Winter 2004

A Quantitative Methodology for Determining the Need for Exposure-Prompted Medical Monitoring

Christopher P. Guzelian
Stanford Law School

Bruce E. Hillner
Virginia Commonwealth University

Philip S. Guzelian
University of Colorado Health Sciences Center

JEROME HALL LAW LIBRARY
INDIANA UNIVERSITY
Maurer School of Law
Bloomington

A Quantitative Methodology for Determining the Need for Exposure-Prompted Medical Monitoring

CHRISTOPHER P. GUZELIAN,* BRUCE E. HILLNER,** PHILIP S. GUZELIAN***

Some toxic exposures to drugs or other environmental chemicals may create an increased risk of future disease for which periodic preventive medical screening might be desirable. However, many of these risks, even if unacceptable as a matter of public health policy, might still not be *significant* enough for medical monitoring (periodic diagnostic screening for latent illnesses or medical conditions) to be an appropriate medical intervention. This somewhat unintuitive, but statistically certain, conclusion can be demonstrated in relatively simple mathematical terms. Accordingly, we introduce Bayes's Rule and decision analysis, a quantitative methodology commonly employed by medical practitioners. A review of current medical practices indicates that physicians decide whether to recommend monitoring for a particular exposed population by knowing the natural history of the disease and by first calculating the predictive value of a positive test ("PPV"), which will be one to five percent or greater for an endorsable monitoring exercise, absent exceptional circumstances. Rather than simply relying on the opinions of retained medical experts, this accessible quantitative method permits judges, jurists, and policymakers to more confidently and objectively decide whether medical monitoring is appropriate and necessary as a result of a specific chemical exposure.

TABLE OF CONTENTS

INTRODUCTION

The last thirty years bear witness to a vast new effort by government, academia, and industry to address the question of adverse human health effects arising from exposures to chemicals in the environment. Regulatory bodies characterize and control the production, use, release, and cleanup of a large number of industrial chemical and waste byproducts. In turn, basic scientists, toxicologists, physicians, and public health officials catalog the potential adverse health effects of chemical exposures and seek to understand their underlying toxicologic mechanisms. The legal profession has not been idle in these endeavors. An entire subspecialty of environmental law, toxic torts, which involves questions of personal injury from chemicals used in the workplace or otherwise present as contaminants in the environment, has arisen as a result of increasing public concern with environmental exposures.

Traditionally, toxic tort case law maintained a straightforward requirement: a plaintiff had to demonstrate present physical injury caused by exposure to industrial chemicals, drugs, or other environmental agents to recover damages or comparable concessions because of tortious exposure.[1] In the past two decades, however, plaintiffs who claim to have been exposed to toxic amounts of environmental chemicals have begun requesting relief where there are no manifest injuries or illness symptoms, but only an ostensibly increased statistical likelihood that disease will develop.

One particular form of relief such exposed plaintiffs seek is "medical monitoring"—periodic diagnostic medical examinations and medical tests intended to diagnose illness or medical conditions before they would be diagnosed in the course of ordinary medical care. Some courts—both state and federal—have refused to recognize medical monitoring prompted by chemical exposure.[2] But

---

1. *E.g., In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 850 (3d Cir. 1990) ("Paoli I") ("[U]nder traditional common law tort doctrine . . . injury needed to be manifest before it could be compensable."). *See generally* Allan T. Slagel, Note, *Medical Surveillance Damages: A Solution to the Inadequate Compensation of Toxic Tort Victims*, 63 IND. L.J. 849 (1988) (arguing for the increased use of medical surveillance as a solution to inadequate tort compensation); Palma J. Strand, Note, *The Inapplicability of Traditional Tort Analysis to Environmental Risks: The Example of Toxic Waste Pollution Victim Compensation*, 35 STAN. L. REV. 575, 576-88 (1983) (describing the evolution undergone by the common law tort system to compensate tort victims).

2. Metro-North Commuter R.R. Co. v. Buckley, 521 U.S. 424, 439-40 (1997) (rejecting particular plaintiff's medical monitoring claim because of lack of compensable injury under Federal Employers' Liability Act and because intermediate court's envisioned potential award of medical monitoring to asymptomatic plaintiff in form of lump sum damages went beyond bounds of "evolving common law" as it now stands); Berg v. E.I. DuPont de Nemours & Co. (*In re* Berg Litig.), 293 F.3d 1127, 1133 (9th Cir. 2002) (holding that the Price-Anderson Act does not permit medical monitoring); Trimble v. Asarco, Inc., 232 F.3d 946, 952, 963 (8th Cir. 2000) (agreeing with the district court that "[t]here exists no pending or prospective legislation to authorize a cause of action or a remedy for medical monitoring, and the court finds it improbable that the Nebraska [state] courts would judicially fashion such a right or remedy"); Ball v. Joy Techs., Inc., 958 F.2d 36, 39 (4th Cir. 1991) (ruling that

others have permitted exposure-prompted medical monitoring in some fashion, either as a remedy or independent cause of action.[3] A survey of court

---

medical monitoring is not cognizable claim under tort law of Virginia or West Virginia absent manifestation of physical injury); Schweitzer v. Consol. Rail Corp., 758 F.2d 936, 942 (3d Cir. 1985); Duncan v. Northwest Airlines, Inc., 203 F.R.D. 601, 608-09 (W.D. Wash. 2001) (holding that Washington does not recognize medical monitoring claims); Purjet v. Hess Oil V.I. Corp., No. Civ. 1985/284, 1986 WL 1200, at *4 (D.V.I. Jan. 8, 1996) (holding actual injury is a prerequisite to medical monitoring in the Virgin Islands); Thomas v. FAG Bearings Corp., Inc., 846 F. Supp. 1400, 1410 (W.D. Mo. 1994); Villari v. Terminix Int'l, Inc., 663 F. Supp. 727, 735 (E.D. Pa. 1987) (allowing medical monitoring only if some physical injury—whether related to feared future diseases or not—is present); Greenberg v. McCabe, 453 F. Supp. 765 (E.D. Pa. 1978); Bakeries, Inc. v. Knipp, 852 So. 2d 712, 718 n.9 (Ala. 2002) ("Alabama does not recognize a cause of action for medical monitoring."); Hinton v. Monsanto Co., 813 So. 2d 827, 832 (Ala. 2001) (holding that medical monitoring is categorically not recognized in Alabama); Badillo v. Am. Brands, Inc., 16 P.3d 435 (Nev. 2001) (refusing to recognize medical monitoring as an independent cause of action in Nevada); Mergenthaler v. Asbestos Corp. of Am., 480 A.2d 647, 651 (Del. 1984) (finding present physical injury is "essential element" of medical monitoring claim arising from asbestos exposure); Ayers v. Township of Jackson, 493 A.2d 1314, 1322 (N.J. Super. Ct. App. Div. 1985) ("Ayers II"), *rev'd*, Ayers v. Township of Jackson, 525 A.2d 287 (N.J. 1987) ("Ayers III").

3. Marine Asbestos Cases v. Am. Haw. Cruises, Inc., 265 F.3d 861 (9th Cir. 2001) (assuming without deciding that medical monitoring is permissible under Jones Act); In re Paoli R.R. Yard PCB Litig., 35 F.3d 717, 785 (3d Cir. 1994) ("Paoli II"); Bldg. & Constr. Dep't v. Rockwell Int'l Corp., 7 F.3d 1487, 1493-94 (10th Cir. 1993); Abuan v. General Elec. Co., 3 F.3d 329, 335 (9th Cir. 1993) (stating in dicta that against the backdrop of its "extremely plaintiff-friendly" toxic tort law, Guam would permit some medical monitoring claims); In re Paoli R.R. Yard PCB Litig., 916 F.2d 829, 850 (3d Cir. 1990) ("Paoli I"); Hagerty v. L & L Marine Servs., Inc., 788 F.2d 315, 318 (5th Cir. 1986) (medical monitoring permitted in mental distress case "[w]ith or without physical injury"), *but see* Hagerty v. L & L Marine Servs., Inc., 797 F.2d 256, 256 (5th Cir. 1986) (Jolly, J., dissenting) (opining that phrase was ambiguous and physical injury must be demonstrated in medical monitoring claim); Friends for All Children, Inc. v. Lockheed Aircraft Corp., 746 F.2d 816 (D.C. Cir. 1984); Carey v. Kerr-McGee Chem. Corp., 999 F. Supp. 1109 (N.D. Ill. 1998) (finding Illinois law can recognize a medical monitoring claim); Gibbs v. E.I. DuPont De Nemours & Co., 876 F. Supp. 475, 478-79 (W.D.N.Y. 1995) (finding that New York recognizes medical monitoring claims); O'Neal v. Dep't of the Army of the United States, 852 F. Supp. 327, 336 (M.D. Pa. 1994) (recognizing medical monitoring generally in Pennsylvania, but concluding that plaintiff's risks, which were unquantified or insignificant, did not meet *Paoli* requirements); Day v. Nat'l Lead of Ohio, 851 F. Supp. 869, 880-82 (S.D. Ohio 1994), *vacated in part on other grounds sub. nom. In re* NLO, Inc., 5 F.3d 154 (6th Cir. 1993); Bocook v. Ashland Oil, Inc., 819 F. Supp. 530, 537 (S.D. W. Va. 1993) (finding medical monitoring would be recognizable under Kentucky law); Cain v. Armstrong World Indus., 785 F. Supp. 1448, 1451-52 (S.D. Ala. 1992) (awarding medical monitoring damages); In re Hanford Nuclear Reservation Litig., 780 F. Supp. 1551, 1566-67 (E.D. Wash. 1991) (recognizing a medical monitoring claim under CERCLA); Stead v. F.E. Myers Co., 785 F. Supp. 56, 57 (D. Vt. 1990); Werlein v. United States, 746 F. Supp. 887 (D. Minn. 1990) (holding that a plaintiff can recover costs for future medical monitoring as tort damages if the plaintiff shows that her exposure to a toxic substance has caused an increased risk of incurring a disease and her doctor recommends medical monitoring), *vacated in part on other grounds*, 793 F. Supp. 898 (D. Minn. 1992); Merry v. Westinghouse Elec. Corp., 684 F. Supp. 847, 849-50 (M.D. Pa. 1988); Johnson v. Armstrong Cork Co., 645 F. Supp. 764, 769 (W.D. La. 1986); Burns v. Jaquays Mining Corp., 752 P.2d 28 (Ariz. Ct. App. 1987); Potter v. Firestone Tire & Rubber Co., 863 P.2d

recommendations shows a pattern of elements generally necessary to a successful cause of action:[4]

TYPICAL ELEMENTS OF A SUCCESSFUL MEDICAL MONITORING CLAIM

1) A "significant" exposure[5]
2) to a hazardous substance with a probable (or proven) link to a human disease[6]

795, 823 (Cal. 1993) (en banc) (recognizing medical monitoring in California); Miranda v. Shell Oil Co., 26 Cal. Rptr. 2d 655, 658-59 (Cal. Ct. App. 1993); Gutierrez v. Cassiar Mining Corp., 75 Cal. Rptr. 2d 132 (Cal. Ct. App. 1998); Doe v. City of Stamford, 699 A.2d 52 (Conn. 1997) (recognizing medical monitoring claim under Connecticut Workers' Compensation statute); Petito v. A.H. Robins Co., 750 So. 2d 103, 106 (Fla. Dist. Ct. App. 1999) (declaring medical monitoring valid cause of action in Florida); Bourgeois v. A.P. Green Indus., Inc., 783 So. 2d 1251, 1261 (La. 2001) ("Bourgeois II"), *overruled on statutory grounds*, 1999 La. Acts 989; Bourgeois v. A.P. Green Indus., 716 So. 2d 355 (La. 1998) ("Bourgeois I"), *overruled on statutory grounds*, 1999 La. Acts 989; Bonnette v. Conoco, Inc., 801 So. 2d 501 (La. Ct. App. 2001); Lilley v. Bd. of Supervisors of La. State Univ., 735 So. 2d 696 (La. Ct. App. 1999) (acknowledging *Bourgeois I* as controlling); Dragon v. Cooper/T. Smith Stevedoring Co., 726 So. 2d 1006, 1012 (La. Ct. App. 1999) (state substantive law can be applied to non-lump-sum medical monitoring claims under federal FELA act); Meyerhoff v. Turner Constr. Co., 534 N.W.2d 204 (Mich. Ct. App. 1995); Elam v. Alcolac, Inc., 765 S.W.2d 42, 209 (Mo. Ct. App. 1988) (dictum); Theer v. Philip Carey Co., 628 A.2d 724, 733 (N.J. 1993) (requiring that the exposure of a plaintiff be direct); Ayers v. Township of Jackson, 525 A.2d 287, 309 (N.J. 1987) ("Ayers III"); Ayers v. Jackson, 461 A.2d 184 (N.J. Super. Ct. Law Div. 1983) ("Ayers I"); Abusio v. Consol. Edison Co., 656 N.Y.S.2d 371 (N.Y. App. Div. 1997); Askey v. Occidental Chem. Corp., 477 N.Y.S.2d 242, 247 (N.Y. App. Div. 1984); Acevedo v. Consol. Edison Co., 572 N.Y.S.2d 1015 (N.Y. Sup. Ct. 1991); Gerardi v. Nuclear Util. Servs., Inc., 566 N.Y.S.2d 1002 (N.Y. Sup. Ct. 1991); Redland Soccer Club, Inc. v. Dep't of the Army of the United States, 696 A.2d 137 (Pa. 1997) (recognizing medical monitoring under Pennsylvania common law); Simmons v. Pacor, Inc., 674 A.2d 232, 239 (Pa. 1996); Wagner v. Anzon, Inc., 684 A.2d 570 (Pa. Super. Ct. 1996); Foust v. SEPTA, 756 A.2d 112 (Pa. Commw. Ct. 2000); *In re* Pa. Diet Drugs Litig., 41 Pa. D. & C.4th 78 (1999); Laxton v. Orkin Exterminating Co., 639 S.W.2d 431 (Tenn. 1982); Hansen v. Mountain Fuel Supply Co., 858 P.2d 970, 979 (Utah 1993) (recognizing medical monitoring in Utah); Bower v. Westinghouse Elec. Corp., 522 S.E.2d 424 (W. Va. 1999).

4. It is quite true that "[m]edical monitoring is not a uniform concept among the states; the elements a plaintiff must prove to establish a right to medical monitoring differ among the states." Dhamer v. Bristol-Myers Squibb Co., 183 F.R.D. 520, 533 (N.D. Ill. 1998). The following elements are merely representative, and the important differences in medical monitoring laws are the major focus of this Article.

