspeptic symptoms (Grossman, Tr. 7759–60). Thus, the possibility that a buffered aspirin tablet may reduce aspirin-associated occult bleeding to a relatively small degree does not suggest that it would reduce incidence of clinically important gastrointestinal side effects (Grossman, Tr. 8012). All forms of aspirin, buffered or unbuffered, pose a potential hazard as regards clinically important gastrointestinal events (Grossman, Tr. 8008–09).

649. The available data are not sufficient to demonstrate that buffered aspirins, such as Bufferin, cause a lower incidence of dyspepsia than plain aspirin.

650. Aspirin, even in single doses, causes damage to the gastric mucosa in the form of lesions, detectable by visual examination and/or on microscopic examination (Grossman, Tr. 7758–59).

651. While the means by which aspirin injures the gastric mucosa, and thus causes adverse effects on the gastrointestinal tract has not been established, at least two mechanisms are **[168]** involved: (1) a topical action (Davenport effect) in which aspirin, as it is absorbed into the gastric mucosa, causes injury in the form of erosion, hemorrhaging and cell damage (seen as lesions on the mucosa); (2) a systemic effect, in which aspirin after entering the bloodstream interferes with the normal mechanisms protecting the gastric mucosa (Grossman, Tr. 7762–64). In Dr. Grossman's opinion, both these mechanisms contribute to gastrointestinal blood loss (Grossman, Tr. 7764).

652. Aspirin can cause massive, life-threatening gastrointestinal bleeding (Grossman, Tr. 7727–28, 7741–43). Associations between ingestion of single doses and massive blood loss have been reported (Grossman, Tr. 7743; CX 514, p. 35393). Dr. Grossman estimated that 5 to 10% of massive gastrointestinal blood loss is due to aspirin ingestion (Grossman, Tr. 7985). Clinically important gastrointestinal blood loss can lead to weakness and shock, usually requires hospitalization, and may require surgical intervention (Grossman, Tr. 7742; CX 514, p. 35391). Severe gastrointestinal blood loss is the most serious ad-

verse side effect of aspirin on the gastrointestinal tract (Grossman, Tr. 7741; CX 514, p. 35391). The mortality risk is high (CX 514, p. 35391). There is between a 5 and 10% mortality rate from severe gastrointestinal bleeding (Grossman, Tr. 7741).

653. The incidence of massive bleeding is not insignificant. There is a recognized higher risk of severe gastrointestinal blood loss among persons with peptic ulcers, and those who have had prior experiences of gastrointestinal blood loss or dyspepsia, and these persons should avoid aspirin (Grossman, Tr. 7764; CX 514, p. 35392).

654. Aspirin may not only present a grave risk to those persons with pre-existing gastric ulcers, by increasing gastrointestinal bleeding, but in large doses may actually cause gastric ulcers (Grossman, Tr. 7727; CX 514, p. 35390). There is evidence that aspirin may produce a specific kind of ulcer, not seen in its absence (Grossman, Tr. 7745–47, 7753–54; CX 514, p. 35390).

655. Gastric ulcers are a serious disease, causing significant morbidity, stomach perforation, obstruction of the flow of food from the stomach, peritonitis, and often requiring surgery on the stomach (Grossman, Tr. 7756).

656. By conservative estimate, most notably reported by Levy in his Boston Collaborative Group studies, aspirin ingestion results in 10 out of every 100,000 users developing a gastric ulcer, requiring hospitalization (Grossman, Tr. 7840; CX 514, p. 35391). The Levy Study also estimated that one-eighth of all gastric ulcers were related to aspirin (CX 514, p. 35390), and Dr. Grossman testified that a history of aspirin ingestion is found in 20 to 30% of individuals with gastric ulcers (Tr. 7756). **[169]**

657. Dr. Grossman is familiar and agreed with the report of the FDA OTC Internal Analgesics Panel as it related to the nature, incidence and severity of aspirin-related side effects (Grossman, Tr. 7782). In this connection, the FDA Panel noted, *inter alia,* that in a recent survey, the adverse effects of aspirin on the gastrointestinal tract were the second most frequent drug-involved adverse effect that was serious enough to require hospitalization. Two out of every 1,000 hospital admissions were attributed to aspirin (CX 514, p. 35392, reporting on the results of a survey by the Boston Collaborative Drug Surveillance Program).

658. Aspirin also interferes with blood clotting, and should be avoided by persons with a history of blood coagulation defects, those receiving anticoagulant drugs, or those with severe anemia (CX 514, p. 35385).

659. The FDA Analgesics Panel has recommended that the following warning appear on all aspirin-containing products, regardless of formulation: "Caution: Do not take this product if you have stomach

distress, ulcers or bleeding problems except under the advice or supervision of a physician" (CX 514, p. 35395).

2. Allergic Side Effects of Aspirin

660. Aspirin may also have respiratory and allergic side effects, including severe and even life threatening attacks to those suffering asthma.

661. An asthmatic attack involves a spasm and resulting constriction of the bronchial tubes. Symptoms include shortness of breath, coughing, and in severe cases, hypoxia (*i.e.*, insufficient delivery of oxygen to red blood cells), shock and occasionally death within minutes of an attack (Stevenson, Tr. 1481; Farr, Tr. 2565–66, 2571–72; CX 514, p. 35398).

662. Ingestion of anywhere from 3 mg. to 650 mg. of aspirin can cause an asthmatic attack among susceptible members of the asthmatic population (Stevenson, Tr. 1480). The severity of the aspirin-induced asthmatic attack depends on the degree of bronchial constriction prior to ingestion of the aspirin. If the bronchial tubes are already partly closed, the attack can be severe or even life threatening (Stevenson, Tr. 1489).

663. Combining aspirin with buffering ingredients, as in Bufferin, will not mitigate aspirin's asthmatic side effects (Farr, Tr. 2576; Stevenson, Tr. 1490–91). While the number of asthmatics in the population is uncertain, as is the number of asthmatics sensitive to aspirin, the incidence of persons susceptible to aspirin-induced asthmatic attacks is not insignificant. Dr. Stevenson cited a 1972 study by Davis **[170]** concluding that 9 million persons were under care for asthma (Stevenson, Tr. 1494).

664. The Tecumseh Study, an epidemiological study of health problems of the residents of a Michigan town, is the best evidence available on the incidence of asthmatics in the general population, and reported that 6% of the townspeople had conditions previously diagnosed as asthma, and another 6% had medical histories consistent with asthma (Stevenson, Tr. 1494).

665. Dr. Stevenson's own study, which "challenged" asthmatic patients not known to be sensitive to aspirin with aspirin, led him to conservatively estimate that 10% of the asthmatic population is sensitive to aspirin (Stevenson, Tr. 1498). A study by Dr. Farr found 17.36% of asthmatics intolerant to aspirin, a figure he believed low because certain high risk subjects were excluded from the study (Farr, Tr. 2589–2605). The FDA Analgesics Panel estimated that between 6 to 20% of asthmatics are sensitive to aspirin (CX 514, p. 35397).

666. Aspirin may also cause dermal allergic reactions, particularly urticaria (hives) and angiodema (giant hives and swelling) (Stevenson,

Tr. 1512; CX 514, p. 35398). Such reactions are not usually life threatening (Stevenson, Tr. 1511; CX 514, p. 35398), but urticaria may be serious if the lining of the stomach is involved, and angiodema may be fatal if swelling takes place in the vocal chords, cutting off breathing (Stevenson, Tr. 1512).

666a. In some persons a few molecules of aspirin will cause a dermal reaction, in others a relationship between dose and severity has been seen (Stevenson, Tr. 1513). By contrast to asthmatic reactions, the incidence of dermal reactions is very small (Stevenson, Tr. 1464).

667. The overall incidence and severity of allergic reactions to aspirin is such that the American Academy of Allergy, a professional organization with a membership of some 2,200 allergists, adopted the following resolution in 1973:

> While recognizing that acetylsalicylic acid (aspirin) is a valuable drug, the American Academy of Allergy recommends that a formulation containing aspirin and advertisements promoting the formulation should clearly indicate that the preparation contains aspirin and that aspirin can be harmful to some persons.

In the same year, the American College of Allergists, another professional organization of allergists, passed a similar resolution (Farr, Tr. 2608–12). [171]

668. The FDA OTC Internal Analgesics Panel stated its agreement with the Academy resolution (CX 514, p. 35398). The Panel has recommended that the following warning should appear on all products containing aspirin:

> This product contains aspirin. Do not take this product if you are allergic to aspirin or if you have asthma except under the advice and supervision of a physician. (CX 514, p. 35399).

669. Disclosure in advertising that Bufferin, Excedrin and Excedrin P.M. contain aspirin would be important to the substantial number of people who for sound medical reasons should avoid aspirin, and may not be aware that these products contain aspirin (Grossman, Tr. 7765–67; Moertel, Tr. 5625–26). There are large numbers of people who should avoid aspirin and are so warned (Grossman, Tr. 7767; Moertel, Tr. 5625–26). Dr. Stevenson testified, for example, that he warns patients he identifies as aspirin sensitive to avoid aspirin, but most asthmatics do not know if they are aspirin sensitive or not, and should avoid aspirin as a precaution (Stevenson, Tr. 1502). Immunologists generally warn asthmatics to avoid aspirin (Farr, Tr. 2601, 2606).

670. However, many patients do not know that an OTC aspirin product which does not contain "aspirin" in its brand name, such as Bufferin and Excedrin, in fact contains aspirin. Because of this prob-

lem, some persons warned not to take aspirin will take it anyway (Stevenson, Tr. 1509; Moertel, Tr. 5625–26; Grossman, Tr. 7766–67; 7879–80).

671. The particular danger posed by aspirin unawareness was made clear, in Dr. Moertel's experience, when large numbers of his patients, whom he warned against aspirin-containing drugs, took aspirin products not knowing their aspirin content. This subsequently caused gastrointestinal bleeding and hospitalization (Moertel, Tr. 5625–26). Dr. Grossman also cited a specific example of a patient suffering from a peptic ulcer, who was warned not to take aspirin, but who developed upper gastric bleeding and later recounted that he had taken Excedrin (Grossman, Tr. 7880).

672. Disclosure of aspirin content on the product label alone is not a sufficient means of alerting persons who should avoid aspirin. In the experience of doctors testifying in this proceeding, patients generally do not read labels on medications carefully, if at all (Grossman, Tr. 7767; Moertel, Tr. 5625–26).

673. A substantial number of consumers do not know, and have not known for a long time, that Bufferin and Excedrin contain aspirin. In a survey of consumers conducted by the Gallup [172] organization in 1964 (CX 333), only 19% of a nationally projectable sample could name aspirin as an ingredient in Bufferin on an unaided basis; 74% of the sample could not name any ingredient in Bufferin (CX 333H). In that same study, when consumers were directly asked whether aspirin was an ingredient in Bufferin, only 46% answered affirmatively (Ross, Tr. 7463–64; CX 333J).

674. In the Vanquish Study (CX 347–48) the predominant response among Bufferin users who were asked to state the number of ingredients in Bufferin was "Don't know" (61%) (Ross, Tr. 7464–66; CX 348Z041); only 41% of those who stated Bufferin contained "more than one ingredient" were able to name aspirin as an ingredient in Bufferin (Ross, Tr. 7464–66; CX 348Z043). The predominant response among Excedrin users who were asked to state the number of ingredients in Excedrin was "Don't know" (56.8%) (Ross, Tr. 7465; CX 348Z041); 33% of those who stated that Excedrin contained "more than one ingredient" could name aspirin as an ingredient in Excedrin (Ross, Tr. 7465–67; CX 348Z043). Only 33.1% of Bufferin users and 25.8% of Excedrin users agreed that "all advertised brands rely chiefly on aspirin to relieve pain," which indicates a general lack of awareness of aspirin as an ingredient in both Bufferin and Excedrin (Ross, Tr. 7467–68; CX 348Z251).

675. In the 1967 and 1970 Oxtoby-Smith studies (CX 1058 and 1059), consumers showed a general lack of awareness of ingredients by the magnitude of their responses to the question "I have little idea of

ingredients in the headache tablets I take." In 1967, approximately 62.3% of male and 46.2% of female Bufferin users agreed with that statement and approximately 61.3% of male and 47.9% of female Excedrin users agreed; in 1970, approximately 63% of male and 49.2% of female Bufferin users agreed with that statement and 59.9% of male and 57.9% of female Excedrin users agreed (Ross, Tr. 7474–75; CX 1058Z460; CX 1059Z179).

676. In the 1972 Pain Reliever Telephone Study (CX 314), only 23% of the consumers surveyed were able to name aspirin as an ingredient in Bufferin (Ross, Tr. 7470–71; CX 314Z006); 70% of the sample surveyed could not name any ingredients in Bufferin (CX 314Z007). For Excedrin, only 21% of the consumers surveyed were able to name aspirin as an ingredient in Excedrin (Ross, Tr. 7471–72; CX 314Z008). Seventy-seven percent of the sample surveyed could not name any ingredients in Excedrin (CX 314Z009).

