MERRELL DOW PHARMACEUTICALS, INC., Petitioner,

v.

Ernest HAVNER and Marilyn Havner on Behalf of their minor child Kelly HAVNER, Respondents.

No. 95–1036.

Supreme Court of Texas.

Argued March 19, 1996.

Decided July 9, 1997.

Order Overruling Rehearing Nov. 13, 1997.

Parents of child who suffered from limb reduction birth defect brought products liability action against manufacturer of prescription drug (Bendectin) ingested by mother during pregnancy. The 214th District Court, Nueces County, Mike Westergren, J., entered judgment on jury verdict awarding actual and exemplary damages to plaintiffs, and manufacturer appealed. After panel initially reversed and rendered judgment, rehearing en banc was granted, and on rehearing, the Corpus Christi Court of Appeals, 907 S.W.2d 535, affirmed as to actual damages, and reversed and rendered as to punitive damages. Application for writ of error was granted, and the Supreme Court, Owen, J., held that: (1) properly designed and executed epidemiological studies indicating that exposure more than doubled risk of injury may be part of evidence supporting finding of causation in toxic tort case; but (2) other factors must be considered, and plaintiff must in addition offer evidence excluding other possible causes of disease with reasonable certainty; and (3) evidence was legally insufficient to establish that child's defect was caused by exposure to drug..

Court of Appeals reversed, and judgment rendered for defendant.

Gonzalez, J., concurred and filed opinion.

Spector, J., concurred and filed opinion.

In determining whether there is no evidence of probative force to support jury's finding, all record evidence must be considered in light most favorable to party in whose favor verdict has been rendered, and every reasonable inference deducible from evidence is to be indulged in that party's favor.

2. Appeal and Error ⇐1001(3)

No evidence point of error will be sustained when (1) there is complete absence of evidence of a vital fact, (2) court is barred by rules of law or of evidence from giving weight to only evidence offered to prove a vital fact, (3) evidence offered to prove a vital fact is no more than a mere scintilla, or (4) evidence conclusively establishes the opposite of the vital fact.

3. Appeal and Error ⇐1001(3)

"More than a scintilla" of evidence exists to support jury finding, and no evidence point of error will be denied, when evidence supporting finding, as a whole, rises to level that would enable reasonable and fair-minded people to differ in their conclusions.

See publication Words and Phrases for other judicial constructions and definitions.

4. Appeal and Error ⇐842(7)

Evidence ⇐570

Expert's bare opinion testimony will not suffice to support factual finding, and substance of testimony must be considered in reviewing legal sufficiency of evidence.

5. Evidence ⇐546

Testimony of expert is generally opinion testimony, and whether such testimony rises to level of evidence is determined under Rules of Evidence.

6. Evidence ⇐546

While rule governing admission of expert testimony deals with admissibility of evidence, it offers substantive guidelines in determining if expert testimony is some evidence of probative value. Rules of Civ.Evid., Rule 702.

substance is capable of causing a particular injury or condition, and there will be objective criteria by which it can be determined with reasonable certainty that a particular individual's injury was caused by exposure to a given substance. However, in many toxic tort cases, direct experimentation cannot be done, and there will be no reliable evidence of specific causation.

In the absence of direct, scientifically reliable proof of causation, claimants may attempt to demonstrate that exposure to the substance at issue increases the risk of their particular injury. The finder of fact is asked to infer that because the risk is demonstrably greater in the general population due to exposure to the substance, the claimant's injury was more likely than not caused by that substance. Such a theory concedes that science cannot tell us what caused a particular plaintiff's injury. It is based on a policy determination that when the incidence of a disease or injury is sufficiently elevated due to exposure to a substance, someone who was exposed to that substance and exhibits the disease or injury can raise a fact question on causation. *See generally Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1320 n. 13 (9th Cir.) (on remand), *cert. denied*, — U.S. —, 116 S.Ct. 189, 133 L.Ed.2d 126 (1995). The Havners rely to a considerable extent on epidemiological studies for proof of general causation. Accordingly, we consider the use of epidemiological studies and the "more likely than not" burden of proof.