5. Although most courts consider a *significant* exposure necessary to proving a need for medical monitoring, few have given script to what this expression means. We surmise that the intent of most courts including this element (and a definition that would be in accord with contemporary medical practice) is to consider any exposure leading to a *significant risk* of a *serious* disease. *See, e.g.*, Redland Soccer Club, Inc. v. Dep't of the Army of the United States, 55 F.3d 827, 846 (3d Cir. 1995) (Significant exposure "refers to an exposure which, either by duration or harm, is sufficient to cause a significantly increased risk, which in turn is sufficient to require a monitoring regime different from that normally required in the absence of such an exposure."). *Cf.* Theer v. Philip Carey Co., 628 A.2d 724, 733 (N.J. 1993) ("[A] cause of action applies only to persons who have been *directly* exposed to hazardous substances.").

3) that creates a
    a.   quantifiable,[7] and
    b.   significant[8]
increase in the exposed individual's risk
4) of developing one or more "serious" diseases,[9]
5) such that there is a demonstrable clinical value of medical monitoring (that is, as a result of early detection and diagnosis, treatment exists that improves either morbidity or mortality statistics).
6) Lastly, courts will only grant medical monitoring if the monitoring is *in excess of or in addition to* a diagnostic regime that is either (i) recommended for a part or all of the population at large by contemporary medical consensus or (ii) necessary as a result of individual risk factors distinct from the defendant's tortious exposure (such as hereditary and

---

6. Some courts require a substance to have a proven link to a serious disease before the court will entertain medical monitoring requests. *See In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 787 (3d Cir. 1994) ("Paoli II"); *In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 852 (3d Cir. 1990) ("Paoli I"); Petito v. A.H. Robins Co., 750 So. 2d 103, 106 (Fla. Dist. Ct. App. 1999); *Bourgeois I*, 716 So. 2d at 360-61; Redland Soccer Club, Inc. v. Dep't of the Army of the United States, 696 A.2d 137, 145 (Pa. 1997). Others, however, only demand proof of a "probable" association to some human disease. Bower v. Westinghouse Elec. Corp., 522 S.E.2d 424, 433 (W. Va. 1999); Thomas v. FAG Bearings Corp., 846 F. Supp. 1400, 1410-11 (W.D. Mo. 1994) (requiring the compound to be "potentially" hazardous). Some courts have an explicit requirement that the compound potentially (or definitely) causes serious disease in humans, rather than simply in tested animals. Hansen v. Mountain Fuel Supply Co., 858 P.2d 970, 979 (Utah 1993) ("We note that the substance must be toxic to humans rather than to other forms of life."). Other courts permit demonstrated toxicity in animals to meet this requirement. *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 781 (3d Cir. 1994) ("Paoli II") ("Here, where the EPA has relied on animal studies to conclude that PCBs are a probable human carcinogen, where there is reason to think that animal studies are particularly valuable because animals react similarly to humans with respect to the chemical in question, and where the epidemiological data is inconclusive and some of it supports a finding of causation, we think that the district court abused its discretion in excluding the animal studies."). Still other opinions are simply ambiguous as to what demonstration of toxicity is required. Merry v. Westinghouse Elec. Corp., 684 F. Supp. 847, 850-51 (M.D. Pa. 1988) (considering only "the toxicity of the substance"); Potter v. Firestone Tire & Rubber Co., 863 P.2d 795, 825 (Cal. 1993) (en banc) (requiring only that substance must be "toxic").

7. O'Neal v. Dep't of the Army of the United States, 852 F. Supp. 327, 336 (M.D. Pa. 1994) ("Because there was no testimony quantifying the probability that any Plaintiff will contract a disease other than cancer, Plaintiffs have failed to prove by a preponderance of the evidence that the well contamination caused them a significantly higher risk of contracting non-cancerous diseases.").

8. *Hansen*, 858 P.2d at 979 ("[T]he plaintiff must prove that the exposure was of sufficient intensity and/or duration to increase his or her risk of the anticipated harm significantly over the plaintiff's risk prior to exposure."); *Bourgeois I*, 716 So. 2d at 360 (Plaintiff must show that he "suffers a significantly increased risk of contracting a serious latent disease."). *Cf. In re* St. Jude Medical, Inc., MDL No. 01-1396 (JRT/FLN), 2003 WL 1589527, at *11 (D. Minn. Mar. 27, 2003) (only requiring "an increased risk of harm" to establish "an injury in fact"); Miranda v. Shell Oil Co., 26 Cal. Rptr. 2d 655, 658 (Cal. Ct. App. 1993) (setting a standard of "the relative toxicity of the chemicals").

9. A "serious" disease has been defined by at least one court as one that results in "significant impairment" or death. *Hansen*, 858 P.2d at 979.

lifestyle factors, as well as alternative sources of exposure).[10] For clarity of nomenclature, we describe this monitoring awardable by courts as "special" medical monitoring.

There is near unanimity among courts that a plaintiff must meet requirements akin to conditions 1, 2, 3(b), 4, and 6 to state a successful claim for medical monitoring. But courts are divided as to whether the plaintiff must show the medical monitoring regime has had demonstrated clinical value (condition 5).[11] And, far more important to this Article's discussion, element (3)(a)—that a plaintiff *quantify* the increase in his risk for a serious disease as a result of defendant's negligent exposure—is almost never a formal requirement, and is even disavowed by some courts.[12] Instead, most judges assess a plaintiff's professed medical monitoring need against "quantitative modifiers" (for example, element 3(b), a "significant" increase in risk).

Judicial use of quantitative adjectives as a proxy for assessing the actual increase in risk is a questionable practice. One pair of authors observes that

---

10. The Pennsylvania Supreme Court and the Third Circuit explained in detail the motivation of the "special" medical monitoring requirement:

> *Paoli II*'s requirement of 'special' medical monitoring implicitly recognizes the longstanding requirement in all tort cases other than those based on the old 'intentional' common law torts for various forms of trespass that a plaintiff must prove an injury before he may recover anything from a defendant. Otherwise, a polluter would become a health care insurer for medical procedures routinely needed to guard persons against some of the ordinary vicissitudes of life. It would convert toxic torts into a form of specialized health insurance. Imposition of liability on this basis seems to go beyond current tort theories of negligence or strict liability by requiring a polluter to pay for medical procedures that the general population should receive. Thus, *Paoli II* requires plaintiffs to show not only that their exposure to toxic substances is greater than normal background levels, but that the increased risk of injury from such exposure warrants medical monitoring against future illness beyond that which is recommended for everyone.

Redland II, 696 A.2d at 144 (citations omitted in original) (quoting *Redland I*, 55 F.3d at 846 n.8). *See also* Lockheed Martin Corp. v. Superior Court, 63 P.3d 913, 922 (Cal. 2003) (quoting Potter v. Firestone Tire & Rubber Co., 863 P.2d 795, 824-25 (Cal. 1993) (en banc)) ("[C]ourts confronting medical monitoring claims may consider 'the relative increase in the chance of onset of disease in the exposed plaintiff as a result of the exposure, when compared to (a) the plaintiff's chances of developing the disease had he or she not been exposed, and (b) the chances of the members of the public at large of developing the disease.'"); *Potter*, 863 P.2d at 825 (quoting *Miranda*, 26 Cal. Rptr. 2d at 660) (holding toxic exposure plaintiffs may recover "only if the evidence establishes the necessity, as a direct consequence of the exposure in issue, for specific monitoring beyond that which an individual should pursue as a matter of general good sense and foresight").

11. *See infra* Part III.B.1.

12. *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 788 (3d Cir. 1994) ("Paoli II") ("Nor do we think that an expert must quantify the increased risk."); Merry v. Westinghouse Elec. Corp., 684 F. Supp. 847, 850 (M.D. Pa. 1988) (criticizing a quantification of risk requirement because "[w]e think this formulation unduly impedes the ability of courts to recognize that medical science may necessarily and properly intervene where there is a significant but unquantified risk of serious disease").

the court[s] self-consciously rel[y] on a series of quantitative modifiers . . . in an effort to reserve liability for truly deserving cases. Anyone familiar with modern American trial practice will understand that, however well-meaning, this reliance on superlatives will not prevent most well-prepared cases from reaching triers of fact. There is no escaping the conclusion that defendants in these medical monitoring cases face potentially crushing liabilities.[13]

Yet no court to our knowledge, even among those that explicitly advocate quantification,[14] has recommended a quantitative methodology useful in determining whether a claim warrants medical monitoring. What is more, it is evident that courts *desire* such a numerical "litmus test" to better judge when an increase in risk due to tortious exposure makes medical monitoring proper.[15] Part of this newfound interest in quantification may be that courts are just now coming to a realization, long understood in medical circles, that medical monitoring is itself a potentially harmful intervention in patients' lives, and must be wielded, speaking strictly from a medical standpoint, quite carefully.[16] Where now the courts are clearly struggling to zero in on the level of risk increase believed necessary to medically warrant monitoring by invoking imprecise, vague, and qualitative assessments—one leading court, for instance, has apparently adopted the double-bounded adjectival approach, decreeing that a plaintiff must demonstrate "a significant but not necessarily likely risk of serious disease"[17]—better legal certainty about the medical propriety of a decision to permit monitoring should, must, and can be attained.

To remedy this evident deficit, we first review the purposes of medical monitoring, as understood by health care specialists. Thereafter, we present a simple quantifiable metric, used widely by medical professionals, to approximately gauge the efficacy of a proposed monitoring regime.

---

13. James A. Henderson, Jr. & Aaron D. Twerski, *Asbestos Litigation Gone Mad: Exposure-Based Recovery for Increased Risk, Mental Distress, and Medical Monitoring*, 53 S.C. L. REV. 815, 845 (2002).

14. *See, e.g.*, O'Neal v. Dep't of the Army of the United States, 852 F. Supp. 327, 334 (M.D. Pa. 1994) (rejecting significance of quantified risk, because "[w]ithout regard to toxic exposure, a person has a 25 per cent chance of contracting cancer during his lifetime . . . . Employing [the plaintiff expert's] estimate, Plaintiffs' lifetime risk of contracting cancer increased because of their well-water exposure from 25 per cent to 25.03 per cent . . . . Plaintiffs' risk of contracting cancer because of their well-water exposure increased by approximately the same amount as if, for instance, they had each smoked seven packs of cigarettes, lived in New York City or Boston for 200 days or driven 500 miles each year"); *Potter*, 863 P.2d at 824-25 ("In determining the reasonableness and necessity of monitoring . . . [a court must consider] the relative increase in the chance of onset of disease in the exposed plaintiff as a result of the exposure, when compared to (a) the plaintiff's chances of developing the disease had he or she not been exposed, and (b) the chances of the members of the public at large of developing the disease.").

15. *C.f.* JAMES A. HENDERSON, JR. ET AL., THE TORTS PROCESS 231-57, 338-53 (6th ed. 2003); Metro-North Commuter R.R. Co. v. Buckley, 521 U.S. 424, 433 (1997) (noting that absent some workable standard, medical monitoring claims could threaten a "flood" of less important cases).

16. *See infra* Part II.

17. Lockheed Martin Corp. v. Superior Court, 63 P.3d 913, 921 (Cal. 2003) (quoting *Potter*, 863 P.2d at 824).

I. DEFINING MEDICAL MONITORING

A. What Medical Monitoring Is

Medical monitoring is the periodic application (medical *screening* is a single application) of a diagnostic medical examination or test to detect a latent disease or condition.[18] The object of monitoring is to find a disease before symptoms arise that would prompt the patient to seek medical care resulting in a typical clinical diagnosis.[19] Thus, a subject being medically monitored is asymptomatic for the target disease—the person does not exhibit easily ascertainable signs of disease and does not complain of symptoms.[20] Finally, because courts hold that monitoring costs can be awarded only for diseases whose risk was "significantly increased" by the tortious exposure,[21] this Article uses interchangeably the terms *medical monitoring* and *exposure-prompted medical monitoring.*[22]

---

18. *See, e.g.*, U.S. PREVENTIVE SERVICES TASK FORCE [hereinafter USPSTF], GUIDE TO CLINICAL PREVENTIVE SERVICES 187-91, 209-18 (2d ed. 1996). Although some courts, just like some medical authorities, have misconstrued the proper definition of medical monitoring, there are legal sources that provide accurate, general definitions of medical monitoring. *See* Allan L. Schwartz, Annotation, *Recovery of Damages for Expense of Medical Monitoring to Detect or Prevent Future Disease or Condition*, 17 A.L.R. 5th 327, 340 (1994) (stating that courts have "defined a medical monitoring claim as a claim for the costs of periodic medical examinations to detect latent diseases or disorders caused by a defendant's culpable conduct, the object of which is to facilitate early diagnosis and treatment of such diseases or disorders").

19. *See, e.g.*, USPSTF, *supra* note 18, at 187-91, 209-18; J. M. G. Wilson & G. Junger, *Principles and Practice of Screening for Disease* (World Heath Organization, Public Health Paper No. 34, 1968), *at* http://whqlibdoc.who.int/php/WHO_ PHP _34.pdf (last visited Jan. 27, 2004).

20. Wilson & Junger, *supra* note 19. While many courts have recognized that medical monitoring, as distinguished from *medical management*, is given only to asymptomatic individuals, some courts still fail to recognize this distinction. Compare Ball v. Joy Techs., Inc., 958 F.2d 36, 39 (4th Cir. 1991), Cook v. Rockwell Int'l Corp., 755 F. Supp. 1468, 1477 (D. Colo. 1991), Potter v. Firestone Tire & Rubber Co., 863 P.2d 795, 822-23 (Cal. 1993), Petito v. A.H. Robins Co., 750 So. 2d 103, 105-06 (Fla. Dist. Ct. App. 1999), Bourgeois v. A.P. Green Indus., 716 So. 2d 355, 359 (La. 1998) ("Bourgeois I"), Meyerhoff v. Turner Constr. Co., 534 N.W.2d 204, 206 (Mich. Ct. App. 1995), Ayers v. Township of Jackson, 525 A.2d 287, 312 (N.J. 1987) ("Ayers III"), Hansen v. Mountain Fuel Supply Co., 858 P.2d 970, 977-78 (Utah 1993), Bower v. Westinghouse Elec. Corp., 522 S.E.2d 424, 430 (W. Va. 1999) (failing to make a proper distinction), with Barnes v. Am. Tobacco Co., 161 F.3d 127, 139 (3d Cir. 1998), Witherspoon v. Philip Morris, Inc., 964 F. Supp. 455, 467 (D.D.C. 1997), Burton v. R.J. Reynolds Tobacco Co., 884 F. Supp. 1515, 1523 n.6 (D. Kan. 1995) ("[A] separate tort action for medical monitoring arises in an instance where a plaintiff . . . may increase his or her susceptibility to contracting a latent disease, but . . . has not yet suffered any injuries."), Abney v. Exxon Corp., 755 So. 2d 283, 290 (La. Ct. App. 1999). Philip Morris, Inc. v. Angeletti, 752 A.2d 200, 250-51 (Md. 2000) (quoting Redland Soccer Club, Inc. v. Dep't of the Army of the United States, 696 A.2d 137, 144 (Pa. 1997) ("Redland II") ("The injury in a medical monitoring case not being a physical one, it is instead construed as the 'costs of periodic medical examinations necessary to detect the onset of physical harm.'")).