677. Dr. Moertel conducted an informal survey of two samples of individuals with whom he came in contact in his duties at the Mayo Clinic in the recent past (Moertel, Tr. 5626; CX 810A-C). The first sample consisted of 100 patients and their family members who came to the cancer treatment center at the Currie Pavillion at the Clinic (Moertel, Tr. 5626). The second sample [**173**] consisted of 100 paramedical personnel who, although nonphysicians, had some responsibility in dealing with medicine and worked in a medical setting (Moertel, Tr. 5626–27). A short form questionnaire, developed by Dr. Moertel, was self-administered by each respondent with the nurse technicians at the clinic available to answer any questions regarding the form. The questionnaire included questions about age, sex, educational level and asked whether a number of drugs printed on the questionnaire contained aspirin. Respondents were simply asked to check "yes" or "no" or "don't know" (Moertel, Tr. 5626–27; CX 810A).

678. Ninety percent of the paramedics correctly identified aspirin as an ingredient in Bufferin; 3% said Bufferin did not contain aspirin; and 7% checked the "don't know" response (Moertel, Tr. 5629; CX 810B). For Excedrin, 84% of the paramedics correctly identified aspirin as an ingredient in Excedrin; 1% stated Excedrin did not contain aspirin; and 15% checked the "don't know" response (Moertel, Tr. 5630; CX 810B).

679. Of the 100 patient/family member sample, 68% correctly indicated that Bufferin contained aspirin; 4% stated Bufferin did not contain aspirin; and 28% did not know. For Excedrin, 65% correctly indicated that Excedrin contained aspirin; 1% stated Excedrin did not contain aspirin; and 34% stated they did not know (Moertel, Tr. 5631; CX 810C).

680. Mr. Ivan Combe, the Chairman of the Council on Family

Health ("CFH") (Tr. 9397) testified that CFH has been preparing and making available to networks, television stations, and magazines, advertisements advising consumers to read the labels of drug products since 1972 (Tr. 939H–17; 9401–02). BMRX 128 contains copies of five such film advertisements (Tr. 9404). BMRX 17A-E are copies of print advertisements which CFH supplied to magazines and newspapers for public service exposure (Tr. 9405–08).

681. The value of the television time allocated to CFH's read-the-label campaign in 1977 was approximately $2.7 million in 1976 and $500,000 in 1975. The value of the broadcast time and print space for CFH's read-the-label campaign during 1974 was approximately $1.25 million (Tr. 9414–24). One print advertisement alone, "Why trust your memory when you can be sure" appeared in a number of major magazines, including Good Housekeeping, Esquire, Family Health, U.S. News & World Report, People and Business Week with an estimated exposure to 4.6 million people (Tr. 9453) (BMRX 17E).

682. The purpose of the read-the-label campaign is to inform the consumer on the proper use of medicines in the interest of safety and efficacy (Tr. 9440). CFH felt that the use of a succinct, incisive message would be most effective in communicating to consumers (Tr. 9442–43; 9448–49). The theory behind [174] the campaign is that, if the public is given the right general advice and they follow it, all of the specifics will be covered (Tr. 9451) including awareness of the ingredients contained in the drug product (Tr. 9456).

683. It is found that the primary purpose of the "read the label campaign" is to educate the consumer to read and heed the label instructions regarding doses in the interest of safety and efficacy of OTC drug products. As such the campaign is important in view of the fact that many OTC drug products, such as analgesics, contain potent drugs and can cause serious harm when misused or abused. However, it is highly doubtful whether the consumer who reads the label for dose information will also read the ingredients information.

684. Thus, the fact that Bufferin and Excedrin contain aspirin is a material fact and is not known to a substantial number of consumers. A failure to disclose that material fact in advertising for these products is misleading and deceptive. Therefore, the existence of aspirin in Bufferin and Excedrin should be disclosed in all advertising for these products.

G. *The Ingredients, Either Individually Or In Combination, In Bufferin, Excedrin or Excedrin P.M. Do Not Relieve Tension*

685. Tension (often used synonymously with "anxiety") exhibits symptoms such as headache, depression, anger, hostility, fear, heart palpitation and perspiration (Rickels, Tr. 6516–17) and is appropriate-

ly treated with antianxiety agents or tranquilizers (Rickels, Tr. 6525; CX 513Z003).

686. From 1961–1970 Bristol-Myers made claims in advertisements that Bufferin, Excedrin, and Excedrin P.M. will relieve tension, stress, anxiety and enable persons to cope with the ordinary stresses of everyday life. It also made claims of efficacy for tension relief in labels of Bufferin and Excedrin (CX 815; CX 816; CX 817; CX 818; CX 820; CX 800).

687. The nonantacid active ingredient in Bufferin is aspirin. The ingredients in Excedrin are acetaminophen, salicylamide, aspirin and caffeine. The ingredients in Excedrin P.M. are acetaminophen, salicylamide, aspirin and methapyrilene fumarate. None of these ingredients, either alone or in combination, are considered to be effective antianxiety agents or tension relievers. An ingredient in Excedrin—caffeine—is contraindicated for the treatment of tension.

688. Headache pain can be a symptom of tension. In such instances, the headache pain is caused by the underlying tension (Rickels, Tr. 6518, 624). Underlying tension may also, however, exist simultaneously with, but independently of, a headache without causing it (Rickels, Tr. 6519–20). In such a case, the **[175]** headache is caused by something other than the underlying tension. In either case, however, this headache pain may act to aggravate underlying tension; *i.e.,* someone becomes tense, or more tense, because he has a headache. This situation is called the "tension-headache-tension" cycle (Rickels, Tr. 6519–20, 6524). In those instances when an individual is suffering from tension, which causes a headache as one of its symptoms, aspirin is neither appropriate nor indicated for the treatment of the underlying tension (Rickels, Tr. 6532–33). As an analgesic, aspirin will relieve the pain of headache and, because that pain is gone, the tension caused by the pain may be lessened. But the aspirin will never treat the tension that caused headache in the first place (Rickels, Tr. 6525–27, 6530). Therefore, the only sense in which aspirin can be considered a tension reliever is that it may indirectly relieve the tension caused wholly by pain, while not affecting the underlying tension (Rickels, Tr. 6528). To consider aspirin a tension reliever would be the same as calling an antibiotic a tension reliever in a situation where an infection causes one to be tense. The antibiotic relieves the infection which, in turn, would relieve the tension caused by having an infection. But neither aspirin nor the antibiotic can be said to be tension relievers because neither has any direct tension relieving properties; neither is helpful in treating tension *per se* (Rickels, Tr. 6528–29).

689. In determining whether there is reason to believe that a drug has tension relieving properties, information derived from a well-controlled, randomized, double-blinded clinical study in a well-de-

fined population is given the most weight (Rickels, Tr. 6499, 6529–30, 6548).

690. Dr. Lanman, Bristol-Myers' Medical Director, stated that Bristol-Myers relied on four published papers, an article, and a textbook as its basis for the claim that aspirin and acetaminophen have tension relieving properties. Specifically, the materials relied upon by Bristol-Myers are: (1) two studies by Krumholtz and Merlis, dated 1964 and 1965; (1) a 1954 medical textbook, *Pharmacology and Medicine,* edited by Drill; (3) a 1957 review article entitled "Current Concepts in Therapy" published in the *New England Journal of Medicine;* (4) a 1957 report by Boyd, Gittinger, and Schwimmer entitled "Sleep Induction With Salicylamide and Acetophenetidin"; and (5) a 1959 report by Boyd, Huppert, Sullivan, and Molinus entitled "Hypnotic Effects of Bufferin" (Lanman, Tr. 12161–74). Bristol-Myers has never funded a study to determine or evaluate the amount or degree of tension relief afforded by Bufferin, Excedrin, or Excedrin P.M. (CX 925J; CX 927B).

691. All of the six sources relied upon by Bristol-Myers are dated. The Drill textbook is dated as early as 1954. Dr. Lanman was asked if Bristol-Myers could supply a more recent edition of this textbook which contains the same purported support, but Dr. Lanman stated that he could not (Lanman, Tr. [**176**] 12169). The other sources are dated from 1957 to 1965. In 1965, a date when *all* materials relied upon by Bristol-Myers in this proceeding were extant, Dr. Beaver completed a comprehensive review of all the sources of evidence—including those solicited directly from Bristol-Myers—on the pharmacologic properties of analgesics (Beaver, Tr. 5897–5900). He specifically found that among the over 1,000 articles and other materials he analyzed, there was "no good evidence" that mild analgesics have tension relieving properties (Beaver, Tr. 5897–98; Lanman, Tr. 12151–54).

692. The two Krumholtz and Merlis studies, the only post-1959 studies cited by Dr. Lanman, while interesting, are not the sort of evidence which scientists in biomedicine generally accept as establishing a proposition. The studies used volunteers and apparently were not randomized (Rickels, Tr. 6572, 6579–80). The authors also do not use the standard index for the measurement of tension relief (Rickels, Tr. 6573–74). The authors reported insufficient data to allow a meaningful analysis of their results (Rickels, Tr. 6572, 6579). Above all, the authors themselves recognized the deficiencies of their data, concluding that further studies were needed to test the efficacy of aspirin as a tension reliever (Rickels, Tr. 6634–35; Lanman, Tr. 12258). Likewise, Dr. Beaver, in his landmark review of literature on mild analgesics, explicitly referred to the Krumholtz and Merlis studies as not providing evidence of tension relieving properties of

aspirin and merely "productive of inconsistent results" (Lanman, Tr. 12152).

693. The 1959 report by Boyd, Huppert, Sullivan, and Merlis does not provide a reasonable basis for the assertion that aspirin has tension relieving properties. The authors, whose reputations for research are not widely known, reported that Bufferin showed an ability to induce hypnotic (somnifacient) effects in the test subjects. However, it is not clear whether the sample of 102 custodial care patients (with a median age of 64) were randomized (Rickels, Tr. 6593). Apparently the authors used test subjects who had physical problems which produced pain. Therefore, it is possible that the reported results may be attributable to the pain relieving properties of aspirin rather than to any tension relieving properties of aspirin (Rickels, Tr. 6593). These methodological problems led Dr. Rickels to state that he had "great doubts about the results," particularly since they purport to show that Bufferin's tension relieving abilities exceeded those of most prescription drugs indicated for the relief of tension (Rickels, Tr. 6591–95).

694. The 1957 report by Boyd, Gittinger, and Schimmer was on the hypnotic effects of a drug called Effisin, which contained salicylamide and acetophenetidin. Salicylamide is an ingredient in Excedrin. Dr. Lanman did not say that the ingredient salicylamide has any tension relieving properties. His position [**177**] is limited to aspirin and acetaminophen (Lanman, Tr. 11509–10, 12149–51). In any event, Dr. Beaver, in his comprehensive review of all the research as of 1965 regarding the properties of mild analgesics (Beaver, Tr. 5897–99), found that there "was no good evidence" that such drugs had any tension relieving properties (Lanman, Tr. 12152).

695. The 1954 textbook by Drill and the 1957 review article published in the *New England Journal of Medicine* do not provide a reasonable basis to show the efficacy of aspirin or acetaminophen as tension relievers. Neither of them involved clinical trials. The FDA Panel on OTC Sedative, Tranquilizer, and Sleep-Aid Drug Products did not consider such textbooks and review articles as evidence of a drug's efficacy in their evaluations (Rickels, Tr. 6547–48). Statements from the medical literature and textbooks relating to the tension relieving ability of analgesics were also not accorded much weight by Dr. Beaver in his comprehensive review of the scientific literature on this point. He found that the statements in the literature were often based on a study of three subjects (who were also the investigators) without the benefit of blinding or placebo controls (Lanman, Tr. 12154).

696. On the other hand, all authoritative studies published after Dr. Beaver's 1965 review article have consistently found that there was

no evidence to show that mild OTC analgesics such as aspirin have tension relieving properties. Bristol-Myers continued to make tension relief claims for Bufferin, Excedrin, and Excedrin P.M. until 1970.

697. In a 1973 well-controlled, double-blinded clinical study of Compoz, Librium, aspirin and placebo in normal doses in patients suffering moderate degrees of tension, aspirin was found not to be significantly different from placebo in terms of its ability to relieve tension (Rickels, Tr. 6500, 6511–13, 6517). This result is consistent with the creditable scientific literature regarding the lack of tension relieving properties of aspirin, and confirms Dr. Beaver's conclusion in 1965 that a therapeutic dose of aspirin cannot be considered a tension reliever (Rickels, Tr. 6517).

698. Further confirmation of this view is found in the FDA Internal Analgesics Panel, which concluded that aspirin is "clearly ineffective" for "nervous tension" (CX 514, p. 35355). Likewise, the FDA Advisory Panel on OTC Sedative, Tranquilizer, and Sleep-Aid Drug Products determined that aspirin was "ineffective" as a "day time sedative" product, which the Panel defined as one that claims "mood-modifying indications such as for the relief of occasional simple nervous tension" (CX 513E, Z002; Rickels, Tr. 6538–39). The Sedative Panel made the same conclusions with respect to acetaminophen and salicylamide (CX 513E; Rickels, Tr. 6540). [178]

699. In 1975 a minority of the FDA Sedative Panel considered methapyrilene (an ingredient in Excedrin P.M.) to be ineffective as a daytime sedative; *i.e.*, a tension reliever. A majority voted to place methapyrilene in Category III, that is to allow manufacturers limited time to develop studies to show the efficacy of methapyrilene as a daytime sedative (Rickels, Tr. 6541–42). While the majority recognized that the research at that time did not show any tension relieving properties for methapyrilene, they felt that the industry should be given an opportunity to identify any population which could benefit from that compound (Rickels, Tr. 6550–51). However, the unanimous opinion of the Panel was that the studies would never show methapyrilene's efficacy for the relief of nervous tension (Rickels, Tr. 6541, 6551). Since no research on this issue has been forthcoming, Dr. Rickels testified that all members of the Panel now believe that methapyrilene should be placed in Category II as a daytime sedative (Rickels, Tr. 6541, 6550).