A

Epidemiological studies examine existing populations to attempt to determine if there is an association between a disease or condition and a factor suspected of causing that disease or condition. *See, e.g.*, Bert Black & David E. Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation*, 52 FORDHAM L.REV. 732, 750 (1984). However, witnesses for the Havners and commentators in this area uniformly acknowledge that epidemiological studies cannot establish that a given individual contracted a disease or condition due to exposure to a particular drug or agent. *See, e.g.*, Michael Dore, *A Commentary on the Shifting Cause-In-Fact*, 7 HARV. ENVTL. L. REV. 429, 431–35 (1983); Steve Gold, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion, and Statistical Evidence*, 96 YALE L.J. 376, 380 (1986). Dr. Glasser, a witness for the Havners, gave as an example a study designed to see if a given drug causes rashes. Even though a study may show that ten people who took the drug exhibited a rash, while rashes appeared on only three people who did not take the drug, Dr. Glasser explained that the study cannot tell us which of the exposed ten got the rash because of the drug. We know that things other than the drug cause rashes.

Recognizing that epidemiological studies cannot establish the actual cause of an individual's injury or condition, a difficult question for the courts is how a plaintiff faced with this conundrum can raise a fact issue on causation and meet the "more likely than not" burden of proof. Generally, more recent decisions have been willing to recognize that epidemiological studies showing an increased risk may support a recovery. Judge Weinstein, whose decision in the Agent Orange litigation has been widely discussed and followed, has observed that courts have been divided between the "strong" and "weak" versions of the preponderance rule. *In re "Agent Orange" Prod. Liab. Litig.*, 611 F.Supp. 1223, 1261 (E.D.N.Y.1985) (citing David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L.REV. 851, 857 (1984)). The "strong" version requires a plaintiff to offer both epidemiological evidence that the probability of causation exceeds fifty percent in the exposed population and "particularistic" proof that the substance harmed the individual. The "weak" version allows verdicts to be based solely on statistical evidence. Rosenberg, *supra*, 97 HARV. L. REV. at 857–58. Judge Weinstein concluded that the plaintiffs in *Agent Orange* were required to offer evidence that causation was "more than 50 percent probable," 611 F.Supp. at 1262, and that the plaintiffs' experts were required to "rule out the myriad other possible causes of the veterans' afflictions," *id.* at 1263.

the ingestion of a particular drug by the mother caused the birth defect. Similarly, an expert's assertion that a physical examination confirmed causation should not be accepted at face value. In *O'Conner v. Commonwealth Edison Co.*, 13 F.3d 1090 (7th Cir.1994), a treating physician testified that he knew what radiation-induced cataracts looked like because they are clinically describable and definable and "cannot be mistaken for anything else." *Id.* at 1106. Nevertheless, his opinion that exposure to radiation caused the plaintiff's cataracts was found to be inadmissible because it had no scientific basis. The literature on which the expert relied did not support his assertion that radiation-induced cataracts could be diagnosed by visual examination. *Id.* at 1106–07. For a good discussion of the evils of "evidence" of this nature, see Bernstein, *supra*, 15 CARDOZO L.REV. at 2148–49. Further, as we discuss in Part VI(A), an expert cannot dissect a study, picking and choosing data, or "reanalyze" the data to derive a higher relative risk if this process does not comport with sound scientific methodology.

The FDA has promulgated regulations that detail the requirements for clinical investigations of the safety and effectiveness of drugs. 21 C.F.R. § 314.126 (1996). These regulations state that "[i]solated case reports, random experience, and reports lacking the details which permit scientific evaluation will not be considered." *Id.* § 314.126(e). Courts should likewise reject such evidence because it is not scientifically reliable. As Bernstein points out, physicians following scientific methodology would not examine a patient or several patients in uncontrolled settings to determine whether a particular drug has favorable effects, nor would they rely on case reports to determine whether a substance is harmful. *See* Bernstein, *supra*, 15 CARDOZO L.REV. at 2148–49; *see also* Rosenberg, *supra*, 97 HARV. L.REV. at 870 (arguing that anecdotal or particularized evidence accomplishes no more than a false appearance of direct and actual knowledge of a causal relationship). Expert testimony that is not scientifically reliable cannot be used to shore up epidemiological studies the risk.