21. *In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 852 (3d Cir. 1990) ("Paoli I"). *See also In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 787 (3d Cir. 1994) ("Paoli II") (affirming *Paoli I*, 916 F.2d 829); Potter v. Firestone Tire & Rubber Co., 863 P.2d 795, 825 (Cal. 1993) (en banc); Petito v. A.H. Robins Co., 750 So. 2d 103, 106 (Fla. Dist. Ct. App.

B. What Medical Monitoring Is Not

There is a lack of consistency among various fields, among various authors, and among various courts in the use of medical monitoring terminology. Medical monitoring, as we mean it, has been described elsewhere with the use of such terms as *biologic monitoring, medical surveillance, medical screening,* or *medical supervision.* Confusion arises because these same terms have been used to classify activities that are decidedly not the medical monitoring we are discussing. One medical author acknowledges that the occupational medicine literature is littered with a confusing mélange of terms lacking consensual validation that describe disease prevention activities.[23] Medical monitoring for the early detection of cancer or any other illness in asymptomatic persons allegedly having received a toxic chemical exposure is clearly distinguished from:

(1) (a) carrying out the usual recommendations concerning health care for the general population;[24] or (b) carrying out recommended monitoring procedures for individuals at special risk due to non-exposure factors, including heredity and lifestyle.[25] Courts too do not grant medical monitoring if a diagnostic testing regime fits wholly within either or both of these two subclasses of testing.[26]

(2) "medical management," which describes any clinical procedure (e.g., additional history taking, physical examinations, laboratory tests, etc.) that may be required for a patient *with a symptom, sign, or diagnosis of disease.* Medical management is strictly different from medical monitoring, which may well amount to performing some of the same test(s) on a patient *not* yet complaining of or diagnosed with disease. The difference between "management" and "monitoring" is not semantics. Interpretation of test results, follow-up protocols, and treatment recommendations will seldom be the same or even similar for the patient who already has complaints or symptoms and the one who does not.[27]

(3) "medical surveillance," which is defined in occupational medicine as the collection, analysis, and dissemination of results for the purpose of prevention.[28] The data collected for surveillance may come from

---

1999); Bourgeois v. A.P. Green Indus., 716 So. 2d 355, 360 ("Bourgeois I") (La. 1988); Redland Soccer Club v. Dep't of the Army of the United States, 696 A.2d 137, 144 (Pa. 1997); Hansen v. Mountain Fuel Supply Co., 858 P.2d 970, 979 (Utah 1993). *Cf.* Lockheed Martin Corp. v. Superior Court, 63 P.3d 913, 921 (Cal. 2003) (requiring "a significant but not necessarily likely risk of serious disease"); Miranda v. Shell Oil Co., 26 Cal. Rptr. 2d 655, 658 (Cal. Ct. App. 1993) (requiring a "relative increase" in risk); Meyerhoff v. Turner Constr. Co., 534 N.W.2d 204, 206 (Mich. Ct. App. 1995) (same).

22. *See, e.g.,* USPSTF, *supra* note 18, at 187-91, 209-18; Wilson & Junger, *supra* note 19.

23. Michael Gochfeld, *Medical Surveillance and Screening in the Workplace: Complementary Preventive Strategies,* 59 ENVTL. RES. 67, 69-71 (1992).

24. *See, e.g.,* USPSTF, *supra* note 18, at 187-91, 209-18.

25. *Id.*

26. *See supra* note 10 and accompanying text.

27. Wilson & Junger, *supra* note 19. *See also* Friends for All Children, Inc. v. Lockheed Aircraft Corp., 746 F.2d 816, 826 (D.C. Cir. 1984) (holding that medical monitoring claims are valid in the absence of any apparent physical injury).

28. Occupational medicine specialists are engaged in many types of disease prevention activities that lack a consensually validated nomenclature and are often subsumed under the term "medical surveillance." In contrast, medical monitoring is a narrow class of activity.

environmental monitoring (testing environmental samples for chemicals to assess the amount of human exposure),[29] *biologic monitoring* (testing such human biological samples as blood, breath, or urine for a toxicant or its metabolite to assess the amount of dose an individual has received),[30] *case finding* ("advice provided to a single patient presenting to a doctor on account of some other problems"[31]), or *epidemiological research* (tabulations of diagnosed illness).[32]

(4) Legally and economically, a cause of action for medical monitoring differs from the cause of action commonly referred to as an "enhanced risk" claim, which "seeks compensation for the anticipated harm itself, proportionately reduced to reflect the chance that it will not occur."[33]

One author defines medical surveillance "as the application of an examination, historical question, or laboratory test to apparently healthy persons with the goal of detecting absorption of intoxicants or early pathology before the worker would normally seek clinical care for symptomatic disease." William E. Halperin et al., *Medical Screeing in the Workplace: Proposed Principles*, 28 J. OCCUPATIONAL. MED. 547, 548 (1986). This definition combines several categories of disease detection programs including epidemiologic surveys, screening, and case finding. However the authors then note that "[p]urists would argue that the detection of intoxicants or their metabolites is biological monitoring and not medical screening, which implies the search for pathology" and that "others would distinguish between testing the workers for biologic markers of susceptibility and detecting effects of exposure." *Id.* Clearly, medical monitoring should be distinguished from this definition of medical surveillance. Indeed, these authors employ a slightly different definition of medical surveillance that would not include medical monitoring when they state, "[s]urveillance is the 'continued watchfulness over the distribution and trends of incidence through the systematic collection, consolidation, and evaluation of morbidity and mortality reports and other relevant data.'" *Id.* at 547-48 (citing A.D. Langmuir, *The Surveillance of Communicable Diseases of National Importance*, 268 NEW ENG. J. MED. 182-92 (1963)), and, "Surveillance comes late in the hierarchy of prevention of occupational disease because the *disease must have already occurred* in order to be counted." *Id.* at 548 (emphasis added). Others in the field use medical surveillance in a more restrictive sense concerning assessment of exposure rather than of pre-symptomatic disease. "Medical surveillance: the longitudinal evaluation of potentially exposed people for early detection of biochemical or pathophysiologic changes indicative of significant exposure." Gochfeld, *supra* note 23, at 76-77.

29. Gochfeld, *supra* note 23, at 76 ("Environmental monitoring: the use of industrial hygiene methods to evaluate a workplace environment with respect to potential exposures").

30. *Id.* (Biological monitoring is "[t]he direct testing for a chemical agent or its metabolite in the body, or in the broader sense (sensu latu) including the measurement of specific biomarkers or specific biochemical or pathophysiologic effects. It should be used only as a component of a medical [monitoring] program."); William E. Halperin & Todd M. Frazier, *Surveillance for the Effects of Workplace Exposure*, 6 ANN. REV. PUB. HEALTH 419 (1985). An example of biologic monitoring (as contrasted with medical monitoring) is the use of blood testing to measure the amount of lead in the blood of exposed workers.

31. John J. McNeil & Paul E. O'Brien, *Screening, Case Finding and Evidence-based Guidelines*, 171 MED. J. AUSTRALIA 344, 344 (1999).

32. *See generally* MODERN EPIDEMIOLOGY (Kenneth J. Rothman & Sander Greenland eds., 2d ed. 1998); DAVID L. SACKETT ET AL., CLINICAL EPIDEMIOLOGY: A BASIC SCIENCE FOR CLINICAL MEDICINE (2nd ed. 1991).

33. *See* Schwartz, *supra* note 18, at 342. *See also In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 850 (3d Cir. 1990) ("Paoli I"); Bourgeois v. A.P. Green Indus., 716 So. 2d 355, 359 n.12 ("Bourgeois I"); Hansen v. Mountain Fuel Supply Co., 858 P.2d 970, 975 n.6 (Utah 1993) (referring to Ayers v. Township of Jackson, 525 A.2d 287, 297-315 (N.J. 1987) ("Ayers III").

II. THE HARMS OF MONITORING

A. Testing Error

To the layperson, medical monitoring may appear a universally sensible way to medically confront situations of alleged increased risk of future disease, particularly, dreaded ailments like cancer.[34] People know the aphorism, "an ounce of prevention is worth a pound of cure," and may conclude, without reflection, that a targeted, early intervention must be worthwhile. Further, a periodic examination by a physician may convey a sense of caring and establish a bond of friendship and trust between doctor and patient even though there is little evidence that unsuspected disease will be detected frequently in the asymptomatic patient.[35] However, it cannot be stressed greatly enough that medical monitoring is a medical intervention into a patient's life, qualitatively similar to other potentially harmful interventions such as starting a patient on a medication or performing surgery. As one physician-scientist has described the proposition:

> [Some] have described screening for symptom-free individuals in terms of an implicit contract. By this view, we imply a promise to asymptomatic people not that their subsequent treatment may work but that it does work; "not that we will simply do our best, but that they will be better off as a result of the screening program. When we impose ourselves on the public in this fashion, we thus require very firm evidence that our early diagnosis and subsequent therapy will do more good than harm."[36]

Because this pseudo-"social contract" requires physicians to "do more good than harm," it becomes necessary to gain some intuition of how much harm any such monitoring will cause relative to however much benefit is gained by stymieing or slowing the progress of latent disease. In essence, a physician must quantify risk so that he is able to balance the costs and benefits of his intervention—medical monitoring.[37]

---

34. *See* Merry v. Westinghouse Elec. Corp., 684 F. Supp 847, 850-51 (M.D. Pa. 1988) (approving medical monitoring, because "there is not a serious question of the value of early detection and treatment of cancer . . . . Given that these persons were exposed to multiple chemicals which [sic] can cause cancer and other diseases, a medical surveillance program is essential.").

35. Sylvia K. Oboler & F. Marc LaForce, *The Periodic Physical Examination in Asymptomatic Adults*, 110 ANNALS INTERNAL MED. 214, 215 (1989).

36. Jerome E. Herbers, Jr., *Screening with Prostate-specific Antigen: Should We Or Shouldn't We?*, 269 JAMA 2212, 2212 (1993). *See also* Michael A. Goldenhersh, *Melanoma Screening: Critique and Proposal*, 28 J. AM. ACAD. DERMATOLOGY 642, 643 (1993) ("[M]ass public screening . . . is an intervention in the lives of healthy people. It must be proved to be beneficial. Especially in a screening situation, in which the doctor has initiated the contact and the patient therefore assumes that he will benefit, it behooves the doctor to know with certainty that the patients to whom the service is offered will indeed benefit."), *citing* SACKETT ET AL., *supra* note 32, at 153-70.

37. *Hansen*, 858 P.2d at 980 ("[I]f a reasonable physician would not prescribe it for a particular plaintiff because the benefits of the monitoring would be outweighed by the costs, which may include, among other things, the burdensome frequency of the monitoring

It is true, as with all proven medical interventions, that medical monitoring carries the potential to be beneficial. However, implementing any kind of medical screening or monitoring program may do more harm than good. For example, even in the familiar contexts of general screening for diabetes,[38] or cancers of the prostate,[39] skin,[40] ovaries,[41] and lung,[42] physicians are rethinking the putative benefit of monitoring asymptomatic populations for disease.[43]

---

procedure, its excessive price, or its risk of harm to the patient, then recovery would not be allowed.").

38. Brent Lindahl et al., *Screening for Impaired Glucose Tolerance*, 22 DIABETES CARE 1988, 1988 (1999) (concluding that "a high-risk screening strategy for [impaired glucose tolerance] targeted solely towards subjects with obesity and/or heredity background of diabetes will fail to detect the majority of subjects . . . in the general population"); Daniel E. Singer et al., *Screening for Diabetes Mellitus*, 109 ANNALS INTERNAL MED. 639, 639 (1988) (concluding that screening for Type II diabetes in non-pregnant individuals is "not recommended" and that screening for gestational diabetes produces only a "small expected benefit").

39. Larry Katzenstein, *Can the Prostate Test Be Hazardous to Your Health?*, N.Y. TIMES, Feb. 17, 1999, at G4 ("[T]here are serious concerns about the [Prostate Specific Antigen blood] test's usefulness and whether the treatment for prostate cancer may be harming more lives than it saves. Despite a recent barrage of high-profile endorsements for the test by Arnold Palmer and H. Norman Schwarzkopf, among others, not one major medical or public-health group endorses the screening. And in recent years, most of the groups that have evaluated the test either oppose its use for routine screening, or do not recommend it. These include the National Cancer Institute, the American College of Physicians, the American College of Preventive Medicine and the United States Preventive Services Task Force."); Otis W. Brawley, *Prostate Carcinoma Incidence and Patient Mortality*, 80 CANCER 1857, 1857 (1997) ("[S]creening and early detection efforts are resulting in the diagnosis of prostate carcinoma in some men who do not need therapy; thus, prostate carcinoma screening can lead to unnecessary treatment for such men. Furthermore, epidemiological data do not demonstrate that screening is decreasing mortality."); Margaret T. Mandelson et al., *PSA Screening: A Public Health Dilemma*, 16 ANN. REV. PUB. HEALTH 283, 298-99 (1995) (extensive survey of the medical literature concluding that "evidence of benefit of early detection on prostate cancer mortality has not been demonstrated"); Herbers, *supra* note 36, at 2212. *Cf.* Charles R. Smart, *The Results of Prostate Carcinoma Screening in the U.S. as Reflected in the Surveillance, Epidemiology, and End Results Program*, 80 CANCER 1835, 1835 (1997) (finding indirect evidence to suggest that prostate carcinoma screening of men age fifty and greater decreased the incidence of distant disease, which influences mortality rates).

40. Susan Dozier et al., *Beachfront Screening for Skin Cancer in Texas Gulf Coast Surfers*, 90 S. MED. J. 55, 55 (1997) (indicating that directed skin cancer screening of an at-risk population is more productive in finding skin cancer than screening of a self-selected population); Goldenhersh, *supra* note 36, at 642-43 ("Although the interest in screening for melanoma is understandable, one must first ask, 'Does melanoma screening save lives?' . . . [Screening] has its drawbacks. It involves time, expense, and inconvenience; it provokes anxiety and causes embarrassment . . . . Screening results in 'overdiagnosis,' generating removal of some nevi, that did not need removal, which further compounds the anxiety and the cost. Furthermore, a significant lesion may be overlooked during screening or a new one may develop shortly thereafter, and the screening examination may give the patient a false sense of security causing him not to return for medical attention later when he needs it.") (footnotes omitted). *See also* USPSTF, *supra* note 18, at 141 ("There is insufficient evidence to recommend for or against either routine screening for skin cancer by primary care providers" or "counseling patients to perform periodic self-examination of this skin.").