700. In 1972, after a review of the published literature and after having considered scientific materials submitted by Bristol-Myers in support of its labeling claims, the National Academy of Sciences-National Research Council (NAS/NRC) Drug Efficacy Study Group specifically considered the claim on Bufferin's label that it was indicated for "mild temporary tension" (CX 511C, F). The Panel found

that while Bufferin was "possibly" effective for the relief of tension, there was "very little evidence that aspirin has any tranquilizing or sedative effect" (CX 511F).

701. The combination of antacids with the aspirin in Bufferin does not change aspirin's inability to relieve nervous tension. Thus Bufferin is not effective for the treatment of nervous tension (Rickels, Tr. 6534).

702. The ingredients in Excedrin (aspirin, salicylamide, acetaminophen and caffeine) either alone or in combination are not effective for the relief of nervous tension (Rickels, Tr. 6532). In fact, the recommended dose of Excedrin contains 130 mg. of caffeine, a dose in excess of the clinical dose of caffeine (100 mg.) prescribed as stimulant (Rickels, Tr. 6530–31). A daily dosage of 8 Excedrin tablets contains 520 milligrams of caffeine, which is just short of the level of 600–650 milligrams of caffeine which are known to cause anxiety (Rickels, Tr. 6530–31). The tension reliever claim for Excedrin is patently inconsistent with the presence of such a high level of caffeine in Excedrin.

703. The principal difference between Excedrin and Excedrin P.M. is that Excedrin P.M., in addition to aspirin, acetaminophen and salicylamide, contains methapyrilene instead of caffeine (Rickels, Tr. 6541; F. 2, *supra*). The substitution of methapyrilene for caffeine, a stimulant, does not make Excedrin P.M. effective for the relief of nervous tension. **[179]** The three ingredients Excedrin P.M. shares with Excedrin are not effective for tension relief. The addition of methapyrilene, an ineffective drug for the relief of nervous tension (F. 699, *supra*), will not alter that result. Thus Excedrin P.M. is not effective for the relief of nervous tension.

704. Respondent Bristol-Myers did not have a reasonable basis for the claims that Bufferin, Excedrin, and Excedrin P.M. relieve nervous tension (Rickels, Tr. 6530; F. 691–700, *supra*). None of the materials offered by Bristol-Myers are sufficient to substantiate such claims. The four research studies offered by respondent all had serious methodological defects and cannot be considered to be well-controlled clinical studies. The other sources offered by respondent were statements found in dated and superseded scientific literature, and thus cannot be accorded much, if any, weight. The inadequacy of these sources is confirmed by Dr. Beaver's comprehensive published review in 1965 of all the sources on this issue which explicitly dismissed the only current evidence relied on by Bristol-Myers and which found that there was no good evidence that mild analgesics have any tension relieving properties. The inadequacy of Bristol-Myers' sources has also been confirmed by three panels of independent scientific experts: the FDA Internal Analgesic Panel, the FDA Sedative Panel, and the NAS/NRC Drug Efficacy Study Group, all of which found that there

was insufficient evidence to support a claim that aspirin has tension relieving properties. It was also the testimony of one of the country's foremost experts in psychopharmacology, Dr. Rickels, that the available scientific evidence does not support any tension relieving claim for mild analgesics. Thus the record clearly shows that during the time respondent disseminated tension relief claims, there was no reasonable basis for the claim that its mild OTC analgesics had tension relieving properties.

H. *Other Representations for Bufferin and Excedrin P.M.*

1. The "Doctors Recommended" Claim for Bufferin Lacked a Reasonable Basis

705. Particularly through use of the phrase "Doctors specify Bufferin for minor pain more than any leading brand of pain reliever you can buy," respondents represented that physicians recommend Bufferin more than any other nonprescription analgesic (Complaint ¶ 17; F. 253, *supra*). This representation was unfair and deceptive because respondents did not possess or rely on competent and reliable evidence sufficient to provide a reasonable basis for it.

706. In substantiation for this claim, Bristol-Myers submitted to complaint counsel, in response to subpoena, documents received in evidence as CX 364–390 and 676 (CX 838A). CX 364–[180]380 comprise 17 portions of the National Prescription Audit (NPA), a national survey of drug prescription activity (CX 838A). CX 381–390 are portions of the National Disease and Therapeutic Index (NDTI), a national survey of drug treatment activity (CX 838E). CX 676 is explanatory material relating to the NPA (CX 838E). Neither the NPA nor the NDTI provide competent and reliable evidence in support of respondent's claim because: (a) the NPA monitors drug *prescription* activity and is not representative of doctors' recommendations for the *nonprescription* drugs that are in issue here; (b) the NDTI, which unlike the NPA is designed to reflect doctors' recommendations of both nonprescription and prescription drugs, shows doctors recommending other nonprescription analgesics more than Bufferin (F. 708, *infra*).

707. NPA is a continuing measure of the flow of drugs from retail pharmacies to consumers via written or telephoned prescriptions. Thus, the basic data underlying the survey are physicians' formal prescriptions (CX 838A, B).

708. The NDTI, unlike the NPA, is designed to measure the variety of ways a doctor might "recommend" a nonprescription internal analgesic. It includes prescription activity, but also includes, *inter alia,* drug issuance by the physician in a hospital, and drug recommendations made by the physician but not formally prescribed (CX

838G). Furthermore, doctors' recommendations and issuance activity are categorized in NDTI reports in terms of "desired actions" (*e.g.*, "pain relief," "antiarthritic"), allowing a more specific determination of whether doctors recommend Bufferin most for relief of "minor pain" (CX 838H-I; Complaint ¶ 17; Ross, Tr. 7378–79). The data for the period covered by the submitted portions of NDTI (October 1967 through June 1971) support the following conclusions:

(a) First, construing "doctors' recommendations" broadly, to include all drug issuance activity by doctors *without* regard to the purposes for which the drugs were issued; as projected by NDTI, the total number of times Tylenol was recommended by doctors far exceeds the total number of times Bufferin was recommended. For example, according to the NDTI projection for the period July, 1970 to June, 1971, Bufferin was issued a total of 758,000 times, while Tylenol was issued 1,774,000 times (Ross, Tr. 7380–81; CX 822Z).

(b) Second, construing "doctors' recommendations" more narrowly, and focusing on the issuance activity by doctors for "desired actions" relating solely to pain relief (the combined categories of "pain relief," "analgesic," [181] "analgesic and pain relief," and "relieve headache"/"relieve headache and antipyretic") the number of times doctors recommended Tylenol again far exceeds Bufferin, as does the number of times doctors recommended Ascriptin or generic aspirin (Ross, Tr. 7381; CX 822Z). For example, for the period July, 1970 to June, 1971, Bufferin was issued 404,000 times for these pain-related "desired actions," while issuances totalled 1,030,000 for Tylenol; 655,000 for Ascriptin; and 5,436,000 for generic aspirin (Ross, Tr. 7381; CX 822Z).

(c) Third, for each individual "desired action" relating to pain ("pain relief," "analgesic," or "relieve headache/antipyretic") Tylenol issuances exceeded Bufferin issuances, as did issuances for generic aspirin (Ross, Tr. 7381; CX 822Y).

709. By any reasonable analysis of the 1971 NDTI, Tylenol and generic aspirin, and often Ascriptin, were "recommended" more often than Bufferin, and thus the claim alleged in Paragraph 17 could not reasonably be based on the NDTI data (Ross, Tr. 7382; CX 822Y, Z).

710. For reasons discussed hereinabove (F. 706–09, *supra*) respondents did not have a reasonable basis for the representation that physicians recommend Bufferin more than any other nonprescription internal analgesic at the time such claims were made (Complaint ¶ 18).

2. Superiority Claims for Excedrin P.M.

711. Respondents represented that it has been established that Excedrin P.M. relieves more pain than aspirin; that it is more effective for nighttime pain than aspirin; and that it is more effective than aspirin because it contains three ingredients (F. 357, *supra*).

712. Bristol-Myers has not sponsored or funded any studies in which subjects with slight to severe pain ingested Excedrin P.M., and which (1) compared the analgesic effects of Excedrin P.M. to the analgesic effects of aspirin; or which (2) compared the sedative, hypnotic or somnifacient effects of Excedrin P.M. to aspirin; or which (3) determined or evaluated the amount or degree of analgesic effects upon subjects who ingested Excedrin P.M. (CX 925J; CX 927H; CX 929D).

713. In order to establish the comparative efficacy of analgesics, well-controlled clinical tests are prerequisite (F. 364–94, *supra*). Bristol-Myers did not generate the only type [182] and quality of evidence which could establish its claims relating to Excedrin P.M.'s superiority. Given that well-controlled clinical studies comparing Excedrin P.M. to aspirin do not exist, it has not been established that Excedrin P.M. is superior to aspirin, as alleged in Complaint Paragraphs 7B(9), (10) and 8.

714. Respondent has also represented that Excedrin P.M. contains a special ingredient, unique to its formulation (Complaint ¶ 23). In fact, the special ingredient is methapyrilene fumarate, an antihistamine available in other OTC medications such as Cope (Complaint ¶ 24; Answer of Bristol-Myers, Paragraph 1). By its admission that methapyrilene is available in other preparations besides Excedrin P.M., Bristol-Myers admitted that the uniqueness representation challenged in Paragraph 23 was false.

3. Claims Concerning the Ingredient in Bufferin and Excedrin

715. Respondent represented that the analgesic ingredient in Bufferin is other than ordinary aspirin (Complaint ¶ 21; F. 258, *supra*) when, in fact, aspirin is the only analgesic ingredient in Bufferin. Therefore, this representation was false (Complaint ¶ 22).

716. Respondent represented that the ingredient giving "long lasting relief" in certain Excedrin advertisements is something other than aspirin and the "antidepressant" is something other than caffeine (Complaint ¶ 21; F. 258, *supra*), when, in fact, the "long lasting relief" ingredient is aspirin (Lanman, Tr. 12150–51), and the "antidepressant" is caffeine (Lanman, Tr. 12150). Therefore, these representations were false (Complaint ¶ 22).

4. Substantial Question

717. Because the superiority claims for Excedrin P.M. have not been scientifically established, as alleged in Complaint Paragraph 7B(9)–(10), according to criteria accepted by experts in the relevant scientific community, those claims were made in the presence of a substantial question among such experts as to their validity, as alleged in Complaint Paragraphs 9B(9)–(10) and 10.

VI. CONSUMER IMAGES OF BUFFERIN AND EXCEDRIN

A. *Introduction*

718. From common sense and daily experience, the Bufferin and Excedrin advertising claims discussed in the preceeding sections and repeated during a long period of time, can reasonably be expected to create and maintain a product image or [183] belief in the consumer's mind reflecting the advertising claims. The various surveys conducted by or for Bristol-Myers and other leading manufacturers of OTC analgesics confirm that conclusion.

719. An advertising penetration study is a survey which assesses the consumer's awareness of advertising claims disseminated and the extent to which these advertising themes have penetrated and remained in the consumer's mind. Unlike copy tests, which focus attention on consumer's short-term recall of advertisements to which they have been exposed, usually within 24 hours, penetration studies have a much longer time reference; they measure consumers' recall of both the fact of advertising and its content over a period of time that generally exceeds several months (Ross, Tr. 7159–60). The "fact" of advertising refers solely to having seen any advertising for the brand. The "content" refers to the substance or specific themes of that advertising. Methods employed in penetration studies are similar to copy tests in that they both generally pose open-ended questions to the respondents with respect to their recall of the fact and the content of the advertising. This open-ended questioning calls for top-of-mind recollection of the advertising; *i.e.,* independent recall about the advertising with no probing for specific content by the interviewer. Consequently, penetration studies represent a lower bound estimate of the nature and amount of consumer's recall of advertising claims and themes (Ross, Tr. 7161). If a structured, closed-ended question were put to respondents testing the presence or absence of recall of a particular theme or content in an ad, the percentage of recall would be substantially higher (Ross, Tr. 7161). Also, since penetration studies reflect a longer period of time than copy tests, there is obviously a greater lapse between the time the consumer is exposed to the advertisements and the time the consumer is asked to recall them;

accordingly, one reasonably would expect the level of response to be lower in the penetration study than in the copy tests (Ross, Tr. 7161).