E

To raise a fact issue on causation and thus to survive legal sufficiency review, a claimant must do more than simply introduce into evidence epidemiological studies that show a substantially elevated risk. A claimant must show that he or she is similar to those in the studies. This would include proof that the injured person was exposed to the same substance, that the exposure or dose levels were comparable to or greater than those in the studies, that the exposure occurred before the onset of injury, and that the timing of the onset of injury was consistent with that experienced by those in the study. *See generally* Thompson, *supra*, 71 N.C. L.REV. at 286–88. Further, if there are other plausible causes of the injury or condition that could be negated, the plaintiff must offer evidence excluding those causes with reasonable certainty. *See generally E.I. du Pont de Nemours & Co. v. Robinson*, 923 S.W.2d 549, 559 (Tex. 1995) (finding that the failure of the expert to rule out other causes of the damage rendered his opinion little more than speculation); *Parker v. Employers Mut. Liab. Ins. Co.*, 440 S.W.2d 43, 47 (Tex.1969) (holding that a cause becomes "probable" only when "in the absence of other reasonable causal explanations it becomes more likely than not that the injury was a result").

In sum, we emphasize that courts must make a determination of reliability from all the evidence. Courts should allow a party, plaintiff or defendant, to present the best available evidence, assuming it passes muster under *Robinson*, and only then should a court determine from a totality of the evidence, considering all factors affecting the reliability of particular studies, whether there is legally sufficient evidence to support a judgment.

Finally, we are cognizant that science is constantly reevaluating conclusions and theories and that over time, not only scientific knowledge but scientific methodology in a particular field may evolve. We have strived to make our observations and holdings in light of current, generally accepted scientific

ence may require reevaluation of what is "good science" in future cases.

V

Certain conventions are used in conducting scientific studies, and statistics are used to evaluate the reliability of scientific endeavors and to determine what the results tell us. In this opinion, we consider some of the basic concepts currently used in scientific studies and statistical analyses and how those concepts mesh with our legal sufficiency standard of review. For an extended discussion of statistical methodology and its use in epidemiological studies, see *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 945–48 (3d Cir.1990). *See also Turpin v. Merrell Dow Pharms., Inc.*, 959 F.2d 1349, 1353 n. 1 (6th Cir.1992); Bailey et al., *Reference Guide on Epidemiology, in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra*, at 138–43, 171–78. We do not attempt to discuss all the multifaceted aspects of the scientific method and statistics, but focus on the principles that shed light on the particular facts and issues in this case.

A

One way to study populations is by a retrospective case-control or case-comparison epidemiological study. For example, this type of study identifies individuals with a disease and a suitable control group of people without the disease and then looks back to examine postulated causes of the disease. *See* Bailey et al., *Reference Guide on Epidemiology, in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra*, at 136–38, 172. Another type of epidemiological study is a cohort study, or incidence study, which is a prospective study that identifies groups and observes them over time to see if one group is more likely to develop disease. *Id.* at 134–36, 173.

An "odds ratio" can be calculated for a case-control study. *Id.* at 175. For example, an odds ratio could be used to show the odds that ingestion of a drug is associated with a particular disease. The odds ratio compares the odds of having the disease when exposed to the drug versus when not exposed. If the to develop the disease under study.

Similarly, the "relative risk" that a person who took a drug will develop a particular disease can be determined in a cohort study. *Id.* at 173, 176. The relative risk is calculated by comparing the incidence of disease in the exposed population with the incidence of the disease in the control population. If the relative risk is 1.0, the risk in exposed individuals is the same as unexposed individuals. If the relative risk is greater than 1.0, the risk in exposed individuals is greater than in those not exposed. If the relative risk is less than 1.0, the risk in exposed individuals is less than in those not exposed. For the result to indicate a doubling of the risk, the relative risk must be greater than 2.0. *See id.* at 147–48.