41. Marilyn M. Schapira et al., *The Effectiveness of Ovarian Cancer Screening: A Decision Analysis Model*, 118 ANNALS INTERNAL MED. 838, 838 (1993) ("Given the limited

The reason for these reconsiderations is simple but unintuitive. All diagnostic testing has some rate of error. Test results can err by being either false-positive (the test result indicates disease even though the person is healthy) or false-negative (the test result indicates the person is healthy even though disease is present). Such inaccuracies create ambiguities for proper medical follow up. For one, even the most unlikely positive result must be pursued to determine if it is a true-positive or a false-positive.[44] The costs of sequential testing to confirm a positive test result can be enormous. Complications stemming indirectly from testing are so important that they merit extended discussion below.[45] Psychological anguish[46] and financial

---

effect on overall life expectancy, it is unlikely that mass screening for ovarian cancer with CA 125 and transvaginal sonography would be an effective health policy."). The U.S. Preventive Task Force has concluded that "there is fair evidence to support the recommendation that [ovarian cancer screening] be excluded from consideration in a periodic health examination." USPSTF, *supra* note 18, at 861, 866.

42. The U.S. Preventive Task Force has concluded that "there is fair evidence to support the recommendation that [lung cancer screening] be excluded from consideration in a periodic health examination." USPSTF, *supra* note 18, at 861, 865; Victor R. Grann & Alfred I. Neugut, *Lung Cancer Screening at Any Price*, 289 JAMA 357, 358 (2003) ("Until more data are available and the NCI randomized trial is completed, physicians, patients, and policy makers should be conservative about accepting this new, as-yet not fully tested, and relatively very expensive strategy of using helical CT scanning in screening for lung cancer."). *See infra* notes 84 & 113.

43. In reflecting on the results of a randomized trial which showed that early and intensive intervention to prevent readmission to the hospital increased costs and may have made patients sicker, a New England Journal of Medicine editorialist concluded that the value of early intervention may be wrong:

> Instead of conferring benefit, closer scrutiny of the patients simply led to more medical care and perhaps to harm . . . . This is not the first time the doctrine of early intervention has been challenged; other challenges to the doctrine have emerged, especially with respect to screening. There is now vigorous debate about the benefit of detecting and treating early-stage prostate cancer, a similar debate is developing with regard to the treatment of ductal carcinoma in situ of the breast. Such screening leads to interventions. In exchange for unclear benefits, it brings aggravation and fear to asymptomatic people who otherwise would believe they were in good health. The early-intervention doctrine has also been challenged recently in the context of surveillance, with the value of increased efforts to detect recurrence being seriously questioned in patients with melanoma and disproved in patients with breast cancer. Our belief in early intervention is due in part to its intuitive appeal: disease found early should be easier to eradicate . . . . The findings of Weinberger *et al.* suggest that our assumptions about early therapeutic intervention should be questioned more broadly.

H. Gilbert Welch, *Questions about the Value of Early Intervention*, 334 NEW ENG. J. MED. 1472, 1472 (1996). *Accord* Barron H. Lerner, *Screening for Cancer: A Downside*, N.Y. TIMES, Mar. 5, 2002, at F6.

44. One cannot discount all positives because of a low likelihood that any given positive is correct. Otherwise, why decide to monitor in the first instance?

45. *See supra* notes 39-43; *see infra* Part II.A-B.

46. *See generally* O. Gustafsson et al., *Psychological Reactions in Men Screened for Prostate Cancer*, 75 BRIT. J. UROLOGY 631 (1995); Caryn Lerman et al., *Adverse Psychologic Consequences of Positive Cytologic Cervical Screening*, 165 AM. J.

burdens[47] too can hardly be discounted. False-negative test results also are problematic because they can improperly reassure the patient, and thus deter patient initiative in bringing new symptoms to medical attention.[48] Indeed, randomized clinical trials demonstrate that fewer individuals request medical monitoring once they observe through these quantified benefits and costs that monitoring is certainly not a costless procedure.[49]

B. The Cascade Effect: Lessons from Brer Rabbit and the Tar Baby

Sometimes diagnostic tests have the potential to exact direct harm, such as damage to the fetus during amniocentesis or colon perforation during colonoscopy.[50] But often, indirect consequences of error are equally or even more vexing than direct harms caused by monitoring.

The monitoring process starts out innocuously enough; often what prompts medical monitoring is not that it is necessarily warranted medically, but that its awarder is anxious or uncertain about the (unquantified) possibility that exposed individuals may become ill.

> The first step . . . appears to be a benign action, such as ordering a diagnostic test; however, the discovery of an unexpected abnormality leads to progressively riskier and costlier interventions that seem simultaneously unnecessary and unavoidable. [T]his situation [is likened] to the story of Brer Rabbit and the tar baby. Brer Rabbit greets the tar baby and takes its failure to communicate as snobbery. He hits the tar baby to teach it a lesson and gets stuck. In an effort to make the tar baby let go, he hits it with the other hand, then kicks it. With each

---

OBSTETRICS & GYNECOLOGY 658 (1991); Caryn Lerman et al., *Cancer Risk Notification: Psychosocial and Ethical Implications*, 9 J. CLINICAL ONCOLOGY 1275 (1991); Caryn Lerman et al., *Psychological and Behavioral Implications of Abnormal Mammograms*, 114 ANNALS INTERNAL MED. 657 (1991); Caryn Lerman et al., *Psychosocial Impact of Cancer Screening*, 7 ONCOLOGY 67 (1993); Theresa M. Marteau, *Psychological Costs of Screening*, 299 BRIT. MED. J. 527 (1989); N.F. Reelick et al., *Psychological Side-Effects of the Mass Screening on Cervical Cancer*, 18 SOC. SCI. MED. 1089 (1984); Jane Wardle et al., *False Positive Results in Ovarian Cancer Screening: One Year Follow-up of Psychological Status*, 10 PSYCHOLOGY AND HEALTH 33 (1994); Jane Wardle & Rosemary Pope, *The Psychological Costs of Screening for Cancer*, 36 J. PSYCHOSOMATIC RES. 609 (1992).

47. Ayers v. Township of Jackson, 493 A.2d 1314, 1323 (N.J. Super. Ct. App. Div. 1985) ("Ayers II") ("Faced with the admitted inability of the expert witness to quantify the increased risk, we cannot rule out the probability that such increase is so microscopically small as to be meaningless. Without some quantifying guidance it becomes impossible to say that defendant has so significantly increased the 'reasonable probability,' that any of the plaintiffs will develop cancer so as to justify imposing upon defendant the financial burden of lifetime medical surveillance for. early clinical signs of cancer."), *rev'd*, Ayers v. Township of Jackson, 525 A.2d 287 (N.J. 1987) ("Ayers III"); Gina Kolata, *Lung Cancer Test Is Much in Demand, But Benefit Is Murky*, N.Y. TIMES, June 21, 2000 at A1 ("The cost of testing all current and former [asymptomatic] smokers . . . would be immense—more than $39 billion, according to the National Cancer Institute.").

48. USPSTF, *supra* note 18, at xxxi.

49. Robert J. Volk & Stephen J. Spann, *Decision-aids for Prostate Cancer Screening*, 49 J. FAM. PRAC. 425, 425 (2000).

50. Steven H. Woolf & Douglas B. Kamerow, *Testing for Uncommon Conditions: The Heroic Search for Positive Test Results*, 150 ARCHIVES INTERNAL MED. 2451, 2451 (1990).

blow he becomes more ensnared rather than closer to a solution.
Similarly, as physicians try to explain each new ambiguous result, they
become more ensnared in their own actions.[51]

The term "cascade effect" describes this chain reaction of unwanted outcomes that
is triggered by good intentions, but that "progresses inexorably to a ridiculous
end."[52]

But how can mere testing foment these "ridiculous ends" of the cascade effect?
Richard Deyo describes one example: electronic monitoring for fetal heart rate
while women are in labor.[53] To achieve accurate readings using such a device, the
mother-to-be must remain relatively inactive in bed. This stasis involuntarily forces
her to slow her labor. The concerned and observant physician, noting that her labor
has slowed, will often intervene to speed up the labor (for example, by rupture of
the membranes). This acceleration of labor in turn causes increased pressure on the
baby's skull, which causes abnormal fetal heart rate readings. The physician-
induced labor acceleration simultaneously increases the pain of contractions for the
mother and may prompt her to request epidural anesthesia. If the doctor then gives
systemic anesthesia, the mother's blood pressure may lower, causing even more
abnormal readings on the electronic heart rate monitor. This cascade effect is
thought to explain why "Cesarean sections rates are 40% higher when electronic
monitoring is used rather than simple auscultation of the fetal heart rate."[54]

Other examples of the perverse outcomes from monitoring include "a lifelong
cascade of clinical events" such as "multiple blood tests, urinary hormone assays,
and repeat imaging tests" for those who screen positive for adrenal gland tumors.[55]
Use of cardiac catheterization to monitor cardiac disease in patients who had

---

51. Richard A. Deyo, *Cascade Effects of Medical Technology*, 23 ANN. REV. PUB.
HEALTH 23, 24 (2002). *See also* K. Patrick Ober, *Uncle Remus and the Cascade Effect in
Clinical Medicine: Brer Rabbit Kicks the Tar Baby*, 82 AM. J. MED. 1009 (1987).

52. James W. Mold & Howard F. Stein, *The Cascade Effect in the Clinical Care of
Patients*, 314 NEW ENG. J. MED. 512, 512 (1986).

53. Deyo, *supra* note 51, at 23.

54. *Id.* at 26. Mold and Stein relate the case history of a gentleman admitted to a
hospital for elective repair of an inguinal hernia. The patient had suffered previously from
coronary disease. The surgeon, after learning of the patient's medical history, was concerned
about complications arising during surgery and requested a cardiology consultation. The
cardiologist, perhaps uncertain about his own clinical assessment of the patient, suggested an
exercise tolerance test. The patient was forced to wait for six hours for the test because of
backlogs. In the interim, he became angry, irritated, and anxious, and suffered mild chest
discomfort. Because of his chest discomfort, the exercise test was not performed and he was
instead transferred to a telemetry unit. The patient became even more irritable and agitated.
He was observed to have undergone some electrocardiogram (heart monitor) changes, and
received appropriate medication. At this point, doctors elected to give him a cardiac
catheterization (which actually showed his heart condition to have improved since his
previous medical consultation). By now, the elective hernia repair—the reason he had come
to the hospital in the first place—had to be postponed for two weeks because of a full
operating room schedule. The surgeons were left hoping to convince the patient that he had
no reason to worry about the delay in his surgery. Mold & Stein, *supra* note 52, at 512.

55. The reason for this lifelong testing is that "benign tumors of the adrenal glands
(adenomas) are quite common and have been reported in almost 9% of patients in autopsy.
Adrenal carcinomas are extremely rare, in contrast, with an incidence of 0.0004% per year.
Unfortunately the readiographic appearance [of the two forms of tumor] . . . is generally
indistinguishable." Deyo, *supra* note 51, at 25.

previously suffered heart attacks has been suggested to actually lead to more deaths than in an unmonitored group with approximately equal coronary artery disease histories.[56] "[T]elling a patient that he or she has a herniated disk may result in illness-related behavior and absence from work," and unnecessary (and potentially fatal)[57] spinal surgery, which is often prescribed for clinically irrelevant MRI findings.[58] Indeed, even where monitoring is nearly universally accepted, such as use of mammography for breast cancer in asymptomatic women, cascade effects can materialize.[59] For reference, some examples of commonly reported cascade effects are summarized in the table below.

EXAMPLES OF OFT-REPEATED CASCADES WITH A SUBSTANTIAL RISK OF ADVERSE OUTCOMES[60]

Initial Test	Potential Consequences
Head, body scans	Endocrine incidentalomas
Electronic fetal monitoring	Unnecessary Cesarean sections
Coronary angiography in low-risk patients	Unnecessary invasive coronary interventions
Spinal MRI in the absence of sciatica	Unnecessary spine surgery
Pulmonary artery catheters	Tampering with cardiac physiology
Persistent testing in persons near end of life	Unwanted aggressive interventions

Cascade effects aside, recommended testing as a part of an individual's general health maintenance program does require, from time to time, that novel monitoring programs should necessarily be developed in response to a person's alleged exposure to chemicals. But as a result of monitoring's potential to exact tremendous harm as well as good, it becomes all the more apparent that the decision to medically monitor must be evidence-based. It must be evaluated by a

56. Edward Guadagnoli et al., *Variation in the Use of Cardiac Procedures After Acute Myocardial Infarction*, 333 NEW ENG. J. MED. 573, 573 (1995).

57. As a general matter, a congressional subcommittee found that 2.4 million unnecessary surgeries, ranging from hysterectomies to pacemaker insertions to coronary bypass operations, occur each year in the United States. 11,900 people die yearly as a direct result of those surgeries, and the cost exceeds $3.9 billion. Lucian L. Leape, *Unnecessary Surgery*, 13 ANN. REV. PUB. HEALTH 363, 363 (1992). Similarly, Starfield reports that annually in the U.S., there are 12,000 deaths from unnecessary surgery, 106,000 deaths from adverse drug effects that are not related to improper patient usage, and 80,000 deaths from hospital-spawned infection. Barbara A. Starfield, *Is U.S. Health Really the Best in the World?*, 284 JAMA 483, 483-84 (2000).

58. Richard A. Deyo, *Magnetic Resonance Imaging of the Lumbar Spine–Terrific Test or Tar Baby?*, 331 NEW ENG. J. MED. 115, 115 (1994). The same effect has been observed in individuals screened and diagnosed as hypertensive. R. Brian Haynes et al., *Increased Absenteeism from Work after Detection and Labeling of Hypertensive Patients*, 299 NEW ENG. J. MED. 741, 743 (1978).

59. "An estimate of 25 percent has been given for the cumulative risk of a false positive result after 10 mammograms and 10 clinical breast examinations." Joann G. Elmore et al., *Ten-Year Risk of False Positive Screening Mammograms and Clinical Breast Examinations*, 338 NEW ENG. J. MED. 1089, 1089 (1998); J. Mark Elwood et al., *Breast Cancer Screening with Mammography*, 341 LANCET 1531, 1531 (1993).