B. *Consumer Recall of Product Claims*

720. Four advertising penetration studies in evidence, CX 310, CX 347, CX 326, and CX 345, contain questions relating to levels of both Bufferin and Excedrin advertising penetration in 1969, 1970, 1971, and 1973. The surveys first asked respondents whether they recalled any advertising for Bufferin and Excedrin (F. 723, *infra;* Tables I, II), and then whether they recalled any specific claims being made for the product (F. 724, *infra,* Tables III, IV). These open-ended questions are lower estimates of recall of the advertising, and do not reflect the maximum number of people who had recall of specific claims. Despite the conservatism in the data produced by open-ended questions, the results of these penetration studies demonstrate that (1) a substantial number of consumers remembered that [184] Bufferin's advertising contained comparative speed and gentleness claims; and (2) a substantial number of consumers remembered that Excedrin's advertising contained comparative superiority claims and tension relieving claims (Ross, Tr. 7163; F. 724, *infra*).

721. Evidence from CX 310, *The 1969 Excedrin Study,* commissioned by Bristol-Myers, confirms that Bufferin's and Excedrin's comparative speed claims were remembered by consumers. This study is the only one of the four penetration studies that contained a closed-ended, or aided, recall question; its results show that consumers accurately remembered the advertising for the brands (Ross, Tr. 7178, 7198). For Bufferin, the study demonstrates that 39% of the total sample associated the claim "Goes to work in half the time" with Bufferin (Ross, Tr. 7178; CX 310Z095). With respect to Excedrin, 57% of the total sample correctly associated the claim "extra-strength pain reliever," and 44% associated the claim "For __ Headache No. 1040," with Excedrin (Ross, Tr. 7198–99). Consumers' distinctive and accurate attributions of these claims to Bufferin and Excedrin, coupled with consumers' correct attributions of other claims to other competing brands, demonstrates that consumers' answers to questions about what advertising they recall are not random comminglings of claims for different products (Ross, Tr. 7178). Rather, consumers are demonstrating that they can correctly recall advertising for the brand about which they are thinking and that they associate the central claims made for Bufferin or Excedrin with each brand (Ross, Tr. 7178, 7198). The magnitude of correct responses to this closed-ended question supports the view that the generally lower percentage responses associated with open-end-penetration questions are underestimates of the

actual registration of advertising in the minds of the public (Ross, Tr. 7161).

722. A meaningful analysis of penetration data to reflect the content of recall should be limited to those consumers who said they recalled advertising for the brand (Ross, Tr. 7171). Percentage of content recall based on the total sample would include those who remembered no advertising for the product at all (Ross, Tr. 7171). Accordingly, whenever a study presented recall data percentaged against the entire sample, Dr. Ross adjusted that percentage by limiting its base to those who recalled the fact of advertising. Through simple division, Dr. Ross produced the relevant figures for recall of Bufferin and Excedrin advertising themes which appear in Tables III and IV, *infra* (Ross, Tr. 7171).

723. The results from the four studies in Table I (Bufferin) and Table II (Excedrin) demonstrate that over a four-year period from 1969 to 1973, between 36.8% and 43% of the samples recalled seeing some Bufferin advertising (Ross, Tr. 7170, 7182, 7187, 7192–7193); and between 36.6% and 66% of the samples recalled seeing some Excedrin advertising (Ross, Tr. 7196, 7200–01, 7202, 7204). **[184a]**

Table I
Percent Of Total Respondents Who Recalled Any Advertising For Bufferin

1969[1]	1970[2]	1971[3]	1973[4]
43%	39.9%	37%	36.8% **[184b]**

[1] CX-310 Z-090, Z-146; Ross, Tr. 7169-7170.

What do you recall being said in any advertising [during the past six months] for Bufferin? What was the main idea that the advertiser was trying to get across?

[2] CX-348 Z; CX-347 Z-121; Ross, Tr. 7182-7185.

Do you recall seeing or hearing any advertising for Bufferin in the past four weeks?

[3] CX-3260, C; CX-1009A; Ross, Tr. 7186-7187.

What does any advertising you have recently seen or heard say about Bufferin?

[4] CX-345 Z-027,-031,-033,-107; Ross, Tr. 7191-7193.

Have you seen or heard any recent advertising for any headache remedies or pain relievers? For which products or brands? Do you remember hearing or seeing any recent advertising for Bufferin?

Table II
Percent Of Total Respondents Who Recalled Any Advertising For Excedrin

1969[1]	1970[2]	1971[3]	1973[4]
66%	48.8%	44%	36.6% **[185]**

[1] CX-310 Z-090, Z-146; Ross, Tr. 7196.

What do you recall being said in any advertising [during the past six months] for Excedrin? What was the main idea that the advertiser was trying to get across?

[2] CX-348 S; CX-347 Z-121; Ross, Tr. 7200-7201.

Do you recall seeing or hearing any advertising for Excedrin in the past four weeks?

[3] CX-326 Z, C; Ross, Tr. 7202.

What does any advertising you have recently seen or heard say about Excedrin?

[4] CX-345 Z-027,-031,-033,-107; Ross, Tr. 7204.

Have you seen or heard any recent advertising for any headache remedies or pain relievers? For which products or brands? Do you remember hearing or seeing any recent advertising for Excedrin?

724. Tables III and IV detail the portion of consumers who remembered comparative speed and gentleness claims in Bufferin advertising, and comparative speed, strength and effectiveness claims in Excedrin advertising, respectively. The data in Tables III and IV are derived from the three studies which inquired into the content of recall, and they are appropriately percentaged against the more meaningful bases of those respondents who recalled Bufferin's or Excedrin's advertising (Ross, Tr. 7171). In examining the extent to which these consumers remembered superior efficacy claims for Excedrin, recall of strength and effectiveness claims should be assessed as aspects of comparative superiority in pain relieving efficacy (Ross, Tr. 7196–98, 7202–03, 7204–05; F. 736, *infra*).

725. In assessing the magnitude of the top-of-mind, completely unaided speed and gentleness recall for Bufferin, and superior efficacy recall for Excedrin, the absolute size of the percentages is not nearly as important as their size relative to the recall of other types of claims (Ross, Tr. 7315–16; 7329; 7330; 7444–46; 7450–51). Significantly more consumers recalled comparative claims than recalled the simple fact that the product was a pain reliever or that it relieved headaches. For example, in CX 310 approximately 4% of consumers recalled Excedrin's advertising claiming it "relieves pain" (CX 310 Z091). In CX 326, 6% recalled that Bufferin relieves "headaches," and 4% recalled it relieves "pain" (CX 3260). Approximately 8% recalled Excedrin relieves "headaches," and 5% recalled it relieves "pain" (CX–326Z001). In CX 345, 4.3% recalled "relieves headaches," and 10.9% recalled "relieves pain" for Bufferin (CX 345Z057). For Excedrin, 8.2% recalled "relieves pain," and 7.1% recalled "relieves headaches" (CX 345Z066). The magnitude of recall of superior efficacy and tension relief claims shown in Tables III and IV should be judged against the context of low levels of recall for general claims (Ross, Tr. 7315–16; 7329; 7330; 7444–46; 7450–51).

726. The advertising penetration data in evidence demonstrates that substantial numbers of consumers remembered Bufferin's superior speed and superior gentleness claims and Excedrin's superior effectiveness and tension relief claims. More than one-third of the consumers interviewed could recall Bufferin advertising; likewise more than one-third interviewed recalled Excedrin advertising, off the tops of their heads on an unaided basis. Among those claims recalled, superiority in terms of speed and gentleness to the stomach were the dominant themes played back for Bufferin, and strength and effectiveness were the dominant themes played back for Excedrin (F. 723–25, *supra;* Ross, Tr. 7194–95; 7205). **[185a]**

Table III

Unaided Advertising Penetration

Based Upon Respondents Who Recalled Any Advertising For Bufferin

CX-310 (1969)[1]

Works Twice As Fast/Faster Acting 42%

Doesn't Effect/Upset Stomach 21%

It's Buffered/Contains Buffering Agents 9.3%

CX-326 (1971)[2]

Speed

Faster/Goes To Headache Faster/ twice as fast/in half the time 38%

Dissolves Faster 24%

Fast 8%

Stomach

Less Stomach Upset 22% 8.1%

No Stomach Upset 3% 16.2%

CX-345 (1973)[3]

Faster Acting 23.4%

Doesn't Upset Your Stomach 14.7%

Relieves Stomach Discomfort 2.2%

More Buffers, Buffered 4.9%

Tension/relieves tension 1.1% **[185b]**

[1] CX 310Z-094; Ross, Tr. 7171-7172.
[2] CX 3260, P, Q; Ross, Tr. 7187-7189.
[3] CX 3452-057, 058, 107; Ross, Tr. 7193-7194.

Table IV
Unaided Advertising Penetration
Based Upon Respondents Who Recalled Any Advertising For Excedrin

CX-310 (1969)[1]		CX-326 (1971)[2]			CX-345 (1973)[3]	
		Competitive Superiority		45.5%		
Extra-Strength Pain Reliever	4.5%	Stronger	20.5%		Stronger/Extra Strength	12.6%
Extra-Strength/Stronger	12.1%	More Relief	2.3%		More Effective Than Aspirin	4.4%
Relieves Headaches/All Kinds of Headaches	13%	Generally Better	9.1%			
Longer Lasting	3%	More Ingredients	4.5%			
Relieves Tension/Nerves/Tension Headaches	10.6%	Tension		2.3%	Relieves Tension	2.7% **[186]**

[1] CX-310Z091; Ross, Tr. 7196-7197.
[2] CX-326Z, Z-001; Ross, Tr. 7202-7203.
[3] CX-345Z066,-067,-107, Ross, Tr. 7204-7205.

C. *Consumers In Substantial Numbers Believe Bufferin and Excedrin Are More Effective Than Aspirin*

727. Six reliable market surveys in evidence, conducted during the period 1967 through 1975, demonstrate that a substantial number of consumers have believed and continue to believe that Bufferin is faster and gentler than aspirin and Excedrin is a more effective pain reliever than aspirin.

728. In a category of products such as OTC analgesic drugs, when consumers believe that the attributes of a particular OTC analgesic make it more efficacious than another product, they also believe that the superiority of that product on those attributes has been supported by adequate scientific evidence (Ross, Tr. 7055). As Dr. Ross testified:

> [It's] reasonable in my judgment for consumers in not insignificant numbers to believe you must have such evidence lurking around or being the basis for such claims or you won't be allowed to make them (Ross, Tr. 7053).

729. The fact that typical marketing research, such as the surveys in evidence in this record, does not ordinarily report the nature or adequacy of scientific support underlying consumer beliefs about the attributes of products does not undermine Dr. Ross' view that consumer beliefs include a component relating to the adequacy of scientific support. Consumer research is not structured to pick up the existence of such a belief, nor do consumers ordinarily express the fact that there is underlying scientific support which led them to hold a belief (Ross, Tr. 7054–55).

730. Thus, despite the absence of explicit survey evidence, it is reasonable to infer, from survey evidence showing that consumers do believe in the claims that Bufferin is faster or gentler, or that Excedrin is stronger or more effective, that they also believe that there is adequate scientific support for these comparative claims.

1. Evidence From Commercial Market Research Conducted in 1967 and 1970

731. An image study tests the competitive position of one product versus another by measuring images or beliefs, and the extent of those beliefs, about a product (Ross, Tr. 7224–25). These images or beliefs can be analyzed first by identifying those attributes which are pertinent to consumers' perceptions of the product and its benefits, and second by measuring the extent to which consumers believe those attributes are relevant to their purchasing decisions (Ross, Tr. 7224–25). One cannot learn the nature of consumer beliefs about a product from their purchases alone. In order to learn the nature of a consumer's beliefs about a product and its attributes, one has to ask the [187]

consumer for a descriptive statement about those beliefs (Ross, Tr. 7226). The five studies (CX 1058, CX 346, CX 310, CX 1059, and CX 347–348), which were conducted at different times between 1967 and 1970 by different research organizations, for different OTC analgesics manufacturers (before the FTC Analgesics Complaints were issued in 1972), using different methodologies and different samples, provide relevant information for coming to a conclusion about the comparative images of both Bufferin and Excedrin relative to aspirin (Ross, Tr. 7229). All five studies are typical of the kinds of studies conducted in market research, and are of greater scope and higher reliability than many studies on attitude and image research that are used as a basis for marketing decisions by business firms (Ross, Tr. 7229–30).

732. Due to the fact that these five studies focus on major branded analgesics and not unbranded "aspirin," the only way to assess consumers' beliefs about comparative effectiveness of Bufferin, Excedrin and aspirin is to use a surrogate for "aspirin": Bayer (Ross, Tr. 7240–41). This method injects a bias which tends to diminish differences in consumer beliefs about the branded aspirin products (Ross, Tr. 7401–02). This bias results from the fact that Bayer is both a well known, heavily advertised, widely used analgesic (Ross, Tr. 7241).

733. The five studies conducted between 1967 and 1970 report the results based upon the entire sample surveyed and upon the users of each brand. Three of these studies (CX 346, CX 310, CX 347–348) also permit analyses of respondents who do not use, or do not regularly use, Bufferin, Excedrin, or Bayer (Ross, Tr. 7231–32).

734. An analysis that separately compares users' beliefs and nonusers' beliefs is preferable to an analysis that simply compares the beliefs of all respondents who gave their opinions about the efficacy of the products, regardless of their usage patterns (Ross, Tr. 7237). Preference for a "user v. user" or "non-user v. non-user" analysis is based upon the fact that the relative rather than the absolute beliefs and images are the subjects of concern in this proceeding. An analysis based on the results of the total sample builds in a bias that obscures relative beliefs and images (Ross, Tr. 7233–38; CX 822A).