Perhaps the most useful measure is the attributable proportion of risk, which is the statistical measure of a factor's relationship to a disease in the population. It represents the "proportion of the disease among exposed individuals that is associated with the exposure." *Id.* at 149. In other words, it reflects the percentage of the disease or injury that could be prevented by eliminating exposure to the substance. For a more detailed discussion of the calculation and use of the attributable proportion of risk, see *id.* at 149–50; Black & Lilienfeld, *supra*, 52 FORDHAM L.REV. at 760–61. *See also* Thompson, *supra*, 71 N.C. L.REV. at 252–56.

The numeric value of an odds ratio is at least equal to the relative risk, but the odds ratio often overstates the relative risk, especially if the occurrence of the event is not rare. For an example of the difference between the mathematical calculation of the odds ratio and the relative risk, see BARBARA HAZARD MUNRO & ELLIS BATTEN PAGE, STATISTICAL METHODS FOR HEALTH CARE RESEARCH 233–35 (2d ed. 1993). In the example given by Munro and Page, the odds ratio was 3.91, while the relative risk was only 3.0 based on the same set of data. *See also* Bailey et al., *Reference Guide on Epidemiology, in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra*, at 149; Thompson, *supra*, 71 N.C. L.REV. at 250 n. 22.

testified. Dr. Swan derived a relative risk estimate of 2.2 for women exposed to Bendectin during the first trimester. She also testified that the relative risk for women who were exposed to Bendectin but not exposed to spermicide was 8.8 and finally, that if women who were exposed to two or more Bendectin prescriptions were considered, without regard to exposure to spermicide, the relative risk was 13 with a confidence interval from 3 to 53. She did not reveal the confidence level used in obtaining these results, and there is no evidence of the confidence level in the record.

The other reanalysis by Dr. Swan was of data in the Cordero study, which was based on information collected by the Center for Disease Control in Atlanta. An abstract she prepared regarding this data was published in the *Journal for the Society of Epidemiological Research* in 1983 or 1984 and states that the original Cordero study found the odds ratio for limb reduction birth defects to be 1.2. Swan concluded, however, that when a different control group is selected, the relative risk estimates are affected. Swan's abstract stated that, "under certain assumptions," which are not identified, "the odds ratio for limb reduction defects" are "a highly significant" 2.8. There is no explanation in the abstract or in Dr. Swan's testimony of the significance level used to obtain the 2.8 result. The result may well be statistically inconclusive at a 95% confidence level. We simply do not know from this record. Without knowing the significance level or the confidence interval, there is no scientifically reliable basis for saying that the 2.8 result is an indication of anything. Further, her choice of the control group could have skewed the results. Although her abstract does not identify what control group she used, Swan testified at trial that she chose births of Downs Syndrome babies. Swan's reanalysis using Downs Syndrome babies as the control group was considered in *Lynch* and in *Richardson–Merrell*, and those courts likewise found it insufficient. *See Lynch v. Merrell–National Labs.*, 830 F.2d 1190, 1195 (1st Cir.1987), *aff'd*, 857 F.2d 823 (D.C.Cir. 1988); *Richardson v. Richardson–Merrell*,

In addition to the statistical shortcomings of the Havners' epidemiological evidence, another strike against its reliability is that it has never been published or otherwise subjected to peer review, with the exception of Dr. Swan's abstract, which she acknowledges is not the equivalent of a published paper. Dr. Swan has published a number of papers in scientific journals, including a study that concluded Bendectin is not associated with cardiac birth defects. Although she has been testifying in Bendectin limb reduction birth defect cases for many years, Dr. Swan has never attempted to publish her opinions or conclusions about Bendectin and limb reduction defects. Similarly, studies by Dr. Glasser have been published in refereed journals, but none of his 32 to 33 publications mentions Bendectin or limb reduction birth defects.