60. Table taken from Deyo, *supra* note 51, at 25.

critical quantitative method, weighing the expected benefits of monitoring against its expected risks, and taking into account the many individual factors that may pertain to an exposed population.[61]

III. HOW TO DECIDE WHETHER TO MEDICALLY MONITOR

Unfortunately, there is no universal protocol for monitoring/screening a population for an array of possible future health conditions. Therefore, each target disease must be evaluated separately through an evidence-based demonstration that the monitoring decision is appropriate.[62] Yet the seriousness of a medical monitoring decision cannot be overestimated. It should be known or reasonably expected that detection of the disease in asymptomatic persons through medical monitoring will provide, on average, a better clinical outcome than conventional diagnosis and treatment in patients presenting with symptoms. Depending upon the amount of exposure and resultant increase in risk of disease incurred,[63] and the disease in question, this may or may not be true.

There are two means by which to determine the propriety of medical monitoring. One always should first refer to existing government and academic medical publications for guidance. If, as is often the case, the reference literature is inapplicable to the decision to monitor a particular population after a particular toxic exposure to chemicals, there are secondary meta-techniques that can either establish or disaffirm the need for monitoring. Each of the two steps is addressed in turn below.

A. *Official Monitoring Recommendations and Authoritative Gold Standards*

Accordingly, and not surprisingly, medicine has built consensus regarding the appropriate methodologies for evaluating the scientific and medical advisability of various proposed screening tests for the public and for groups at special risk. The United States Preventive Services Task Force (a panel of 300 experts convened by the U.S. Public Health Service) and its Canadian counterpart adopted well-designed, randomized controlled studies that have resolved whether persons who receive an intervention experience a better overall clinical outcome than those who

---

61. Donald M. Berwick, *Screening in Health Fairs: A Critical Review of Benefits, Risks and Costs*, 254 JAMA 1492, 1497-98 (1985); David M. Eddy, *Comparing Benefits and Harms: The Balance Sheet*, 263 JAMA 2493, 2493-2505 (1990); Stephen A. Feig, *Mammographic Screening of Women Aged 40-49 Years: Benefit, Risk and Cost Consideration*, 76 CANCER 2097, 2102-05 (1995); Robert H. Pantell et al., *Costs and Benefits of Lead Screening*, 270 JAMA 2054, 2054 (1993); Nicholas J. Wald & Hillary Watt, *When Marginal Costs and Benefits Should be Used in Screening*, 312 BRIT. MED. J. 1041, 1042 (1996); Wilson & Junger, *supra* note 19.

62. Bano v. Union Carbide Corp., 2003 WL 1344884, at *9 (S.D.N.Y. Mar. 18, 2003) ("Requiring medical monitoring is an extraordinary remedy requiring extensive factual research [to determine whether it is to be recommended in a given instance] . . . .").

63. Jackson v. Purdue Pharma Co., 2003 WL 21356783, at *4 (M.D. Fla. Apr. 11, 2003) ("Each potential plaintiff received differing amounts of [exposure and subsequent increase in risk] and will have a different need for medical monitoring. Some may require little or no monitoring and treatment, while others may require a great deal.").

do not receive it.[64] These two organizations assessed published clinical evidence and graded[65] its recommendations for monitoring for nearly 100 preventable target diseases and conditions, and approximately 175 screening tests, interventions, and immunizations.[66] Yet neither the Preventive Services Task Force Guide, nor a comparable survey by the American College of Physicians Screening for Healthy Adults was ever intended to provide an exhaustive list of screening recommendations.[67] Nor do they advise about medical monitoring for persons claiming chemical exposure.[68] Absent such official recommendations, randomized controlled studies—so-called "gold standards"—are often ideal and provide dispositive evidence for (or against) monitoring a disease in general populations. Often these published academic studies post-date the most recent versions of official government preventive screening guides, and reflect the latest advances in diagnostic medicine.

But wherever official guides or available gold standards either (a) do not discuss the advisability of monitoring for a disease or condition or (b) recommend that no monitoring be undertaken for a general population, they are of no use in determining the efficacy of monitoring select populations at increased risk of disease because of specific chemical or environmental exposure. Conversely, if the medical literature *does* recommend monitoring for the population at-large, then courts should not allow a monitoring claim to proceed, because the monitoring is not "special."[69]

---

64. The Task Force's most recent edition of the GUIDE TO CLINICAL PREVENTIVE SERVICES is available at http://odphp.osophs.dhhs.gov/pubs/guidecps/ (last visited Jan. 23, 2004).

65. The Task Force graded the strength of its recommendations for or against preventative interventions as follows:

> A: There is good evidence to support the recommendation that the condition be specifically considered in a periodic health examination.
> B: There is fair evidence to support the recommendation that the condition be specifically considered in a periodic health examination.
> C: There is insufficient evidence to recommend for or against the inclusion of the condition in a periodic health examination, but recommendations may be made on other grounds.
> D: There is fair evidence to support the recommendation that the condition be excluded from consideration in a periodic health examination.
> E: There is good evidence to support the recommendation that the condition be excluded from consideration in a periodic health examination.

USPSTF, *supra* note 18, at 861.

66. *Id.* at xii-x; Richard S. A. Hayward et al., *Preventive Care Guidelines*, 114 ANNALS INT. MED. 758, 758-59 (1991).

67. "The tables are not intended as a complete list of all that should occur during the periodic health examination." USPSTF, *supra* note 18, at lx.

68. "The preventive services examined in this report . . . include only those preventive services that might be performed by primary care clinicians . . . in the context of *routine* health care." *Id.* (emphasis in original). Groups with chemical exposure are only one of virtually limitless hypothetical special risk groups that the guide did not specifically address.

69. *See supra* note 10. Occasionally a gold-standard or official guide will provide monitoring recommendations for people with individual risk factors unrelated to environmental exposure. In such instances, the court must determine whether the plaintiffs in its present case present the same risk factors. If they do, then the plaintiffs should already

Accordingly, the upshot is simply this: If either a gold standard or official advisement that is relevant for a specific group and a specific disease recommend monitoring in the absence of unusual chemical exposure, then a court should not grant medical monitoring. If, however, no applicable informing standard exists, or the guide or standard recommends against monitoring for the general population, these eventualities would not be equivalent to a determination of whether monitoring is appropriate for a specific group of plaintiffs who have suffered an atypical exposure.

B. Deciding Whether to Monitor Absent an Official Recommendation or Gold Standard

Requisitioning, designing, and performing such gold standard studies when none exists, unfortunately, is often met with appreciable logistical problems and costs.[70] As an alternative, experts in monitoring/screening often simulate the outcome of a controlled clinical trial were it to be conducted.

Before examining these second-best prescripts for medical monitoring, it follows from basic principles of toxicology that exposure-prompted monitoring cannot be justified unless it has first been demonstrated that an individual (i) has been *exposed* to chemical(s) in the environment (that is, has come in physical contact with a chemical); (ii) has received a *dose* of these chemicals (that is, the chemical has penetrated the body's barriers to uptake and has been absorbed internally); (iii) has, as a result, incurred an increased *risk* of developing one or more diseases in the future; and (iv) no resultant injury has manifested itself yet.[71] Only after these threshold conditions have been met[72] can one address whether to permit exposure-prompted medical monitoring.

---

have been undergoing monitoring, and the monitoring is not "special," as defined in this article. Accordingly, the court cannot sanction monitoring.

70. Stephen G. Pauker & Jerome Kassirer, *Decision Analysis*, 316 NEW ENG. J. MED. 250, 250 (1987) ("To deal with . . . the tough problems, a physician can search for a properly designed, double-blind controlled study that examined patients of the same age, sex, and race and with the same conditions in the same stage; use an algorithm developed for such patients; use the problem-oriented approach to data gathering and hope that the solution to the problem will emerge; or ask for the help of one or more consultants. The frustrations encountered with all these approaches are familiar to all.").

71. Recall that if a dose has resulted in a manifest injury (i.e., the chemical has been found to be the cause of a diagnosed pathophysiologic condition), then any future required medical testing or treatments would not, by definition, be considered as part of a medical monitoring program. Rather, it would constitute *medical management*. Recall too that this medical distinction has a legal parallel; medical monitoring is considered a stand-alone cause of action for increased risk of future illness in a plaintiff without present illness. In contrast, medical management is compensation for visible injuries caused by a tortious chemical exposure. *See supra* note 27.

72. A discussion of how toxicologic data may be analyzed in a medical causation analysis is largely beyond the scope of this Article. *Exposure* may be assessed by environmental testing (e.g., concentration of a chemical in a water sample) coupled with data on individual factors for a given plaintiff (e.g., water use, age, habits, etc.) as might be obtained from a questionnaire, a personal interview, medical or employment records, a medical examination, or a combination of these and other techniques. While an accurate exposure assessment is indispensable for determining the magnitude of the risk of future disease and, hence, the need for medical monitoring (if any exists), none of this prefatory

To determine the suitability of medical monitoring in a specific case, three core principles, first formulated in a broader context in 1968 by Wilson and Junger for the World Health Organization,[73] must be considered:

ESSENTIAL PRINCIPLES OF MEDICAL MONITORING

(1) Given the *natural history* of the diseases or conditions at issue, will early intervention as a result of medical monitoring improve the clinical outcome for each specific condition? Some cancers and many noncancerous diseases cannot be more effectively treated, even if they are detected early through a

---

work constitutes medical monitoring. The *dose* that may result from an exposure is best determined through the use of biologic testing—that is, quantifying the actual amount of a chemical agent that has entered and been retained in the body. An accurate dose assessment also is indispensable for determining the magnitude of the risk of future disease and, hence, the need for medical monitoring. Such biologic monitoring too, is specifically not medical monitoring. *See supra* note 30. Evaluating the magnitude of the *risk* (if any exists) as a result of an exposure and subsequent dose may involve the process of determining toxicologic causation or a mathematical risk assessment.

73. Wilson and Junger were among the first to assemble a list of prerequisites for the evaluation of screening programs on behalf of the World Health Organization:

- The natural history of the disease, including the development of the latent phase, to clinical disease must be sufficiently known.
- There must be an identifiable latent or early symptomatic stage of the disease.
- There must be an effective treatment for patients suffering from localized disease.
- There must be a generally accepted strategy allowing determination of which patients should be treated and which ones should remain untreated.
- Management of the disease in early stages must have favorable impact on prognosis.
- The tests must be acceptable to the screened population.
- The technique to be used for screening must be effective.
- The disease under study is an important health problem.
- The facilities for further diagnosis and treatment must be available.
- The expenses of the screening must be acceptable.

Wilson & Junger, *supra* note 19, at 19. The Journal of Medical Screening similarly promotes two axioms:

- The early detection of the disease should not be an end in itself. The identification of either trivial or untreatable conditions can cause anxiety and waste resources with no useful practical results. Screening should be concerned only with the detection of preventable diseases or disorders that would otherwise cause significant suffering, disability, or death.
- The value of a screening test needs to be determined before it is introduced into practice. It is important to determine quantitatively the avoidance of disability or premature death that screening will achieve. The benefits can then be set against both the financial costs and the "medical" costs (anxiety, discomfort, adverse effects of investigations, and treatment) so that a dispassionate judgment can be reached.

Nicholas J. Wald, *Editorials*, 1 J. MED. SCREENING 1 (1994).

> medical monitoring program, rather than detected through regular medical channels.[74] The U.S. Preventive Services Task Force has written that "[p]ersons with disease who are detected early should have a better clinical outcome than those who are detected without screening (effectiveness of early detection)."[75]
>
> (2) What are the "*sensitivity*" and "*specificity*" of the proposed medical monitoring tests? The tests must be able to detect the target condition earlier than without screening and with sufficient *accuracy* to avoid producing large numbers of false-positive and false-negative results (efficacy).[76]
>
> (3) What is the expected *prevalence* of each specific condition in the population to be monitored? The prevalence must be high; the rarer the disease or condition, the more common a false test result.[77]

Acceptance of these principles, which are discussed in detail immediately below, is predominant across medical disciplines worldwide.[78] Exposure-prompted medical monitoring permits no exception to this decisionmaking process.[79]

---

74. *See supra* notes 39-43.

75. USPSTF, *supra* note 18, at 187-91, 209-18.

76. *Id.*

77. ATSDR's Final Criteria for Determining the Appropriateness of a Medical Monitoring Program Under CERCLA, 60 Fed. Reg. 38840, 38842 (July 28, 1995) ("The population being screened should be at a significantly high risk for the undiagnosed disease (i.e., the disease should have a sufficiently high prevalence in the population).").

78. For an instructive overview of these issues, see Craig Nielsen & Richard S. Lang, *Principles of Screening*, 83 SCREENING 1323 (1999). Many other authors (too numerous to list exhaustively here) agree with these requirements. *See, e.g.*, USPSTF, *supra* note 18, at xlii-xlvii; JOCELYN CHAMBERLAIN & SUE MOSS, EVALUATION OF CANCER SCREENING (1995); DAVID M. EDDY, A MANUAL FOR ASSESSING HEALTH PRACTICES & DESIGNING PRACTICE POLICIES: THE EXPLICIT APPROACH (1992); HAROLD PETER FERRER, SCREENING FOR HEALTH: THEORY AND PRACTICE (1968); LEE A. LILLARD, PREVENTIVE MEDICAL CARE: STANDARDS, USAGE AND EFFICACY (1986); ALAN S. MORRISON, SCREENING IN CHRONIC DISEASE (2d ed. 1992); NUFFIELD PROVINCIAL HOSPITALS TRUST, SCREENING IN MEDICAL CARE: REVIEWING THE EVIDENCE (1968); Loring G. Dales et al., *Evaluating Periodic Multiphasic Health Checkups: A Controlled Trial*, 32 J. CHRONIC DISEASES 385 (1979); Louis J. Denis, *To Screen or Not to Screen?*, 4 PROSTATE SUPPLEMENT 63 (1992); David M. Eddy, *Clinical Decision Making: From Theory to Practice. Principles for Making Difficult Decisions in Difficult Times*, 271 JAMA 1792 (1994); M. Feinleib & M. Zelen, *Some Pitfalls in the Evaluation of Screening Programs*, 19 ARCHIVES ENVTL. HEALTH 412 (1969); Goldenhersh, *supra* note 36, at 642; Leon Gordis, *Evaluating the Evidence for the Effectiveness of Prevention*, 5 J. GEN. INTERNAL MED. S14 (1990); Halperin & Frazier, *supra* note 30, at 419; Richard S. Hosek et al., *Bias in Case-Control Studies of Screening Effectiveness*, 143 AM. J. EPIDEMIOLOGY 193 (1996); Roman Jaeschke et al., *Users' Guides to the Medical Literature. III. How to Use an Article About a Diagnostic Test. A. Are the Results of the Study Valid?*, 271 JAMA 389 (1994); Roman Jaeschke et al., *Users' Guides to the Medical Literature. III. How to Use an Article About a Diagnostic Test. B. What are the Results and Will They Help Me in Caring for My Patients?*, 271 JAMA 703 (1994); Jesse R. Lee, *Medical Monitoring Damages: Issues Concerning the Administration of Medical Monitoring Programs*, 20 AM. J. L. & MED. 251 (1994); John M. Lee, *Screening and Informed Consent*, 328 NEW ENG. J. MED. 438 (1993); J. Michael McGinnis & Steven H. Woolf, *Background and Objectives of the U.S. Preventive Services Task Force*, 5 J.