735. The bias built in a total sample analysis is a consequence of the well recognized phenomenon that users of a product are apt to rate it more favorably than do nonusers (Ross, Tr. 7233). This bias, called user bias or user "halo," disproportionately favors Bayer, the brand that was used more often than Bufferin or Excedrin by the total population at the time the studies were done. Since there were many more Bayer users in the total sample of consumers surveyed than there were users of the challenged brands, the percentage of the total sample that said favorable things about Bayer can be expected to [188] be disproportionately higher (Ross, Tr. 7401–02). This dispropor-

tionate usage of Bayer resulted in more frequent favorable ratings of Bayer by the total sample, and it obscured the relative beliefs—the true differences in beliefs—about Bufferin or Excedrin and Bayer (Ross, Tr. 7233, 7401–02). Separate analysis of relative beliefs among users of these products, and among nonusers of these products, balances the effects of Bayer users' favorable ratings of their product (Ross, Tr. 7234–7238). This technique is frequently used to hold constant the effects of differential product usage within a sample on the relative images of two brands (Ross, Tr. 7237–38; 7243).

736. None of the commercial image studies explicitly questioned consumers about the general pain relieving "efficacy" of the analgesics studied. However, the attribute of strength has been shown to have a strong, logical relationship to a pain reliever's "effectiveness" (Ross, Tr. 7056–59).

736a. Tables V and VI compare users' beliefs of the product attributes "speed" and "gentleness," respectively, for Bufferin and Bayer. Table VII displays a similar user belief comparison for Excedrin and Bayer with respect to a number of strength- and efficacy-related attributes. All three tables were derived from the five commercial image studies in evidence that were conducted during the period 1967–1970. They show that, during that period, a significantly greater portion of Bufferin users believed Bufferin was fast/gentle than Bayer users believed Bayer was, and a significantly greater portion of Excedrin users believed Excedrin was stronger or more effective than Bayer users believed Bayer was.

737. Results from three of the five studies done between 1967 and 1970 (CX 346, CX 310, CX 347–348) were also analyzed from the point of view of respondents who were not current users or current "most often" users of a brand. These results are shown in Tables VIII, IX, and X, *infra*. This "non-users" analysis was another effort to remove to the extent possible, the favorable bias that affects the ratings of all brands by virtue of the fact that those who rate them are also users of them. Analysis of beliefs and images among "non-users" removes this bias by actually removing the favorably biased users' ratings from the analysis. This contrasts with the "user v. user" analysis discussed in F. 735–736a, *supra*, which holds the bias constant by limiting the analysis to users' ratings (Ross, Tr. 7238).

738. Another advantage of the analysis of comparative beliefs and images among nonusers is that it more directly addresses the role of advertising as a source of the beliefs and images analyzed. By definition, the opportunity for usage or prior experience, to contribute to the comparative images of "non-users" is diminished or eliminated (Ross, Tr. 7238). As with the results of the user analyses presented in Tables V, VI, [188a]

Table V
Beliefs About Bufferin And Bayer
Percentages Based Upon Users Of Each Product
Speed

1967 CX-1058[1]	1967 CX-346[2]	1969 CX-310[3]	1970 CX-1059[4]	1970 CX-347/348[5]
Buf./Bayer	Buf./Bayer	Buf./Bayer	Buf./Bayer	Buf./Bayer
For Fast Relief	Relieves Pain Most Quickly	Speedy	For Fast Relief	Gives Fast Acting Relief
59.2/55.2%(N.S.)	73/65%(N.S.)	77/67%***	67.7/55.6%*	68.4/65.2%(N.S.) **[188b]**

[1] CX-1058Z486; CX-807Z101; Ross, Tr. 7269-7270; CX-822L. Dr. Ross' analysis was based upon the responses of males and females. The percentages displayed above simply combine those responses into a total percentage for Bufferin and Bayer users.
[2] CX-346Z150,-151; Ross, Tr. 7245, CX-822B.
[3] CX-310Z148,-071,-072; Ross, Tr. 7267, CX-822K.
[4] CX-1059Z233,-257; CX- 807Z101; Ross, Tr. 7269-7270, CX-822L. *See* Footnote[1] supra regarding the meanings of these percentages.
[5] CX-348Z225,-229 (CX-347Z126); Ross, Tr. 7260, CX-822H.

N.S. = Sig. < .05
* = Sig. > .05
** = Sig. > .01
*** = Sig. > .001

Table VI
Beliefs About Bufferin And Bayer
Percentages Based Upon Users Of Each Product
Gentleness

1967 CX-1058[1]	1967 CX-346[2]	1969 CX-310	1970 CX-1059[3]	1970 CX-347/348[4]
Buf./Bayer	Buf./Bayer	Buf./Bayer	Buf./Bayer	Buf./Bayer
Doesn't Upset The Stomach	Never Upsets Stomach	N.A.	Doesn't Upset The Stomach	Never Upsets Stomach
76/66.5%*	67/68% (N.S.)		81.5/68.4%**	59.6/32%***
Irritates The Stomach			Irritates The Stomach	
4.6/9.3% (N.S.)			1.6/13.5%*** **[188c]**	

[1] CX-1058Z481,-491; CX-807Z101; Ross, Tr. 7269-7270, CX-822L. Dr. Ross' analysis was based upon the responses of males and females. The percentages displayed above simply combine those responses into a total percentage for Bufferin and Bayer users.

[2] CX-346Z150,-151; Ross, Tr. 7245, CX-822B.

[3] CX-1059Z226,-250, (CX-807Z101); Ross, Tr. 7269-7270, CX-822L. *See* Footnote[1] supra regarding the meanings of these percentages.

[4] CX-348Z225,-229 (CX-347Z126); Ross, Tr. 7260, CX-822H.

N.S. = Sig. < .05
* = Sig. > .05
** = Sig. > .01
*** = Sig. > .001

Table VII
Beliefs About Excedrin and Bayer
Percentages Based Upon Users Of Each Product

1967 CX-1058[1]

	Exc./Bayer
For Fast Relief	70.1/55.2%**
Gives Long Lasting Relief	74.9/53.1%***
Good For Severe Headaches	75.9/47.9%***
Strong	69.5/19.1%***
Gives Complete Relief	69.5/51.0%***

1967 CX-346[2]

	Exc./Bayer
Relieves Pain Most Quickly	88/65%***
Relieves Pain For A Long Period	71/44%***
Good For Severe Headaches	77/52%***
Very Strong Product	75/37%***

1969 CX-310[3]

	Exc./Bayer
Speedy	85/67% (N.S.)
Long Lasting	62/33% (N.S.)
Mainly For Severe Headaches	35/7% (N.S.)
Too Strong	3/1% (N.S.)
Complete Relief	49/31% (N.S.)

1970 CX-1059[4]

	Exc./Bayer
For Fast Relief	78.9/55.6%***
Gives Long Lasting Relief	76.7/55.6%***
Good For Severe Headaches	72.2/48.1%***
Extra Strength	82.7/19.5%***
Strong	73.7/27.1%***
Gives Complete Relief	69.9/57.1%* **[188d]**

1970 CX-347/348[5]

	Exc./Bayer
Gives Fast Acting Relief	78.7/65.2% (N.S.)
Gives Longer Lasting Relief	60.0/35.2%***
Good For Severe Headaches	69.0/39.9%***
Is Extra Strength	64.5/12.3%***

[1] CX-1058 Z478,-480,-485,-486,-490,-491; Ross, Tr. 7430-7431, CX-823N.
[2] CX-346 Z152,-153; CX-823A.
[3] CX-310 Z148,-071,-072; Ross, Tr. 7427-7430, CX-823M.
[4] CX-1059 Z226,-229,-233,-238,-241,-245 (CX-344 Z101); Ross, Tr. 7430-7432, CX-823N.
[5] CX-347 Z126, CX-348 Z225,-226,-233,-244; Ross, Tr. 7410-7411, CX-823H.

N.S. = Sig. < .05
* = Sig. > .05
** = Sig. > .01
*** = Sig. > .001

Table VIII
Beliefs About Bufferin And Bayer
Percentages Based Upon Non-Users Of Each Product
Speed

1967 CX-346[1]	1969 CX-310[2]	1970 CX-347/348[3]
Buf./Bayer	Buf./Bayer	Buf./Bayer
Relieves Pain Most Quickly	Speedy	Gives Fast Acting Relief
29/26% (N.S.)	61/48%***	42.6/41.1% (N.S.) **[188e]**

[1] CX-346Z059,-060,-150,-151; Ross, Tr. 7249, CX-822C.
[2] CX-310Z148,-071,-072; Ross, Tr. 7267, CX-822K. Percentages are based on beliefs about Bufferin by Bayer users, and vice versa.
[3] CX-348Z225,-229 (CX-347Z126); Ross, Tr. 7260, CX-822H.
N.S. = Sig. < .05
* = Sig. > .05
** = Sig. > .01
*** = Sig. > .001

Table IX

Beliefs About Bufferin And Bayer

Percentages Based Upon Non-Users Of Each Product

Gentleness

1967 CX-346[1]

Buf./Bayer

Never Upsets Stomach

39/43% (N.S.)

1969 CX-310

Buf./Bayer

N.A.

1970 CX-347/348[2]

Buf./Bayer

Never Upsets Stomach

35.3/16.5%*** **[188f]**

[1] CX-3467059,-060,-150,-151; Ross, Tr. 7249, CX-822C.

[2] CX-3487225,-229 (CX-3477126); Ross, Tr. 7260, CX-822H.

N.S. = Sig. < .05

* = Sig. > .05

** = Sig. > .01

*** = Sig. > .001

Table X
Beliefs About Excedrin And Bayer
Percentages Based Upon Non-Users Of Each Product

1967 CX-346[1]

Exc./Bayer

Relieves Pain Most Quickly
25/26% (N.S.)

Relieves Pain For A Long Period
21/21% (N.S.)

Good For Severe Headaches
38/29%**

Very Strong Product
33/15%***

1969 CX-310[2]

Exc./Bayer

Speedy
61/36%*

Long Lasting
43/23% (N.S.)

Mainly For Severe Headaches
39/4% (N.S.)

Too Strong
25/2% (N.S.)

1970 CX-347/348[3]

Exc./Bayer

Gives Fast Acting Relief
46.0/41.1%*

Gives Longer Lasting Relief
27.2/17.9%***

Good For Severe Headaches
38.4/25.6%***

Is Extra Strength
40.3/11.6%*** **[189]**

---

[1] CX-346Z060,-061,-062,-066; Ross, Tr. 7404-7405, CX-823B.
[2] CX-310Z148,-071,-072; Ross, Tr. 7424, CX-823M. Percentages are based on beliefs about Excedrin by Bayer users, and vice versa.
[3] CX-347Z126, CX-348Z225,-226,-233,-244; Ross, Tr. 7140-7411, CX-823H.

N.S. = Sig. < .05
* = Sig. > .05
** = Sig. > .01
*** = Sig. > .001

and VII, "non-users" of Bufferin and Bayer believe Bufferin superior in speed and gentleness to Bayer; Excedrin and Bayer nonusers believe Excedrin superior in strength and pain relieving efficacy to Bayer.

739. CX–346, the 1967 *Assets & Liabilities Study,* is the only one of the five 1967–1970 studies which permits a comparison of both Bufferin's image and Excedrin's image with that of an "aspirin" product other than Bayer. While Bayer ratings were also included in the study and analyzed (Tables V through X), respondents were asked to rate Norwich and "store's own brand" as well on the same dimensions as Bufferin and Excedrin. These comparisons further confirm the superior speed and gentleness image of Bufferin and the superior strength and effectiveness image of Excedrin (F. 740, *infra;* Tables XI, XII; Ross, Tr. 7252–53, 7404–05).

740. An analysis of the nonexclusive users of Bufferin, Excedrin, Norwich, and store's own brand aspirin in the 1967 *Assets and Liabilities Study* (CX 346) demonstrates that Bufferin's image is superior to Norwich's and store brand's images on the relevant attributes speed and gentleness and Excedrin's image is superior to Norwich and store's own brand on speed, strength and severe headache (Ross, Tr. 7250–53, 7404–05, Tables XI, XII).

741. Additional data from the *1969 Excedrin Study* (CX 310) provide "reasons for using" Bufferin, and consequently support the results relating to beliefs about pertinent attributes of the product (Tables V, VI, IX). First, the most important performance element Bufferin users gave as their reason for initial trial of the product was "safety" (19%, including references to stomach upset) (CX 310Z060). Further, of all of the reasons stated for switching to Bufferin, the three that stood out most were "no upset stomach" (23%), "saw/heard advertising" (17%), and "faster acting" (13%) (Ross, Tr. 7264–66; CX 310Z067–068; CX 822J). These respondents believed Bufferin to be a speedier product that was gentler to the stomach than the brand they previously used (Ross, Tr. 7264–65), and a substantial portion attributed their reason for switching to their brand to the images they formed from the "advertising" they saw for Bufferin (Ross, Tr. 7265–66).