As already discussed, there are over thirty published, peer-reviewed epidemiological studies on the relationship between Bendectin and birth defects. None of the findings offered by the Havners' five experts in this case have been published, studied, or replicated by the relevant scientific community. As Judge Kozinski has said, "the only review the plaintiffs' experts' work has received has been by judges and juries, and the only place their theories and studies have been published is in the pages of federal and state reporters." *Daubert*, 43 F.3d at 1318 (commenting on the same five witnesses called by the Havners). A related factor that should be considered is whether the study was prepared only for litigation. Has the study been used or relied upon outside the courtroom? Is the methodology recognized in the scientific community? Has the litigation spawned its own "community" that is not part of the purely scientific community? The opinions to which the Havners' witnesses testified have never been offered outside the confines of a courthouse.

[14] Publication and other peer review is a significant indicia of the reliability of scientific evidence when the expert's testimony is in an area in which peer review or publication would not be uncommon. Publication in

related argument is that history tells us that the scientific community has been slow at times to accept valid research and its results. While these observations are true, history also tells us that valid and reliable research and theories are generally accepted quickly within the scientific community when sufficient explanation is provided and empirical data are adequate. *See* Black et al., *supra*, 72 TEX. L. REV. at 779–82 (discussing Galileo, Pasteur, DNA, and continental drift).

[16] Others have argued that liability should not be allocated only on the basis of reliable proof of fault because legal rules should have the goals of "risk spreading, deterrence, allocating costs to the cheapest cost-avoider, and encouraging socially favored activities," and because " 'consumers of American justice want people compensated.' " Rochelle Cooper Dreyfuss, *Is Science a Special Case? The Admissibility of Scientific Evidence After Daubert v. Merrell Dow*, 73 TEX. L.REV. 1779, 1795–96 (1995) (quoting Kenneth R. Feinberg, *Civil Litigation in the Twentieth–First Century: A Panel Discussion*, 59 BROOK. L.REV. 1199, 1206 (1993)). It has been contended that "[f]or some cases that very well may mean creating a compensatory mechanism even in the absence of clear scientific proof of cause and effect" and that "[d]eferring to scientific judgments about fault only obscures the core policy questions that are addressed by the laws that the court is applying." *Id.* We expressly reject these views. Our legal system requires that claimants prove their cases by a preponderance of the evidence. In keeping with this sound proposition at the heart of our jurisprudence, the law should not be hasty to impose liability when scientifically reliable evidence is unavailable. As Judge Posner has said, "[l]aw lags science; it does not lead it." *Rosen v. Ciba–Geigy Corp.*, 78 F.3d 316, 319 (7th Cir.), *cert. denied*, —— U.S. ——, 117 S.Ct. 73, 136 L.Ed.2d 33 (1996).

B

The Havners relied on *in vivo* animal studies to support the conclusion that Bendectin

Adrian Gross, a veterinarian and a veterinary pathologist who had worked at the FDA from 1964 to 1979, served as the Chief of the Toxicology Branch at the Environmental Protection Agency from 1979 to 1980, and thereafter was a Senior Science Advisor at the EPA. Dr. Gross confirmed that the FDA and EPA consider animal studies in assessing the potential human response to drugs or pesticides. He testified that what will affect an animal is likely to affect humans in the same way and that the only reason animal studies are done is to predict if the drug at issue will have an adverse effect on humans.

Dr. Gross reviewed a number of animal studies that had been conducted on Bendectin. He described studies on rabbits exposed to Bendectin in which he saw "a lot of malformed kits." Gross testified about another study of rabbits that he found statistically significant. He opined that the probability that the malformations in this study occurred by chance were six in 10,000. With respect to another animal study on rabbits, he stated that the probability that the drug was harmless was less than one per 1,000,000. He listed studies on monkeys, rats, and mice showing "highly significant deleterious harmful effects as far as birth defects are concerned." Based on these animal studies, Dr. Gross was of the opinion that Bendectin was teratogenic in humans, which means that it causes birth defects. However, he conceded that the dosage levels at which Bendectin became associated with birth defects in rats was at 100 milligrams per kilogram per day, which would be the equivalent of a daily dosage of 1200 tablets for a woman weighing 132 pounds.