1. Principle 1: Natural History of Disease

Regardless of a monitoring program's aim—whether it is meant to encourage primary or secondary prevention[80]—early diagnosis presupposes an orderly "natural history" of disease. The natural history of a disease (that is, its chronological "stages") includes: (i) biologic onset; (ii) early diagnosis possible; (iii) usual clinical diagnosis; and, finally, (iv) outcome.

Unless there is a *"critical point"* in the target disease's natural history—defined as the point where early diagnosis is possible, if at all—before which therapy is either more effective or easier to apply than it would be afterward,[81] monitoring cannot alter the natural history of the disease or condition. These considerations pertain not only to general populations, but also to individual patients or limited groups.[82]

GENERAL INTERNAL MED. S11 (1990); T. McKeown, *Unvalidated Procedures Have No Place in Screening Programs, in* CONTROVERSY IN INTERNAL MEDICINE II 93 (Franz J. Ingelfinger ed., 1974); Medical Practice Committee, American College of Physicians, *Periodic Health Examination: A Guide for Designing Individualized Preventive Health Care in the Asymptomatic Patient*, 95 ANNALS INTERNAL MED. 729 (1981); L. Whitby, *Screening for Disease—Definitions and Criteria*, 2 LANCET 819 (1974); Wilson & Junger, *supra* note 19; Steven H. Woolf et al., *Assessing the Clinical Effectiveness of Preventative Maneuvers: Analytic Principles and Systematic Methods in Reviewing Evidence and Developing Clinical Practice Recommendations*, 43 J. CLINICAL EPIDEMIOLOGY 891 (1990).

79. For example, these same three principles were specifically adopted by the Agency for Toxic Substances and Disease Registry [hereinafter ATSDR] for exposure-prompted disease:

> • The early detection through screening should be known to have an impact on the natural history of that disease process.
> • There should be an accepted screening test that meets the requirements for validity, reliability, estimates of yield, sensitivity, specificity, and acceptable cost.
> • The population being screened should be at a significantly high risk for the undiagnosed disease (i.e. the disease should have a sufficiently high prevalence in the population).

Final Criteria for Determining the Appropriateness of a Medical Monitoring Program Under CERCLA, 60 Fed. Reg. No. 145, 38840, 38842 (July 28, 1995).

80. Monitoring/screening aims to detect two different conditions. Monitoring can facilitate *primary prevention* by detecting a *risk factor* (e.g., elevated cholesterol) underlying a target condition for which it is believed a person is at risk (e.g., coronary artery disease). But a second form of monitoring enables *secondary prevention* by an early detection procedure (e.g., colonoscopy) of *pre-clinical disease* (e.g., colon cancer) for which treatment intervention will provide a favorable outcome.

81. *See generally* SACKETT ET AL., *supra* note 32.

82. For example, Eddy notes that even in exposure-prompted, at-risk cases, physicians should adhere to principles of medical monitoring:

> The available data indicate that changing the timing of detection before the appearance of overt signs or symptoms or before the detection of a sign as an incidental finding during evaluation of another problem does not change the natural history of the disease. The nature of lung cancer appears to be such that the probability or timing of death from lung cancer is not influenced by the timing of detection and treatment within this time window . . . . If the consensus is that "large-scale" screening is not appropriate, a more difficult question concerns what individual physicians should do when they have patients at risk (so-called "case

Consider one example of the importance of establishing a critical point in a target disease's natural history. At first glance, it would seem sensible to try to detect early recurrence of malignancy in very high-risk individuals previously treated for leukemia. However, a series of studies has shown that bone marrow biopsy examinations in asymptomatic, previously afflicted individuals did not improve survival or response to therapy compared to individuals who presented with recurrence because of symptoms.[83] Stated in terms of natural history, individuals in this study whose recurrence was discovered had already passed their critical point, such that no change in prognosis could occur, despite medicine's best efforts to effect the earliest detection possible.

Another example of a disobliging natural history is a "high risk" medical condition for which exposure-prompted medical monitoring intuitively might seem strongly appropriate: cigarette smoking. Smokers are a population at relatively high risk for lung cancer because of years of high concentration exposure to a complex mixture of hundreds of carcinogens. However, based on data from controlled clinical trials of screening, use of annual chest x-rays to screen asymptomatic smokers *is not recommended* by the Preventive Services Task Force expert panel.[84]

Indeed, because cancer screenings are programs offered to a large group of apparently healthy individuals rather than a sensitive tool for risk stratification in a disease for which earlier detection provides for easier, more effective treatment, the American Cancer Society, American College of Radiology, National Cancer Institute, US Preventative Services Task Force, and Canadian Task Force on the Health Examination do not recommend general cancer screening.[85]

Given such examples, one might logically resolve that, absent a "critical point" in the disease/condition's natural history, monitoring claims should be disregarded in legal proceedings. However, a few courts have reached the opposite conclusion—that monitoring, even if unable to change the natural history of a disease, may be awarded.[86] Despite the overwhelming weight of medical literature

---

finding"). Although it might seem harmless and inexpensive enough for a physician to recommend that a particular smoker be screened, if all physicians systematically recommended screening, however, the result would be large-scale screening, which is not recommended.
David M. Eddy, *Screening for Lung Cancer*, 111 ANNALS INTERNAL MED. 232, 236-37 (1989).

83. Catherine Haworth et al., *Routine Bone Marrow Examination in the Management of Acute Lymphoblastic Leukaemia of Childhood*, 34 J. CLINICAL PATHOLOGY 483, 483 (1981).

84. USPSTF, *supra* note 18, at 135; Gina Kolata, *Questions Grow Over Usefulness of Some Routine Cancer Tests*, N.Y. TIMES, Dec. 30, 2001, at A1; Kolata, *supra* note 47, at A1 ("[I]t is unclear whether more people will be harmed than helped by [screening for smoking] because the [screening's] effectiveness has never been rigorously tested. One early study has found that 9 of every 10 patients who have suspicious test findings will not have cancer—but they will have to have additional expensive tests, and sometimes surgery, to find that out.").

85. Daniel R. Wolpaw, *Early Detection in Lung Cancer*, 80 MED. CLINICS N. AM. 63, 77 (1996). *Cf.* Peter J. Littrup, *Future Benefit and Cost-Effectiveness of Prostate Carcinoma Screening*, 80 CANCER 1864, 1869 (1997) ("The 'window of curability' is much greater for patients with prostate carcinoma than for those with lung carcinoma and the rate of metastases for an apparently localized tumor is much less with prostate carcinoma; thus making the potential for early curative therapy much greater.").

86. Chief among these motivations appears to be "peace of mind" for the monitored population, even if no treatment can be offered to those who test positive. *See* Bower v.

to the contrary, we are reluctant to dismiss these courts' logic outright. Courts are charged with a greater policy role than are physicians, whose motivation is only to improve the chances of altering the disease's natural history. A court's extra-medical considerations may well motivate monitoring where medicine would not ordain it. It should be understood, however, that whenever a court permits monitoring not expected to change the target disease's natural history, the court has based its decision exclusively and dispositively on non-medical grounds. Such an approach is emphatically not endorsed by medical experts,[87] and is rejected by a wide majority of courts.[88]

Westinghouse Elec. Corp., 522 S.E.2d 424, 434 (W. Va. 1999) ("Chief Justice Calogero gave a poignant justification for permitting [monitoring] even in instances where there is no proven treatment: One thing that . . . a plaintiff might gain even in the absence of available treatment is certainty as to his fate, whatever it might be. If a plaintiff has been placed at an increased risk for a latent disease through exposure to a hazardous substance, absent medical monitoring, he must live each day with the uncertainty of whether the disease is present in his body. If, however, he is able to take advantage of medical monitoring and the monitoring detects no evidence of disease, then, at least for the time being, the plaintiff can receive the comfort of peace of mind. Moreover, even if medical monitoring did detect evidence of an irreversible and untreatable disease, the plaintiff might still achieve some peace of mind through this knowledge by getting his financial affairs in order, making lifestyle changes, and, even perhaps, making peace with estranged loved ones or with his religion. Certainly, those options should be available to the innocent plaintiff who finds himself at an increased risk for a serious latent disease through no fault of his own.") (quoting Bourgeois v. A.P. Green Indus. ("Bourgeois I"), 716 So. 2d 355, 363 (La. 1988) (Calogero, C.J., concurring)). A second proffered "justification" is that medicine in the near future may make technical and scientific progress, and hence medically validate that monitoring which had theretofore been pointless from a medical standpoint. *Id.*; Redland Soccer Club v. Dep't of the Army of the United States, 696 A.2d 137, 146 n.8 (Pa. 1997) ("[W]e do not require a plaintiff to show that a treatment currently exists for the disease that is the subject of medical monitoring . . . . To do so would unfairly prevent a plaintiff from taking advantage of advances in medical science.").

  87. *See supra* note 78.
  88. *In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 852 (3d Cir. 1990) ("Paoli I") ("Monitoring and testing procedures [must] exist which make the early detection and treatment of the disease possible and beneficial"); *Bourgeois I*, 716 So. 2d at 361 ("[A] plaintiff must show that there is some medical benefit to be gained through early detection of the disease. In other words, plaintiff must show that an existing treatment, administered before the illness becomes apparent to a layperson, is effective in curing or ameliorating the consequences of the illness. Unless such treatment is available, then there is nothing for plaintiff to gain from a hastened diagnosis . . . .") (citing Hansen v. Mountain Fuel Supply Co., 858 P.2d 970, 979-80 (Utah 1993)). *See also In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 787 (3d Cir. 1994) ("Paoli II") (affirming *Paoli I*); Thomas v. FAG Bearings Corp., 846 F. Supp. 1400, 1410 n.8 (W.D. Mo. 1994) (questioning "whether medical monitoring would be of any benefit since there are only 'preliminary ideas' on the treatment of TCE-related impairments") (citing *Paoli I*, 916 F.2d at 852); Villari v. Terminix Int'l, Inc., 663 F. Supp. 727, 735 n.6 (E.D. Pa. 1987) ("At trial, the [plaintiffs] will also have to 'demonstrate the probability . . . that the treatment will be performed . . . .'") (quoting Greenberg v. McCabe, 453 F. Supp. 765, 773 (E.D. Pa. 1978)), *aff'd mem.*, 594 F.2d 854 (3d Cir. 1979)); O'Neal v. Dep't of the Army of the United States, 852 F. Supp. 327, 336 (M.D. Pa. 1994) (affirming *Paoli I*); Bourgeois v. A.P. Green Indus., Inc., 783 So. 2d 1251, 1255 n.2 (La. 2001) ("Bourgeois II") (requiring there be "some demonstrated clinical value in the early detection and diagnosis of the disease"), *overruled on statutory grounds*, 1999 La. Acts 989; Potter v. Firestone Tire & Rubber Co., 863 P.2d 795, 823 (Cal. 1993) (en banc) (demanding

2. Principle 2: Test Accuracy

Two characteristics comprise a monitoring test's "accuracy": *sensitivity* and *specificity*. *Sensitivity* is defined as the proportion of persons with the condition who are correctly identified by the screening test (positive in disease), and *specificity* is the proportion of persons without the condition who correctly test negative.[89]

A perfect test would reflect 100% sensitivity in detecting disease and 100% specificity in confirming the absence of disease. However, no test, even those that afford an illusion of uncanny accuracy,[90] is perfect in separating disease from non-

---

demonstration of "clinical value of early detection and diagnosis") (citing Miranda v. Shell Oil Co., 26 Cal. Rptr. 2d 655 (Cal. Ct. App. 1993)); Petito v. A.H. Robins Co., 750 So. 2d 103, 106-07 (Fla. Dist. Ct. App. 1999) (requiring that "the prescribed monitoring regime is reasonably necessary *according to contemporary scientific principles*") (emphasis added).

89. USPSTF, *supra* note 18, at xliii.

MEASUREMENTS OF THE VALIDITY OF CLINICAL TEST RESULTS

TEST PARAMETER	DEFINITION
Sensitivity	The sensitivity of the test result is defined as the proportion of persons with a condition who correctly test positive when screened, i.e., it is the ability of the test to detect the condition. A test with good sensitivity will detect the disease in nearly all cases; a test with poor sensitivity will falsely indicate the absence of disease in a significant percentage of cases when it is in fact present (i.e., have a high false-negative rate).
Specificity	The specificity of a test is defined as the proportion of persons without the condition who correctly test negative when screened, i.e., it is the ability of the test to exclude the condition. A test with good specificity will have few positive results in persons without the disease; a test with poor specificity will falsely suggest the presence of disease in a significant percentage of healthy individuals (i.e., have a high false-positive rate).
Reproducibility	The reproducibility of the test is the ability of a test to obtain the same result when repeated over and over again.
Accuracy	The accuracy of the test is defined as the proportion of all test results, both positive and negative, that are correct. It encompasses both the sensitivity and specificity of the test.
Efficacy	The efficacy of a test is its measured ability to detect the target condition earlier than without screening and with sufficient accuracy to avoid producing large numbers of false positives and false negatives.
Effectiveness of Early Detection	The property of identifying early on persons with the disease, thus producing a better clinical outcome than those whose disease is not detected early.

90. One must always question whether evolving technology truly affords greater accuracy in diagnosing the condition, or risks triggering an unintended cascade effect, as discussed previously. *See* Deyo, *supra* note 51, at 23 ("Health professionals and laypersons alike tend to equate new medical technology with better-quality health care, assuming that newer is better. Unfortunately [these technologies'] use in routine care sometimes proves

disease.[91] The nature of a test and statistical principles[92] dictate the sources of error and the ability of diagnostic tests to detect disease.

Test sensitivity affects the potential benefits of screening, namely reduction in morbidity and mortality. A test with high sensitivity will generate few false-negative test results (that is, miss few people who actually have the disease). A test with inadequate sensitivity will compromise a program's effectiveness because a large proportion of tested persons who actually have the condition will escape detection and will therefore receive a false sense of reassurance.