742. The *1970 Vanquish Study* (CX 347–348) reports additional data to support the fact that people used their particular brands to obtain benefits that were consistent with the benefits they sought from a headache remedy in general (Ross, Tr. 7258–59). That is, Bufferin users, to a significant degree more than Bayer users, believed that "contains buffers" (60.3%/11.9%) and "doesn't upset your stomach" (84.6%/75.9%) were reasons for using their brand; and Bufferin users, to a significant degree more than Bayer users (32%/2.4%), stated "doesn't upset stomach/buffers" was a reason for using their regular brand **[189a]**

Table XI
Beliefs About Bufferin And Aspirin
Percentages Based Upon Non-Exclusive Users Of Each Brand

1967 CX-346[1]
Buf./Norwich/Store
Relieves Pain Most Quickly
15/5/1%
Never Upsets Stomach
23/12/11% **[189b]**

[1] CX-346Z059,-060; Ross, Tr. 7252-7253; CX-822E.

Table XII
Beliefs About Excedrin And Aspirin
Percentages Based Upon Non-Exclusive Users of Each Brand

1967 CX-346[1]
Exc./Norwich/Store
Relieves Pain Most Quickly
21/5/1%
Relieves Pain For A Long Period
16/2/-1%
Very Strong Product
27/2/-3%
Good For All Kinds Of Pain
16/11/7%
Good For Severe Headaches
31/6/3% **[190]**

[1] CX-346Z060,-061,-062,-064,-066; Ross, Tr. 7404-7405; CX-823C, D.

specifically "most often for headaches" (Ross, Tr. 7258–59). These data demonstrate that Bufferin users chose their brand because they believed it was gentler to the stomach than aspirin, *i.e.,* it would prevent or diminish stomach upset (Ross, Tr. 7258–7259). This study also shows that Bufferin users more often than Bayer users believe that "headache remedies and pain relievers work too slowly," indicating that users of Bufferin are convinced that some brands work faster than others, and that this is their reason for choosing Bufferin over other brands.

743. Results of the 1969 *Excedrin Study* (CX 310) demonstrate that Excedrin users use their brand for the precise attributes which it advertises: comparative strength and effectiveness. Of all the reasons cited for switching to Excedrin, "faster acting" (37%) and "stronger, more powerful" (19%) were the most frequent reasons given by Excedrin users (Ross, Tr. 7417–18; CX 823K). Moreover, a substantial portion of Excedrin users (16%) listed "saw/heard advertising" as their reason for switching to Excedrin, and 44% cited advertising as their reason for coming to Excedrin in the first place. Excedrin users claim to suffer from "severe headaches" far more than Bayer users do

(50%/24%). This represents the largest difference for ailments treated by users of the two brands (Ross, Tr. 7420–23; CX 823L). Excedrin users, more than Bayer users (93%/59%), chose their own brand for treatment of "severe headaches"; and Bayer users were far more inclined to use Excedrin for "severe headaches," than were Excedrin users inclined to take Bayer (13%/1%) (Ross, Tr. 7424, CX 823L). These data show that both Excedrin and Bayer users believe Excedrin is a stronger pain reliever than aspirin, and that Excedrin users suffer from, and use their brand to relieve, specifically those ailments for which it advertises relief (Ross, Tr. 7420–24). Further, Excedrin users claim to suffer from ailments which reflect greater pain than do Bayer users, *e.g.,* twice as many Excedrin users than Bayer users felt that their "headaches are more severe than other people's headaches"; and six times as many believed, "Headache remedies and pain relievers don't work for me unless they are extra strong." Accordingly, Excedrin users think their brand contains what they want in an analgestic—extra strength (Ross, Tr. 7420–23).

744. The *1970 Vanquish Study* (CX 347–348) demonstrates further that Excedrin users use their brand for the attributes it advertises. The most frequent reasons why Excedrin users used their brand most often for headaches were "gives fast/quick relief" (28.8%), "works faster than others" (18.2%) (Ross, Tr. 7407–08; CX 823F). Of the reasons for using "headache remedies," those which stood out the most for Excedrin users were "provides quick relief" (90.3%), "be extra strong" (57.4%), "stronger than plain aspirin" (64.5%), and "provides long lasting relief" (85.2%) (Ross, Tr. 7408–10; CX 823F). Other [191] opinions reveal that Excedrin users, more than Bayer users, believe some brands work faster than others, and other brands work too slowly (Ross, Tr. 7412–13; CX 8231). These data indicate again that Excedrin users want and believe their brand has superior speed, strength and effectiveness, over other brands (Ross, Tr. 7414–17).

2. Evidence Of Current Consumer Beliefs In Bufferin's And Excedrin's Superiority Is Supplied By The Leavitt Study, CX 349

745. The *Leavitt Study* was an adequately designed, carefully administered consumer study performed for the Federal Trade Commission's staff by Dr. Clark Leavitt and the Gallup Organization. The study was a telephone survey, which employed well-controlled, reasonably randomized procedures to contact 780 consumers, who were asked to rate the pain relieving efficacy, speed, strength and gentleness of aspirin, Anacin, Bufferin and Excedrin.

746. Approximately 98% of the 780 respondents interviewed by Gallup Organization had heard of all of the four products surveyed. Dr. Leavitt did not analyze data from the 17 respondents, or 2% who

had not heard of all of the survey products (Leavitt, Tr. 6199). This was a reasonable approach (Leavitt, Tr. 6191–95). The exclusion of these seventeen (17) respondents did not impact upon the reliability of Dr. Leavitt's analysis because, in fact, analyses of results based upon all 780 interviews produced results virtually identical to those obtained when the 17 were excluded (F. 756, *infra*).

747. Whenever a respondent was unwilling or unable to rate a production the four-point scale presented to him in Questions 2 through 5, the interviewer was instructed to code "Don't Know" on the questionnaire (Leavitt, Tr. 6185; CX 349W). The pretesting of the questionnaire had disclosed that some respondents might be unwilling to rate a product because they did not personally use it (Crespi, Tr. 2270), and the questionnaire was modified to address this possibility by changing the preamble to Questions 2 through 5 to begin "Whether or not you have ever used them. . . ." During the actual interviews, respondents' reasons for not rating a product were not sought out by the interviewers, who had been instructed not to deviate from or to explain the wording on the questionnaire—only to repeat it (F. 165, *supra*).

748. Comparing a consumer's images about different products is an acceptable alternative to eliciting a statement of his own comparative images of these products. The former approach has an important advantage in that it permits an analysis of the degree or intensity of the comparative perceptions underlying a direct comparative statement. The Leavitt Study adopted the former approach and differs from the five commercial marketing studies in evidence in this respect. [192]

749. Comparative beliefs in the Leavitt Study are assessed by confining analysis to those respondents who expressed an opinion about both Bufferin or Excedrin and aspirin (Ross, Tr. 7279–80). This approach is based upon the view that a "Don't know" response about Bufferin or Excedrin, on the one hand, or about aspirin, on the other, reflects the lack of a basis for any comparative image concerning the three products (Ross, Tr. 7279–80). Therefore, exclusion of "Don't know" responses from an analysis of comparative images is appropriate for two reasons: (1) a "Don't know" response, by definition, is a lack of opinion; and (2) it is virtually impossible to position a "Don't know" response on the four point scale along with "extremely," "very," "fairly," and "not" (Ross, Tr. 7279–80). This was a reasonable approach.

750. While inclusion of "don't know" responses as part of an analysis of comparative images is not as meaningful, and may lead to erroneous conclusions, Dr. Ross analyzed the data from the Leavitt Study based on both the total sample, including "Don't knows," and

the subsample of respondents who rated both products (*i.e.*, excluding "Don't knows") (Ross, Tr. 7279). These dual analyses were performed to see if conclusions about comparative images differed depending upon the approach adopted (Ross, Tr. 7280).

751. When expressed as percentages based upon all 763 respondents analyzed by Dr. Leavitt in CX 349, the raw results are depicted in Tables XIII and XIV, *infra*. There are four "independent" percentages in each row of these tables, *i.e.*, the percentages in each row represent completely independent groups of respondents, and each response appears once, and only once, in each row (Crespi, Tr. 2352; Leavitt, Tr. 6203–04). These percentages are reasonably projectable to the population of adults who live in homes with telephones and who are aware of these products (Leavitt, Tr. 6193; 6246–47). At the 95% level of confidence, given a sample of approximately 750 people, the percentages could vary by approximately plus or minus 4% (Crespi, Tr. 2346–47). These results, generally speaking, show that approximately one out of every four Americans in telephone households who are aware of these products believes Bufferin is faster and gentler than aspirin and that Excedrin is more effective than aspirin.

752. Tables XV and XVI show the same comparative beliefs, but the percentages are based upon the subsample who rated both products as indicated. Regardless of the sample base used, Tables XIII–XVI clearly demonstrate that a significant number of consumers believed Bufferin was faster and gentler than aspirin, and Excedrin is faster, stronger, and more effective than aspirin (Ross, Tr. 7435–36; CX 822M; CX 823P). **[192a]**

Table XIII
Beliefs About Bufferin And Aspirin
Percentages Based Upon The Total Sample[1]

	Rated Both Products			Did Not Rate Both Products[2]	
	Rated Bufferin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Bufferin		Total
Speed	22.4%(171)	28.7%(219)	5.4%(41)	43.5%(332)	763 = 100%
Gentleness	24.9%(190)	28.1%(214)	7.2%(55)	39.8%(304)	763 = 100% **[192b]**

[1] CX-349Z018,-019; Ross, Tr. 7273-7275; CX-822M.
[2] Respondents were coded "Don't Know" for either Bufferin, aspirin or both.

Table XIV
Beliefs About Excedrin And Aspirin
Percentages Based Upon The Total Sample[1]

	Rated Both Products			Did Not Rate Both Products[2]	
	Rated Excedrin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Excedrin		Total
Effectiveness	22.1%(169)	25.0%(191)	3.7%(28)	49.2%(375)	763 = 100%
Speed	25.2%(192)	20.5%(156)	4.2%(32)	50.2%(383)	763 = 100%
Strength	23.6%(180)	22.5%(172)	2.8%(21)	51.1%(390)	763 = 100% **[192c]**

[1] CX-349Z015,-016,-017; Ross, Tr. 7435-7436; CX-823P.
[2] Respondents were coded "Don't Know" for either Excedrin, aspirin or both.

Table XV
Beliefs About Bufferin And Aspirin
Percentages Based Upon The Total Sample Who Rated Both Products[1]

	Rated Bufferin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Bufferin	Total
Speed	39.7%(171)	50.8%(219)	9.5%(41)	431 = 100%
Gentleness	41.4%(190)	46.6%(214)	12.0%(55)	459 = 100% **[192d]**

[1] Table XIII.

Table XVI

Beliefs About Excedrin And Aspirin

Percentages Based Upon The Total Sample

Who Rated Both Products[1]

	Rated Excedrin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Excedrin	Total
Effectiveness	43.6%(169)	49.2%(191)	7.2%(28)	388 = 100%
Speed	50.5%(192)	41.1%(156)	8.4%(32)	380 = 100%
Strength	48.3%(180)	46.1%(172)	5.6%(21)	373 = 100% **[193]**

[1] Table XIV.

753. Tables XVII and XVIII reflect the comparative images of Bufferin and aspirin, and Excedrin and aspirin, respectively, among the nonusers of Bufferin and Excedrin who rated both brands. Since a comparison of nonusers removes user bias with respect to Bufferin and Excedrin, the comparisons shown in Tables XVII and XVIII are more conservative than those shown in Tables XIII–XVI (Ross, Tr. 7274–75; 7435–36). In any event, Tables XVII and XVIII also demonstrate that a substantial number of consumers believed Bufferin is faster and gentler than aspirin and that Excedrin is superior to aspirin in terms of speed, strength, and effectiveness (Ross, Tr. 7273–78; 7435–36; CX 822M, CX 823P). *Cf.* Tables XIX–XXII, *infra.*

754. Tables XIX and XX reflect the comparative images of all those respondents who used *neither* Bufferin *nor* aspirin, or who used *neither* Excedrin *nor* aspirin. Tables XXI and XXII reflect the images among the same subsample, but percentages are based upon those who rated both products only (Ross, Tr. 7290–98; 7302; 7437; 7439–40). The analysis reflected in Tables XIX–XXII removes user bias completely because it removes aspirin usage as well as usage of Bufferin and Excedrin. Thus, their analysis reflects the prevalence and nature of comparative images among those persons who had images which, by definition, could not be affected by usage (F. 738, 753, *supra;* Ross, Tr. 7284–85). The results demonstrate that a significant number of this subsample of respondents believe that Bufferin is faster and gentler than aspirin, and that Excedrin is faster, stronger, and more effective than aspirin regardless of whether the percentage base includes the "Don't knows" (Ross, Tr. 7296, 7300–02, 7437, 7439–7400).

755. Finally, Tables XXIII through XXVI reflect the comparative images of nonusers of all four products surveyed in the Leavitt Study. This analysis is even more conservative in terms of eliminating all possible sources of user bias. The results in Tables XXIII–XXVI also demonstrate for both Bufferin and Excedrin that their superior image over aspirin persists in this most conservative analysis (Ross, Tr. 7298–7300, 7303, 7438, 7440).