The Havners assert in their briefing before this Court that the accepted technique for determining if a substance is a teratogen in humans is to look at all information, including epidemiological data, animal data, biological plausibility, and *in vitro* studies. Dr. Swan confirmed that these are the relevant sources of information in determining teratogenicity. *See also* Brent, *Comment on Comments on "Teratogen Update: Bendectin,"*

(1) consistent, reproducible findings in human epidemiological studies; (2) development of an animal model; (3) embryo toxicity that is dose related; and (4) consistency with basic, recognized concepts of embryology and fetal development). Thus, scientific methodology would not rely on animal studies, standing alone, as conclusive evidence that a substance is a teratogen in humans. *See Raynor v. Merrell Pharms., Inc.*, 104 F.3d 1371, 1375 (D.C.Cir.1997) (noting that the only way to test whether data from nonhuman studies can be extrapolated to humans would be to conduct human experiments or to use epidemiological data); *Elkins v. Richardson–Merrell, Inc.*, 8 F.3d 1068, 1071 (6th Cir.1993) (holding that expert opinion indicating a basis of support in animal studies is admissible but is simply inadequate to permit a jury to conclude that Bendectin more probably than not causes limb defects); *Lynch*, 830 F.2d at 1194 (asserting that *in vivo* and *in vitro* animal studies singly or in combination do not have the capability of proving causation in human beings in the absence of any confirming epidemiological data); *see also Brock*, 874 F.2d at 313 (recognizing that animal studies are of very limited usefulness when confronted with questions of toxicity); *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 197 (5th Cir.1996) (quoting and following *Brock* in toxic tort case).

We further note that with respect to the *in vivo* studies about which Dr. Gross testified, their reliability as predictors of the effect of Bendectin in humans is questionable because of the dosage levels. Dr. Gross offered no explanation of how the very high dosages could be extrapolated to humans. Other courts have rejected animal studies that relied on high dosage levels as evidence of causation in humans. *See, e.g., Turpin v. Merrell Dow Pharms., Inc.*, 959 F.2d 1349 (6th Cir.1992) (reasoning that to eliminate drugs toxic to embryos at high dosage levels would eliminate most drugs and many useful chemicals on which modern society depends heavily) (citing James Wilson, *Current Status of Teratology, in* HANDBOOK OF TERATOLOGY 60 (1977)). Gross also failed to explain why the published studies from which he extract-

The *in vivo* studies identified in this case cannot support the jury's verdict.

C

Dr. Stuart Allen Newman also relied on animal studies to support his opinion that Bendectin is a teratogen in humans. Dr. Newman holds a doctorate in chemical physics and is a professor at New York Medical College. He has published over fifty articles, although none contain the opinions or conclusions to which he testified in this case.

The studies Newman reviewed were *in vitro* studies, which are based on tests conducted on cells in a test tube or petri dish. Doxylamine succinate was placed directly on the limb bud cells of animals including chickens and mice. The development of cartilage was affected. Newman acknowledged that in these studies, the researchers who had conducted them concluded only that doxylamine succinate was potentially capable of inducing genetic damage and that it should be tested on other systems. But Newman testified that if you find an effect that prevails across a number of different species, "you can be awfully sure that the same thing will prevail in humans."

[17] Newman opined that Kelly Havner's defect was due to loss of portions of the skeleton that *could* with scientific certainty have been caused by a teratogen that affected the embryo. Similarly, he testified that the findings of one study, the Hassell/Horigan Study, indicated to him that doxylamine succinate *can* interfere with chondrogenesis, which is the process of certain cells turning into cartilage. We note that testimony to the effect that a substance "could" or "can" cause a disease or disorder is not evidence that in reasonable probability it does. *See, e.g., Parker v. Employers Mut. Liab. Ins. Co.*, 440 S.W.2d 43, 47 (Tex.1969); *Bowles v. Bourdon*, 148 Tex. 1, 219 S.W.2d 779, 785 (1949). Newman testified, however, that based on the Hassell/Horigan and other animal studies, he concluded with a reasonable degree of medical certainty that doxylamine succinate is a teratogen for cartilage development and