Test specificity, conversely, controls true-negative results that can provide reassurance (which for many people is the main reason to be screened). A test with inadequate specificity will generate many false-positive test results, inordinately alarming healthy individuals. False-positive results cause anxiety, "label" a patient, and require additional follow-up testing that may be inconvenient, unpleasant, and may potentially cause iatrogenic complications.

As the above paragraphs have suggested, the decision to monitor turns importantly on whether the test's accuracy—constituted by its sensitivity and specificity—is sufficiently high. The Agency for Toxic Substances and Disease Registry (ATSDR) has stated the importance of available tests' accuracy in stark, cautionary terms:

> There may be serious consequences in the use of screening tests with poor sensitivity and/or specificity. Persons with false negative results may have delays in diagnosis and treatment. False positive results can

---

futile or even harmful."); Deyo, *supra* note 58, at 115 ("With improvements in the resolution of imaging studies, such as magnetic resonance imaging (MRI) of the lumbar spine, more and smaller abnormalities are detected, and the apparent prevalence of disease increases. Yet the anatomical problems identified may not be clinically relevant . . . . Studies suggest that imaging studies only weakly predict either the need for surgery or its outcome. Social and psychological factors are often key components of the patient's responses to back pain."). *Cf.* STANLEY MILGRAM, OBEDIENCE TO AUTHORITY: AN EXPERIMENTAL VIEW 113-115 (1974) (conducting experiment demonstrating that study subjects were willing to engage in objectionable behavior, such as administering potentially dangerous shocks to another person, as long as there is some veneer of scientific legitimacy to the command to engage in such behavior).

91. Harold C. Sox, Jr., *Probability Theory in the Use of Diagnostic Tests: An Introduction to Critical Study of the Literature*, 104 ANNALS INTERNAL MED. 60, 60 (1986). *Cf.* SCOTT SAGAN, THE LIMITS OF SAFETY 28 (1991) ("The belief that intelligent design and management will result in complex organizations that are capable of safely operating hazardous monitoring is an illusion according to [normal accidents theory].").

92. Statistical principles particularly apply to biochemical tests, such as those performed on blood specimens. Standard convention defines "normal" blood samples as being within two standard deviations of the tested population. Using this definition, 95% of the population is "normal," 2.5% of individuals have blood markers that are too high, and 2.5% is too low. Using a twelve-stage ("multi-phasic") chemistry test, the chance of all twelve tests being normal is only 54% (each independent test stage's error rate is multiplied by all eleven others to calculate the aggregate error rate). Therefore, when performing all twelve tests on a healthy individual, the converse applies: 46% of the time, a false alarm will be given, based solely on statistical grounds. Ronald D. Cebul & J. Robert Beck, *Biochemical Profiles: Applications in Ambulatory Screening and Preadmission Testing of Adults*, 106 ANNALS INTERNAL MED. 403, 407 (1987).

result in follow-up testing that is uncomfortable, expensive and potentially harmful.[93]

There are two other key points regarding the sensitivity and specificity of a given test for a given disease. The first is that sensitivity and specificity ordinarily remain unchanged[94] regardless of how many people that comprise the group to be monitored have the disease, or whether they are taken from the general population or from a special risk group, such as persons exposed to chemicals from environmental or "occupational" sources.[95]

The second observation is that more independent rounds of testing a given population for given diseases (called "re-sampling") will not improve the accuracy of testing results. This conclusion is apparent from the following table, which tabulates the number of false positives that are predicted when a healthy population is subjected to unnecessary medical monitoring:

FALSE POSITIVE RATE VERSUS POPULATION SIZE AND NUMBER OF TESTS APPLIED
Specificity = 95%

Size Of Study Population (All Subjects Are Without Disease)	1 Test*	5 Tests*	10 Tests*	15 Tests*
100	5	23	40	54
300	15	68	120	161
500	25	113	201	268
700	35	158	281	376
1000	50	226	401	537
2500	125	565	1003	1342

* Number of people within the total population testing falsely positive. For this table, each test is assumed to be statistically independent of the others.

---

93. Final Criteria for Determining the Appropriateness of a Medical Monitoring Program Under CERCLA, 60 Fed. Reg. 38840, 38842 (July 28, 1995).

94. "[T]he sensitivity and specificity do not [ordinarily] change, as one deals with different groups of people. Sensitivity and specificity, however, CAN CHANGE if the population tested is dramatically different from the population [typically screened and by which these values were fixed]." Medical College of South Carolina, *Sensitivity and Specificity*, at http://www.musc.edu/dc/icrebm/sensitivity.html. In addition, our discussion is predicated on the assumption that a chosen test for a given disease reflects optimal sensitivity and specificity. Yet, often a test's sensitivity can be improved at the expense of specificity, or vice versa. To fix the optimal specificity and sensitivity, statisticians traditionally employ Receiver Operating Characteristic (ROC) curves. *See generally* http://www.anaesthetist.com/mnm/stats/roc/. For more rigorous discussions of ROC curve analysis, see James A. Hanley & Barbara J. McNeil, *A Method of Comparing the Areas Under Receiver Operating Characteristic Curves Derived from the Same Cases*, 148 RADIOLOGY 839 (1983); James A. Hanley & Barbara J. McNeil, *The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve*, 143 RADIOLOGY 29 (1982); Charles E. Metz, *Basic Principles of ROC Analysis*, 8 SEMINARS NUCLEAR MED. 283 (1978); Mark H. Zweig & Gregory Campbell, *Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine*, 39 CLINICAL CHEMISTRY 561 (1993).

95. The ATSDR specifically recognized that medical monitoring of residential populations potentially exposed to chemicals from environmental sources carries risks just as any medical monitoring program does. Final Criteria for Determining the Appropriateness of a Medical Monitoring Program Under CERCLA, 60 Fed. Reg. 38840 (July 28, 1995).

The table demonstrates that if 2500 individuals were tested with all fifteen tests sequentially, 1342 distinct patients would be flagged with "abnormal" results at least once. Even if all individuals who test positive after the first fifteen tests were isolated and retested in an effort to eliminate false positives, at least 67 healthy people would be told they still have an abnormal result upon retesting (5% x 1342 = 67) if the error rate is random[96] and not a characteristic of the individual. To complicate matters, "abnormal" test results are often attributable to ordinary biological variations in a tested population. The usefulness of sequential testing for certain conditions that are prey to such natural and harmless variations is thus further vitiated, because the confidence with which a physician can pronounce an "abnormal" result to be truly abnormal is decreased.[97]

As noted above, an erroneous positive result may cause unnecessary fear and concern in the individual, "label" him as potentially ill, and usually require additional (and more invasive) diagnostic testing, potentially climaxing in a dreaded cascade effect. In short, when a largely healthy population of individuals is screened by a series of tests, a large number of false positives will be generated, just as is the case with application of just one test, even if the sequential testing used is highly specific (that is, have a low false positive rate).[98] Administering a series of tests as part of a monitoring program or redoing tests, exacerbates the number of erroneous diagnoses and missed cases;[99] as repeated testing of the population occurs over time (for example, annually, biennially, etc.), the number of individuals in the population erroneously tested as positive for a disease will likewise increase in cumulative fashion.

Some tests are used to screen for more than one disease or condition. Those that do have multiple sensitivities and specificities associated with them—one pairing for each disease. Hence, one speaks of sensitivity and specificity as existing (i) for a given test (ii) for a given disease or condition. This point is particularly important in legal settings. Occasionally, jurists may be tempted to leave the "details" of monitoring for experts to later determine, and to "decide whether to monitor" as a primary matter.[100] However, this logic is fallacious. It is

---

96. If the error rate in some of the tests were not always random (as is the case for many enzyme tests in which a normal value is defined as the range which includes 95% of the normal population), then anywhere between 67 and 1342 individuals would be told their retesting had provided an abnormal result again.

97. James A. Eastman et al., *Variation of Serum Prostate-Specific Antigen Levels: An Evaluation of Year-to-Year Fluctuations*, 289 JAMA 2695, 2696 (2003) (observing that "natural variations in PSA level may confound our ability to use PSA testing as a successful screening tool" for malignant prostate cancer).

98. The simplified exercise in the Table presumes an entirely healthy population. Re-sampling would be doubly complicated, where individuals with disease also present in the population. This is because one would have to contend not only with false positives, but false negatives too. Thus, at every stage of testing—including re-sampling—some individuals with disease will inevitably test negative.

99. One author has stated that the "shotgun-testing phenomenon" has led to the cynical saying that "the only normal person is someone who hasn't had enough tests. Unfortunately, both patients and naïve physicians may fail to appreciate the statistical basis for the normal ranges of chemical tests, underestimate the likelihood of false-positive results, and fail to recognize that many abnormalities will not represent disease." Richard A. Deyo, *supra* note 51, at 30. *See also* M. Rang, *The Ulysses Syndrome*, 106 CAN. MED. ASS'N J. 122 (1972).

100. *Cf. In re* St. Jude Medical, Inc., 2003 WL 1589527, at *11 n.14 (D. Minn. Mar. 27, 2003) ("Keeping in mind that at this stage plaintiffs only need properly allege and support

impossible to decide the propriety of monitoring without first knowing what diseases are being screened for and which particular tests will be used to screen (and thus how accurate and successful the screening will be). Consequently, test accuracy must be known *before monitoring begins.*[101] Sensitivities or specificities are properly calculated by "gold standard" studies or similar randomized clinical trials. Usually the results of any such studies, if they exist, are available in the medical literature.[102] We suggest that the supervising jurist for a lawsuit or policymaking committee request that the pertinent expert witnesses or scientific advisors cite medical studies yielding relevant accuracy statistics that are reflective of contemporary medical consensus in the opinions of those experts.[103]

3. Principle 3: The Critical Importance of Prior Knowledge of Disease Prevalence

The overall balance between the benefits of detecting a pre-symptomatic condition and the adverse consequences of false results depends largely on the frequency of the condition in the population tested, referred to as the *"prevalence."* The physician confronted with the results of a screening test must try to answer two questions:

> (1) Given a positive test, what is the probability that the targeted disease is present?[104]
> (2) Given a negative test, what is the probability the targeted disease is absent?[105]

---

their [medical monitoring] claim, not prove it, the Court has avoided conducting a lengthy substantive analysis of plaintiff's experts, ensuring only that 'the basis of the expert opinions are not so flawed that they would be inadmissible as a matter of law.'").

101. Ober states this principle well: "We must consider the potential impact of a test or intervention *before* it is performed . . . . Failure to consider this will lead to more adverse effects of testing, including increased anxiety and the potential generation of false-positive results, which will then trigger a cascade of further intervention." Ober, *supra* note 51, at 1012.

102. If the accuracy of a given test is unavailable, it follows that it is impossible to know how safe the monitoring protocol will be. In this case of complete uncertainty, it seems clear that a plaintiff would not have carried his burden of proof in demonstrating the utility of the test for medical monitoring. Thus, his claim would be rejected if no alternate tests exist that are verifiably useful.

103. If the values submitted by various experts or advisors are significantly disparate, a jurist is advised to consult an independent third party expert to reach a reasonable estimation of a given test's accuracy in screening for a given disease. Absent this solution, the jurist may simply have to strike a "compromise" number between the opposing experts' submitted values. *Accord* Lockheed Martin Corp. v. Superior Court, 63 P.3d 913, 921-22 (Cal. 2003) ("[R]eliable medical expert testimony may establish the reasonableness and necessity of medical monitoring. Expert medical opinion, however, does not always constitute substantial evidence . . . . No reason appears why in the medical monitoring context the court should depart from [the] settled understanding that 'an expert's opinion which rests upon guess, surmise or conjecture, rather than relevant, probative facts, cannot constitute substantial evidence.'").

104. The statistic used to answer this question is the Positive Predictive Value (PPV). *See infra* notes 114-24 and accompanying text.

105. The statistic used to answer this question is the Negative Predictive Value (PPN). *See infra* notes 114-24 and accompanying text.

These questions cannot be answered simply by knowing the test characteristics (sensitivity and specificity) for a given disease. However, by an application of Bayes's Theorem,[106] which combines the probability of disease (that is, the chance that a given person in the monitored population has, unrecognized, the target disease) and the sensitivity and specificity for a given test used to detect or exclude that specific disease, the above questions can be answered.

A two-by-two matrix represents the issue well:

		DISEASE PRESENT?	
		Yes	*No*
TEST RESULT?	*Positive*	TRUE POSITIVE	FALSE POSITIVE
	Negative	FALSE NEGATIVE	TRUE NEGATIVE

The four categories of testing outcome (True Positive, False Positive, False Negative, True Negative) are calculated from knowledge of two analytically separable components: Prevalence of disease and Accuracy of the test (Sensitivity and Specificity). Prevalence will differ for each disease and for each population to be monitored. Sensitivity and Specificity are inherent characteristics of each test to be employed.

For any medical test, even those with outstanding characteristics (sensitivity = 95%; specificity = 95%),[107] the fraction of positive test results that are correctly positive does not remain the same across various populations. The true positive fraction rises dramatically as the number of people in the screened population who have the disease (that is, the Prevalence of disease) rises.

Thus, to monitor, (i) one must know the prevalence of unrecognized disease prior to testing, and (ii) that prevalence must be sufficiently high to merit testing. These are not intuitive concepts. Take the first issue: Ex ante knowledge of disease prevalence is an indispensable requirement for making the decision whether to monitor. Many object that this puts the cart before the horse; what if no prevalence is as yet calculable?[108] Can medical monitoring not serve both purposes—to elicit the necessary prevalence statistics and to begin diagnosis of potentially ill subjects? This thinking contends that medical monitoring is needed to determine how many people will contract the disease as a result of a chemical exposure in the first place. And with the massive testing arsenal available to physicians, it is hard to resist intuition's tug to fulfill simultaneously this aim as well as the aim of early detection.

This urge must be resisted. Inverting the traditional adage that lighting a candle is preferable to darkness, William A. Silverman has said if we have possibly ventured into a fireworks factory, it is better "*to curse the darkness than to light the*

---

106. Bayes's Theorem is named for Reverend Thomas Bayes, an 18th century minister and mathematician, who developed a method of calculating conditional probabilities.

107. By comparison, most medical screening tests have much lower accuracies. For example, "[the] [S]ensitivity of chest x-ray [screening for lung cancer] ranges from 40% to 50%" and specificity is less than 50%. USPSTF, *supra* note 18, at 136.