756. As indicated in Table XXVII, the fact that Dr. Leavitt discarded data from seventeen (17) respondents who were not aware of all four products surveyed has no impact upon the results of this study. A comparison of results based upon either the 763 respondents analyzed by Dr. Leavitt or all 780 respondents interviewed reveals virtually identical results.

757. The Leavitt Study, together with the five commercial image studies discussed in this section, provides convincing confirmatory evidence that a significant segment of the consuming public over the years has held the beliefs that Bufferin is faster and gentler than aspirin and Excedrin is faster and more effective than aspirin. **[193a]**

Table XVII
Beliefs About Bufferin And Aspirin
Percentages Based Upon Non-Users Of Bufferin
Who Rated Both Products [1]

	Rated Bufferin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Bufferin	Total
Speed	28.5%(63)	59.7%(132)	11.8%(26)	221 = 100%
Gentleness	29.3%(71)	55.8%(135)	14.9%(36)	242 = 100% **[193b]**

[1] CX-349Z018,-019; Ross, Tr. 7273-7278; CX-822M.

Table XVIII
Beliefs About Excedrin And Aspirin
Percentages Based Upon Non-Users of Excedrin
Who Rated Both Products[1]

	Rated Excedrin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Excedrin	Total
Effectiveness	28.1%(65)	61.5%(142)	10.4%(24)	231 = 100%
Speed	36.7%(80)	55.5%(121)	7.8%(17)	218 = 100%
Strength	39.1%(88)	49.3%(111)	11.6%(26)	225 = 100% **[193c]**

[1] CX-349Z015,-016,-017; Ross, Tr. 7435-7436; CX-823P.

Table XIX
Beliefs About Bufferin And Aspirin
Percentages Based Upon Non-Users Of Both Products[1]

	Rated Both Products			Did Not Rate Both Products	Total
	Rated Bufferin Higher Than Aspirin	Rated Bufferin Equal To Aspirin	Rated Aspirin Higher Than Bufferin		
Speed	16.9%(43)	26.4%(67)	1.6%(4)	55.1%(140)	254 = 100%
Gentleness	19.7%(50)	24%(61)	3.9%(10)	52.4%(133)	254 = 100% **[193d]**

[1] CX-349W,X (Qu. #2,3,9); Ross, Tr. 7290-7298; CX-8220,R.

Table XX
Beliefs About Excedrin And Aspirin
Percentages Based Upon Non-Users Of Both Products[1]

	Rated Both Products			Did Not Rate Both Products	
	Rated Excedrin Higher Than Aspirin	Rated Excedrin Equal To Aspirin	Rated Aspirin Higher Than Excedrin		Total
Effectiveness	13.7%(41)	23%(69)	1.3%(4)	62%(186)	300 = 100%
Speed	16.3%(49)	19.7%(59)	1%(3)	62.3%(187)	300 = 100%
Strength	14.3%(43)	21%(63)	.7%(2)	64%(192)	300 = 100% **[193e]**

[1] CX-349W,X (Qu. #3,4,5,9); Ross, Tr. 7437-7439-40; CX-823R,U,X.

Table XXI
Beliefs About Bufferin And Aspirin
Percentages Based On Non-Users Of Both Products
Who Rated Both Bufferin And Aspirin[1]

	Rated Bufferin Higher	Rated Both The Same	Rated Aspirin Higher	Total
Speed	37.8%(43)	58.8%(67)	3.5%(4)	114 = 100%
Gentleness	41.3%(50)	50.4%(61)	8.3%(10)	121 = 100% **[193f]**

[1] Table XIX.

Table XXII
Beliefs About Excedrin And Aspirin
Percentages Based On Non-Users Of Both Products
Who Rated Both Excedrin And Aspirin[1]

	Rated Excedrin Higher	Rated Both The Same	Rated Aspirin Higher	Total
Effectiveness	36%(41)	60.5%(69)	3.5%(4)	114 = 100%
Speed	44.1%(49)	53.2%(59)	2.7%(3)	111 = 100%
Strength	39.8%(43)	58.3%(63)	1.9%(2)	108 = 100% **[193g]**

[1] Table XX.

Table XXIII
Beliefs About Bufferin And Aspirin
Percentages Based Upon Non-Users Of Aspirin, Bufferin, Excedrin, Or Anacin[1]

	Rated Both Brands			Did Not Rate Both Brands	
	Rated Bufferin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Bufferin		Total
Speed	18.7%(27)	22.9%(33)	1.4%(2)	56.9%(82)	144 = 100%
Gentleness	23.6%(34)	18.8%(27)	2.8%(4)	54.9%(79)	144 = 100% **[193h]**

[1] CX-349W,X (Qu. #2,9); Ross, Tr. 7298-7300, 7303; CX-822P,S.

Table XXIV
Beliefs About Excedrin And Aspirin
Percentages Based Upon Non-Users Of Aspirin, Bufferin, Excedrin, Or Anacin[1]

	Rated Both Products			Did Not Rate Both Products	
	Rated Excedrin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Excedrin		Total
Effectiveness	14.6%(21)	18.1%(26)	.7%(1)	66.7%(96)	144 = 100%
Speed	11.1%(16)	22.2%(32)	1.4%(2)	65.3%(94)	144 = 100%
Strength	10.4%(15)	23.6%(34)	.7%(1)	65.3%(94)	144 = 100% **[193i]**

[1] CX-349W,X (Qu. #3,4,5,9); Ross, Tr. 7438-7440; CX-823S,V,Y.

Table XXV
Beliefs About Bufferin And Aspirin
Percentages Based Upon Non-Users Of Aspirin, Bufferin, Excedrin, Or Anacin Who Rated Both Bufferin And Aspirin[1]

	Rated Bufferin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Bufferin	Total
Speed	43.6%(27)	53.2%(33)	3.2%(2)	62 = 100%
Gentleness	29.1%(34)	3.4%(4)	67.5%(79)	117 = 100% **[193j]**

[1] Table XXIII.

Table XXVI
Beliefs About Excedrin And Aspirin
Percentages Based Upon Non-Users Of Aspirin, Bufferin, Excedrin, Or Aspirin Who Rated Both Excedrin and Aspirin[1]

	Rated Excedrin Higher Than Aspirin	Rated Both The Same	Rated Aspirin Higher Than Excedrin	Total
Effectiveness	43.8%(21)	54.2%(26)	2.1%(1)	48 = 100%
Speed	12.0%(16)	64.0%(32)	4.0%(2)	50 = 100%
Strength	41.7%(15)	94.4%(34)	2.8%(1)	36 = 100% **[193k]**

[1] Table XXIV.

Table XXVII
Comparison of Percentage Results Based Upon 763* and (780)** Respondents

Bufferin

	Bufferin Superior To Aspirin	Bufferin Equal To Aspirin	Aspirin Superior To Bufferin
Speed	22.4%(22.4%)	28.7%(29.1%)	5.4%(5.4%)
Gentleness	24.9%(24.9%)	28.1%(28.3%)	7.2%(7.2%)

Excedrin

	Excedrin Superior To Aspirin	Excedrin Equal To Aspirin	Aspirin Superior To Excedrin
Effectiveness	22.1%(22.1%)	25.0%(24.9%)	3.7%(3.9%)
Speed	25.2%(23.9%)	20.5%(20.4%)	4.2%(4.1%)
Strength	23.6%(23.6%)	22.5%(22.4%)	2.8%(2.7%) **[194]**

* Tables XIII, XIV.
** Ross, Tr. 7273-7275, 7439-7441; CX-822Q, T; CX-823T, W, Z.

D. *Respondent's Advertising Played A Substantial Role In Creating And Reinforcing Consumers' Beliefs In Their Superiority Over Aspirin*

758. Several factors play a role in the creation and reinforcement of beliefs (often used interchangeably with "images") about products. Obviously, the most important factors are product usage and advertising (Ross, Tr. 7483–84, 7486). "Word of mouth," is also recognized as a source of product beliefs. However, "word of mouth" is a derivative factor: it derives from product usage and advertising (Ross, Tr. 7484).

759. The fundamental role of advertising is to call consumer's attention to the attributes of a product and to create favorable expectations about the performance of that product (Ross, Tr. 7486–87, 7496). For consumers who have already tried the product, advertising serves to reinforce those expectations by reminding consumers about the benefits of the product (Ross, Tr. 7487). Hence, advertising plays an important role in both the initial trial of a product and continued use of the product (Ross, Tr. 7487).

760. It is difficult to distinguish between the role of advertising, on the one hand, and product usage, on the other, in creating beliefs about products (Ross, Tr. 7488). The extent to which product usage will act as a distinct source of a belief about a product depends upon the difference between consumers' perception of product performance and their ability to "evaluate" product performance (Ross, Tr. 7488–89). The perception of product performance simply refers to a consumer's description of his or her own perception of the use experience. In contrast, the ability to "evaluate" refers to the ability to accurately measure or assess the true performance and differences between products (Ross, Tr. 7489–90).

761. Advertising is less important as a source of beliefs, and usage more important, in those cases where consumers' usage experience permits them to "evaluate" a product's performance. Usage in such situations provides the opportunity to confirm or disconfirm the expectations about product performance induced by advertising (Ross, Tr. 7493). A pocket calculator is such an example.

762. On the other hand, when consumer use does not permit "evaluation" of true product performance, consumer beliefs are, to a significant degree, the result of expectations induced by advertising. In such cases, usage experience does not provide the opportunity to confirm or disconfirm the expectations about product performance that advertising induces (Ross, Tr. 7494). A drug is a good example. The inability of consumers to evaluate the true pharmacological performance of a drug is supported by classical psychological research which shows that [**195**] user "perceptions" of the performance of

drugs are significantly influenced not by actual product performance but by extraneous information, such as advertising (Brock, Tr. 8557–58).

763. In the case of mild OTC analgesic products, such as Bufferin and Excedrin, product usage plays a minor role in creating product beliefs because consumer's ability to "evaluate" the pharmacological performance of the drug is affected by such factors as the placebo effect, the subjective nature of pain, and by the fact that each experience with pain is different. With respect to *comparative* product images of different OTC analgesics, usage is even less a factor as a source. In addition to the factors already named, consumers know the identities of the products they take for pain relief. Hence, their differential, advertising-induced expectations for each product's performance operate to influence their "perceptions" of these product's comparative performance. Consequently, consumers cannot "evaluate" the comparative performance of mild OTC analgesics on an unblinded basis (*i.e.,* when consumers know the products they are taking) (F. 399, *supra*).

764. Because consumers cannot "evaluate" the performance of OTC analgesics, their use experiences with the product cannot serve to disconfirm advertising-induced expectations about product performance (Brock, Tr. 8598–8602). This makes advertising more important a factor than usage as a source of product images regarding OTC analgesic products. Consumer research studies in evidence and the testimony of experts support this conclusion.

765. Furthermore, the market research in evidence shows that both users and nonusers hold essentially the same beliefs about the performance attributes of Bufferin and Excedrin (F. 736a–40, 752–55, *supra*). This absence of a difference in belief structure between users and nonusers is further support for the conclusion that the advertising for Bufferin and Excedrin has played a significant role in creating and reinforcing beliefs about those products.

766. Bufferin's advertised attributes of superior speed and gentleness are important to consumers who choose Bufferin (CX 347Z039; Brock, Tr. 8692). The themes of superior speed and gentleness compared to aspirin have been important aspects of Bufferin's advertising since at least 1960 (CX 816; CX 800). Likewise, Excedrin's advertised attribute of superior effectiveness is important to consumers who choose Excedrin (CX 347Z039; Brock, Tr. 8695), and this theme has been an important aspect of Excedrin advertising since it was introduced around 1960 (CX 818; CX 801).

767. From 1960 to 1973, Bristol-Myers spent over $171 million advertising Bufferin and over $98 million advertising **[196]** Excedrin (F. 5, *supra*). Advertisements disseminated during this period portrayed

Bufferin as a product that was faster and gentler than aspirin and Excedrin as a more effective pain reliever than aspirin (CX 816; CX 818; CX 800; CX 801). During this period Bufferin's advertising-to-sales ratio was about 30% and Excedrin's about 32% (CX 660; CX 661). There is evidence that advertisements representing Bufferin as faster than aspirin were disseminated between 1950 and 1976 at least 6,122 times on national and/or spot television programs and at least 28 times in magazines with national circulations (CX 816; CX 800). There is evidence that advertisements representing Bufferin as gentler than aspirin were disseminated between 1961 and 1976 at least 5,569 times on national and/or spot television programs and at least 11 times in magazines with national circulations (CX 816; CX 800). From 1960 to 1976 advertisements representing Excedrin as a more effective pain reliever than aspirin were disseminated at least 1,395 times on national and/or spot television programs and at least 116 times in magazines with national circulations (CX 818; CX 801H–801Z006).

768. The basic literature in both marketing and psychology shows that various well-known principles of persuasion can, if successfully used in communications, play a significant role in creating lasting beliefs, including beliefs about products (Brock, Tr. 8592–93). Dr. Brock, an expert in the applications of techniques of persuasion, analyzed a reasonably representative sample of Bufferin and Excedrin ads in evidence to ascertain the extent to which principles of persuasion were employed. He found a consistent and effective use of these techniques in them (Brock, Tr. 8593–96). Among the most prevalent techniques or principles known to be effective and used in the advertising of Bufferin and Excedrin are: (1) Linking a product with important human values. By linking the product with something important to the consumer (*e.g.,* relief from pain, maintenance of livelihood), the consumer is less likely to accept contrary information about the product; (2) The use of source credibility, such as using medical experts or studies to support medical claims, to enhance the believability of the message; and (3) Repetition (Brock, Tr. 8593–95).

769. In fact, in the sample of Bufferin and Excedrin ads which were analyzed in detail in his testimony, Dr. Brock found frequent use of at least ten distinct principles of persuasion. Among them were: (1) the linking of the product with an important human value (CX 34; CX 39A; CX 79A; CX 104; CX 717G; CX 125A, CX 148; CX 729); (2) the use of highly credible sources, such as doctors and medical reports (CX 3A; CX 78A; CX 79A; CX 717G; CX 153A; CX 164; CX 173; CX 176A; CX 204); (3) the repetition of claims or themes (CX 32A; CX 39A; CX 87A; CX 722; CX 153A); (4) the arousal of an apparent conflict in the communication and then the offering of the product as the solution

to this conflict; (CX 22A; CX 32A; CX 87A; CX 722; CX 162A); (5) the [197] making of claims that cannot be refuted by the consumer through his experience (CX 22A; CX 32A; CX 39A; CX 74A; CX 87A); (6) the use of "open-minded manipulation," a technique designed to induce attitude formation or change by asking the viewer to consider the possibility of different points of view (CX 93);[6] (7) the use of metaphors—that is, suggesting the product is like something else with which the viewer is familiar (CX 93)[6]; (8) endowing the commercials with trappings of scientific precision (CX 94; CX 132A; CX 729); (9) describing the message or product as being a scarce commodity making the message more valuable to consumers (CX 82; CX 153A; CX 164); and (10) presenting the product as successfully used by many other consumers ("social comparison principle") (CX 104 CX 148A) (Brock, Tr. 8597–8614, 8627–31). Dr. Brock found that repeated use of these principles of persuasion made product attributes both more salient and beliefs about them more stable in the minds of consumers (Brock, Tr. 8614).

770. There are several methods of ascertaining whether the use of persuasion techniques in advertising have been successful in creating a lasting impact on consumers. These methods include analysis of consumers' acceptance or other immediate reactions to advertising, analysis of the effect of advertising on consumers' intention to purchase or use the products, and analysis of any delayed impact or penetration of the advertising messages (Brock, Tr. 8614–15).

771. An important measure of the success of a communication is the extent to which an individual accepts it (Brock, Tr. 8615). Thus, it is important to look at consumers' immediate reaction to a communication, such as their own feelings of being convinced, informed, or persuaded by the message. Such measures indicate the extent to which the communication was effective in creating an impact in the form of beliefs (Brock, Tr. 8615–16). The ASI copy tests in evidence measured in part consumers' reactions to Bufferin and Excedrin advertisements. Consumers were asked to select from a list of positive and negative adjectives those that best described their feeling about the commercials they had just viewed. In doing so, consumers consistently found the tested commercials for Bufferin and Excedrin to be "informative," "convincing" and "effective." These results confirm the fact that the persuasive techniques used in the Bufferin and Excedrin ads were having an impact, and support the view that this advertising could reasonably be viewed as playing a significant role in forming beliefs about both Bufferin and Excedrin (Brock, Tr. 8619–21, 8633–35).

[6] Dr. Brock referred to CX 94 as an illustration of the principle. CX 93 in evidence is identical to CX 94 (Brock, Tr. 8596–97).

772. Another important measure of the success of persuasion techniques in advertising is the extent to which it influences [198] consumers' intentions to purchase the advertised product (Brock, Tr. 8614–15). The ASI copy tests in evidence also provide information that permits an analysis of the relation between advertising for Bufferin and Excedrin and purchase intention. In those ASI tests consumers were asked about their preferences for various analgesic products, both before and after they viewed various Bufferin and Excedrin commercials. The results showed a small increase in preference for Bufferin and Excedrin after viewing the commercials (CX 828; CX 830; Brock, Tr. 8624, 8635). This increase is significant because of two factors operating against any change in preference at all: the desire to be consistent and the desire to resist the direction of persuasion (Brock, Tr. 8623).

773. An analysis of advertising penetration, or delayed impact also supports the role of advertising as an important factor in forming beliefs about Bufferin and Excedrin. The various advertising penetration studies in evidence demonstrate that a significant number of consumers remembered the superior speed and gentleness claims for Bufferin and the superior effectiveness claims for Excedrin off the top of their heads (F. 718–26, *supra*). Three of these penetration studies (CX 310, 325, 345) can be analyzed to ascertain the effects of advertising over time and whether advertising is influencing use of the product (Brock, Tr. 8638). Data from CX 310 shows that the advertising for both Bufferin and Excedrin was one of the most frequent reasons for initial trial of the product (CX 310Z60; Brock, Tr. 8639–40). Data from these three penetration studies also show that consumers in general had high awareness of the advertising for these products (CX 301Z090; CX 325O, Q, Z, Z001, CX 354V, W; Brock, Tr. 8640, 8647–48) and significant recall of the superior speed and gentleness claims for Bufferin and the superior effectiveness claims for Excedrin (CX 310Z90–95; Brock, Tr. 8641, 8645). These studies show a strong penetration of advertising themes for Bufferin and Excedrin, and a significant connection between the advertising and the beliefs about those products (Ross, Tr. 7510–11).

774. The Bufferin and Excedrin advertisements themselves, the ASI research, the advertising penetration studies and the expert testimony in this record taken together tend to confirm that the advertising for Bufferin played a substantial role in creating and reinforcing consumers' beliefs that Bufferin is gentler to the stomach and a faster pain reliever than aspirin and the belief that Excedrin is a more effective pain reliever than aspirin (Brock, Tr. 8650; Ross, Tr. 7510–11).

E. *The Evidence Regarding Tension Relief Image Of Bufferin, Excedrin and Excedrin P.M. Is Equivocal And Inconclusive*

775. The tension relief claims for Bufferin, Excedrin and Excedrin P.M. began during the early 1960's and ceased by 1970, **[199]** some 10 years ago. Tension relief claims for Excedrin ceased in 1969. The image studies in the record is equivocal and inconclusive on the issue of whether a substantial number of consumers hold "tension reliever" images regarding Bufferin or Excedrin. In these circumstances, it cannot be reasonably inferred from the fact of advertising dissemination a fact of tension relief image among consumers regarding Bufferin, Excedrin or Excedrin P.M.

776. Dr. Ross reviewed the image studies in evidence in this proceeding for the purpose of coming to a conclusion about the nature of people's images of Bufferin (Tr. 7227). Dr. Ross did *not* state that, in his opinion, such image of Excedrin as a tension reliever, which he felt to exist, was created or reinforced by Excedrin advertising. Dr. Ross did *not* state that, in his opinion, such image of Bufferin as a tension reliever, which he felt to exist, was created or reinforced by Bufferin advertising.

777. Dr. Brock reviewed the evidence relevant to the question of an image among consumers for tension relief for these products and declined to give an opinion because he found the evidence "unclear and sparse" (Tr. 8724–25). More people in the Leavitt Study (1.8%) stated that aspirin was good for the relief of tension than stated that either Bufferin (1.7%) or Excedrin (0.9%) was good for the relief of tension (Tr. 6219; CX 350Z008).

778. Dr. Ross admitted that the data indicates that not insubstantial numbers of consumers regard unadvertised plain or store brand aspirin as efficacious for the relief of tension (Tr. 8217–18, 8221). In fact, in a 1964 Gallup survey, 24% of the people surveyed stated that simple aspirin relieved nervous tension (CX 333K).

779. Dr. Ross claimed to find evidence for the existence of a consumer image of Bufferin as a tension reliever in the data collected in CX 345 (Tr. 7311–12). However, none of the data recorded in CX 345 have any relevance to Bufferin advertising, because the only advertisements, which, in Dr. Ross' opinion, made tension relief claims (CX 816A–C) were run exclusively on the West Coast (CX 881O–W), and no sampling was done for CX 345 on the West Coast (Tr. 558; BMF 1220).

780. As part of the basis for his opinion that Bufferin has an image as a tension reliever, Dr. Ross looked to the data contained in CX 310 at Z056 and Z072 (Tr. 7321–23). However, the interviewing for CX 310 was conducted during the period June 6, 1969 through July 20, 1969

(CX 310L) and the "Sensitive People" campaign did not begin running on the West Coast (CX 880W–881B; CX 8810–W) until mid-June 1969 (CX 800K–L). Thus, very few of the people interviewed for CX 310 could have seen that advertising and the few that might have seen it would have [**200**] been exposed to it for, at most, one month. Dr. Ross also relied on CX 1058 and CX 1059 for his opinion regarding an image of Bufferin as a tension reliever and the source of that image. The data on which Dr. Ross relied for these two studies are summarized on CX 822V and W. Those data show that a tension relief image existed for Bufferin in 1967, prior to the Sensitive People campaign, and that image had actually *decreased* slightly in 1970 after the alleged tension ads had run (CX 822V and W). Dr. Ross testified that he could not attribute any meaning to an increase or decrease in the percentages for tension in CX 1058 and CX 1059, image studies which bracketed the period during which the Sensitive People campaign was aired (Tr. 8338, 8463–64).

781. Dr. Leavitt pretested the Leavitt Study questionnaire, CX 349, and the results of the pretest left him confident that his questionnaire was capable of eliciting tension relief responses for Bufferin and Excedrin (Tr. 6274–75), noting "it's certainly possible to get the tension or relaxation or whatever word they happen to use, that kind of response from this question" (Tr. 6278–79), especially if that attribute was salient or important to the consumer (Tr. 6279).

782. Questions six through eight in the Leavitt Study, found at CX 349X, attempted to determine the percentage of the population who felt that Bufferin and Excedrin were good for things "other than pain." (Tr. 8344). According to Dr. Ross' calculations, only 14 respondents, representing 1.8% of the total sample, stated that they thought Bufferin was good for the relief of tension (Tr. 8343). Of those 14, 11 were Bufferin users and only 3 were Bufferin nonusers as defined by the study (Tr. 8347–48). Thus, only 6/10 of 1% of the Bufferin nonusers interviewed for the Leavitt Study indicated a belief that Bufferin was good for the relief of nervous tension (Tr. 8348).

F. *The Record Does Not Contain A Convincing Showing That Consumers' Beliefs About Bufferin and Excedrin Will Endure Unless Corrected*

783. While the recall of specific copy points made in Bufferin and Excedrin advertising may continue for three to nine months after those claims are made, product images or beliefs about Bufferin and Excedrin can endure long after the specific information that led to their formation is forgotten (Ross, Tr. 7509–10).

784. The stability and durability of consumers' product image that Bufferin is gentler and faster acting than aspirin or that Excedrin is

a more effective pain reliever than aspirin, depend on such factors as the sharpness of those images, consumers' usage of the product, the powerful principles of persuasion used in advertising that led to the formation of the beliefs, and the salience of the beliefs (Brock, Tr. 8652–60). [201]

785. There is expert testimony that the two marketing studies in evidence (CX–346 and CX–349) provide data which show that consumers have relatively "sharp" beliefs of both Bufferin and Excedrin: most consumers have definite, as opposed to diffuse, opinions regarding the attributes of these products (Brock, Tr. 8665–67). The sharper the belief, the longer it will endure (Brock, Tr. 8652). According to Dr. Brock, analyses of these marketing studies, conducted in 1967 and late 1975, show that this "sharpness" of beliefs about Bufferin's and Excedrin's superiority has remained high and relatively unchanged for a long period of time. In Dr. Brock's view, this finding supports the conclusion that beliefs about Bufferin and Excedrin are stable and durable ones (Brock, Tr. 8665–67).

786. The beliefs that Bufferin is gentler and faster acting than aspirin and that Excedrin is a more effective pain reliever than aspirin are also salient to consumers. They stand out from beliefs about other attributes of the product's performance (CX 349; CX 346Z150, Z152; Brock, Tr. 8663–64). According to Dr. Brock, this high level of salience, as shown by the market research studies in evidence, has remained consistently high over the time period analyzed, 1967 to 1975 (Brock, Tr. 8664).

787. The quality and consistency of salience and relative sharpness in consumer's product images of Bufferin and Excedrin suggest that they are powerful and durable (Brock, Tr. 8679). According to Dr. Brock, the fact that these beliefs have been shaped by the use of powerful principles of persuasion in advertising makes it even more likely that they will endure (Brock, Tr. 8659). Furthermore, because consumers cannot "evaluate" product differences among mild OTC analgesics, their future usage of Bufferin or Excedrin will not disabuse them of these beliefs created in substantial part by the advertising for those products.

788. Complaint counsel's expert witnesses testified that, assuming that respondents were to cease the challenged advertising claims about Bufferin and Excedrin, the product images that Bufferin is faster and gentler than aspirin and Excedrin is more effective than aspirin will persist indefinitely in the minds of consumers who use the product (Ross, Tr. 7513–14; Brock, Tr. 8698). For nonusers, Dr. Ross testified that beliefs about these attributes will endure for at least one year based upon averaging across the marketing literature which focuses upon the *sales* effects of advertising. Dr. Ross recognized that