108. See *infra* Part III.B.4 for a description of why in some instances a "best estimate" of exposure-inclusive disease prevalence may not be readily computable.

wrong candle."[109] If the aim is to ascertain the incidence of disease, then epidemiologic techniques (for example, distributing questionnaires regarding diagnosed disease) is all that is required, while avoiding many of monitoring's potential harms.[110] With epidemiology, one is simply attempting to ascertain the number of already diagnosed illnesses in a population over time. It is unnecessary and may be inappropriate to make these diagnoses earlier than they would be made in the ordinary course of medical care.[111] In contrast, medical monitoring is only appropriate where the goal is to diagnose disease before it is clinically apparent.[112] Medical monitoring is not an appropriate means to achieve hypothesis confirmation.[113]

---

109. WILLIAM A. SILVERMAN, RETROLENTAL FIBROPLASIA: A MODERN PARABLE 89 (1980) ("The disturbing consequences of impatient action which I have reviewed recall an apocryphal saying in factories which manufacture fireworks: *It is better to curse the darkness than to light the wrong candle.*") (emphasis in original).

110. *See* SACKETT ET AL., *supra* note 32, at 153-70; Anthony B. Miller, *An Epidemiological Perspective on Cancer Screening*, 28 CLINICAL BIOCHEMISTRY 41, 41 (1995) ("Although it is relatively easy to demonstrate that screening results in earlier detection of cancer, survival is a biassed [sic] measure of its effectiveness. The only valid design to study the efficacy of screening is the randomized trial.").

111. *See supra* Part II.

112. Gochfeld is especially contemptuous of misusing medical monitoring to elicit prevalence statistics:

> [L]awyers seek damages to pay for the periodic examination in perpetuity of clients who have been exposed to hazardous wastes. Considering that the utility of such a program depends upon the magnitude of exposure and the underlying prevalence of the diseased condition being sought, it is apparent that one must have a good understanding of the community risk before invoking a medical [monitoring] program. In fact, it would be reasonable to initiate a[n] [epidemiological] screening for determining the magnitude of exposure and the frequency of effect and to use this information as a basis to determine whether an ongoing medical [monitoring] program is warranted.

Gochfeld, *supra* note 23, at 76.

113. The court said in a New York case:

> [P]laintiffs' experts take the view that medical monitoring in essence would be a research project. [One] testified that the proposed medical monitoring "is a means by which . . . a test or experiment or study might be conducted" in order to test his "hypothesis that there are additional injuries" caused by [the drug] Rezulin. [Another], in a similar vein, [admitted] that the scientific jury is still out on the "question of whether Rezulin is, in fact, an important factor in the development and/ or progression of liver disease" and that medical monitoring, in the form of "careful study of these patients, with case control" was appropriate to answer it. *[But] the Court is not convinced that medical monitoring, at least on a class-or subclass-wide basis, is medically* indicated . . . . In short, it is so far from clear that informed physicians, unaffected by litigation considerations, would recommend routine monitoring on the basis of former Rezulin use that the Court cannot conclude that a medical monitoring action . . . would be an appropriate remedy even if plaintiffs prevailed.

In re Rezulin Prods. Liab. Litig., 210 F.R.D. 61, 73-74 (S.D.N.Y. 2002) (emphasis added);

The second concept becomes clearer by considering the following limiting situation: assume a population of 10,000 includes no one with the disease being sought.[114] Assume also that the relevant test's specificity and sensitivity are both 95%. Then everyone who tests positive (500 people) must be a false positive test result because the entire population is healthy (specificity of 95% means 5% x 10,000 healthy persons = 500 false positive test results). To determine how helpful their testing was in categorizing those with and without disease, statisticians calculate a ratio termed the "positive predictive value" ("PPV"). The PPV is the answer to the first question posed above; it is the probability that any given individual who receives a positive test result actually has the disease.[115] In the present example, the PPV would be zero (true positives [0] ÷ all positives [500]). Thus, the test of this given population for a given disease failed to detect any disease—*not because the test was inaccurate, but because the prevalence of disease was nil.*

The motivation of this extreme example can be generalized. Suppose there is just *one* person with disease among the 10,000. The PPV would still be nearly zero (true positives [1] ÷ all positives [501]). Again, the value of such testing is questionable insomuch as so many positives yield so few true positives. But as the *prevalence* of disease increases (assuming the same size population of 10,000 and the same tests), an increasing fraction of the positive tests will be true positives and the PPV will increase, indicating that monitoring is more helpful at identifying truly diseased individuals in such situations.

As a final quantitative effort to make more transparent the importance of sufficiently high prevalence, suppose (purely for illustrative purposes) that a single

---

O. Gefeller & J. Windeler, *Risk Factors for Cervical Cancer: Comments on Attributable Risk Calculations and the Evaluation of Screening in Case-Control Studies,* 20 INT'L J. EPIDEMIOLOGY 1140, 1141 (1991) ("[I]n the evaluation of screening [the efficacy of monitoring], the results can only be judged in the context of prospective randomized trials."). Gina Kolata has also argued that

> [some] could not conceive of waiting well over a decade for a randomized clinical trial to enroll patients, screen them and follow them as the study wound to a close . . . .
>
>   . . . .
>
> Dr. Barnett Kramer of the National Institutes of Health, knows the emotional pull of patients' stories. . . . [But] [h]e does not [use medical monitoring in place of clinical studies] and must rely instead on cold logic and reasoning. . . .
>
> Dr. Kramer remembers the last time that doctors thought they had a way of preventing lung cancer deaths. During the 1960's and 1970's, many recommended that smokers and former smokers have regular X-rays, and some also ordered tests of their sputum to look for cancer.
>
> Finally, researchers managed to do large clinical trials that asked whether such screenings saved lives. The results were eye-opening. Although more cancers were found in the screened group, and at earlier stages, the screenings did nothing to cut the death rate from lung cancer.

Kolata, *supra* note 47, at A1; *see also supra* notes 28, 101.

114. Obviously, in reality this fact would not generally be known in advance; the very purpose of monitoring is to ascertain who has the disease among an otherwise undistinguishable population.

115. Note that the PPV—the probability of having a disease, given a positive test—is *not* the same as calculating the sensitivity of a test, the probability of having a positive test, given disease.

test having sensitivity equal to 90% and specificity equal to 95% is used to screen for a disease in a group of 10,000 individuals. For this population, however, assume now that the disease incidence is 20 in 10,000 (0.2% prevalence). Of the 20 people who have "Disease Present," 18 will test positive (*Sensitivity of 90% x 20 = 18*). By subtraction, 2 will test negative (*20 total with disease - 18 true positive test results*). This leaves 9980 people with "Disease Not Present" (*10,000 - 20 = 9980*). The number of these healthy people who correctly test negative is 9481 (*Specificity of 95% x 9980 = 9481*). Finally, 499 healthy people wrongly will test positive (*9980 – 9481 = 499*). The PPV (*true positives ÷ all positives*) is then calculated as 3.48% (*true positives [18] ÷ all positives [517]*). These calculations can be represented readily in the familiar two-by-two matrix:

	DISEASE PRESENT?			
	Yes	*No*		
TEST POSITIVE	18	499	18/517 (3.48%)	POSITIVE PREDICTIVE VALUE
TEST NEGATIVE	2	9481	9481/9483 (99.98%)	NEGATIVE PREDICTIVE VALUE

Sensitivity =90%; Specificity = 95%; prevalence = 0.2% (20/10,000)

Thus, when the incidence of disease is 20/10,000 (in the context of environmental chemical exposures acceptable risks are typically at least ten-fold lower), for every 10,000 individuals screened, 19 of the 20 individuals with the disease will be detected on average. However, to achieve this result a very real and probably unacceptable price had to be paid in the process. Approximately 500 healthy people will also test positive and will be falsely alarmed.[116] The low PPV reflects this high false-positive test fraction. A test with a PPV this low is of little use to a clinician monitoring a population for the appearance of disease.[117] The physician is obligated to follow up *all* those with positive test results. Thus, the group that tests positive—in this example, a population comprised overwhelmingly of healthy people—will be subjected to unnecessary additional tests and procedures with all the attendant risks, including potential cascade effects.

The dramatic effect of the pre-screening prevalence on the PPV, one that is unintuitively much greater than the effects of improved sensitivity or specificity on the PPV,[118] can be visualized easily by repeating this calculation for various values

---

116. One person with the disease will also be sent home, falsely assured that he has no illness. The false confidence that person places in his test result may delay his return for diagnosis when noticeable symptoms finally arise.

117. Stephen G. Pauker, *Deciding About Screening*, 118 ANNALS INTERNAL MED. 901, 902 (1993).

118. "For any given sensitivity and specificity, the PPV increases and decreases in accordance with the prevalence of the target condition in the screened population. If the target condition is sufficiently rare in the screened population, even tests with excellent sensitivity and specificity can have low PPV in these settings . . . ." USPSTF, *supra* note 18, at xliv.

of disease prevalence.[119] The resulting diagram indicates that the PPV remains quite low and does not achieve even 50% until disease prevalence is about 1%:[120]

Diagnostic Yield of Hypothetical Medical Monitoring[121]
(Hypothetical Test with 100% Sensitivity & 99% Specificity)

Percent Prevalence of Disease in the Study Population
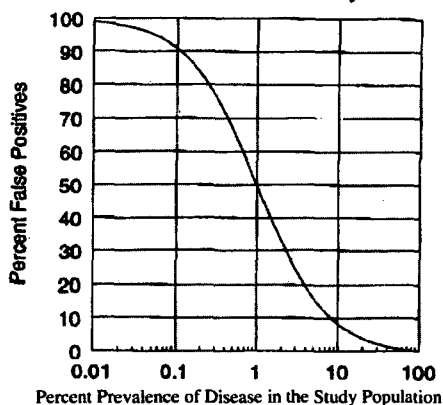
There is no published strict demarcation between acceptable and unacceptable values for the PPV; any value may require *decision analysis* for its justification.[122] Still, in our view, it appears from a study of the contemporary medical literature that it will seldom be acceptable to undertake a medical monitoring program if PPV values are less than between 1% and 5%, regardless of the seriousness of the disease, the opportunity for intervention upon early detection, the cascade effects that may follow false positive tests, and many other case-specific factors.[123]

---

119. We assume 10,000 tested with the same test having the same sensitivity and specificity of 95% each. Test sensitivity and specificity are held constant to isolate and demonstrate the appreciable effect that prevalence has on the PPV.

120. *PhRMA/FDA/AASLD: Drug-Induced Hepatotoxicity White Paper Postmarketing Considerations* 26 (2000), *available at* http://www.fda.gov/ cder/livertox/postmarket.pdf.

121. *Id.*

122. *See infra* Part IV.

123. *See, e.g.*, Chiara Benedetto et al., *A Two-Stage Screening Test for Pregnancy-Induced Hypertension and Preeclampsia*, OBSTETRICS & GYNECOLOGY 1005, 1005, 1008 (1998) (noting that use of Doppler ultrasound examinations alone to monitor for pregnancy-induced illnesses is associated with a 27% PPV, a value "too low to support using it as a routine screening test," but pairing the Doppler test with systolic and diastolic midline estimating statistics of blood pressure rhythm favorably raises the PPV to 63%); Loren G. Miller et al., *How Well Do Clinicians Estimate Patients' Adherence to Combination Antiretroviral Therapy?*, 17 J. GEN. INTERNAL MED. 1, 9-10 (2002) (estimating PPV of monitoring methods for determining whether HIV patients are adhering to medication regiments to be 76-83%, but nonetheless concluding "estimates of adequate adherence may be inaccurate. Better methods for identifying nonadherent patients are needed."); Satwant K. Samra et al., *Evaluation of a Cerebral Oximeter as a Monitor of Cerebral Ischemia During Carotid Endarterectomy*, 93 ANESTHESIOLOGY 964, 969-70 (2000) (concluding that a 33.3% PPV when monitoring for stroke risk during carotid endarterectomy is unfavorably "low"). *Compare* Linda L. Humphrey et al., *A Breast Cancer Screening: Summary of the Evidence for the U.S. Preventative Services Task Force*, 137 ANNALS INTERNAL MED. 344, 351-52, 355-56 (2002) (noting ongoing medical debate whether mammographies of women aged 40-49 with a PPV estimated between 1% and 4% are recommendable as a general breast cancer

Whatever the general PPV "cutoff," it should be evident from the foregoing discussion that it is impossible to decide whether medical monitoring for a given population is recommendable unless one knows the disease prevalence and the PPV.[124] Indeed, even when the first two principles of monitoring—(i) an amenable natural history of the disease and (ii) a sufficient testing accuracy—have been fulfilled, a determination to monitor is expressly contingent on an adequately large PPV value.[125]

4. Principle 4: The Elements of Prevalence

Fulfilling the requirement that the prevalence of disease within a population is high enough to create an acceptable PPV will inevitably involve some calculation of the "risk" of the disease. While no all-encompassing definition of risk may exist,[126] typical parlance, at least in medicine, conveys two ideas: (i) a characterization of the type of harm that is known to occur (for example, bleeding is a "risk" of surgery) and (ii) a prediction of the frequency of occurrence of the harm (for example, the "risk" of significant bleeding from performance of a liver biopsy is about two to six in every one thousand).[127] It is the latter definition of risk that is pertinent in medical monitoring. The prevalence of disease will be the sum

---

monitoring strategy), *with* Mark Helfand et al., *Screening for Skin Cancer*, 20 AM. J. PREVENTIVE MED. 47, 52-54 (2001) (concluding that melanoma monitoring with current technologies, which have PPVs of <1%, is not warranted in light of accompanying emotional and financial impacts); Deyo, *supra* note 51, at 30-31 (noting that Watson-Schwartz testing for acute intermittent porphyria ("AIP") yields an insufficiently low PPV, 0.2%, to merit monitoring). We note that our proposed 1-5% PPV cutoff is intentionally overinclusive for recommendable monitoring. We have set the general cutoff low because there is, in actuality, no universally applicable PPV cutoff. Cutoff determinations are highly situational, and depend on the risks and benefits of follow-up procedures, costs, and so forth. Nonetheless, we are confident concluding that, absent the exceptional case, our 1-5% PPV range is an appropriate lower bound to meritorious monitoring. We stress, however, that jurists should be careful not to conclude that simply because monitoring in a given case could achieve a PPV of at least 1-5%, monitoring is automatically sanctified. Indeed, for particular diseases or conditions, a much higher PPV may be required before general medical consensus would condone monitoring.

124. *See, e.g.*, K. Kerlikowske et al., *Positive Predictive Value of Screening Mammography by Age and Family History of Breast Cancer*, 270 J. AM. MED. ASS'N 2444 (1993).

125. This mathematical exercise helps to rationalize the difference between performing a test on a patient who presents for clinical evaluation and performing the same test as part of a medical monitoring program for a given population. When diagnostic tests are selected and performed in response to specific patient complaints and symptoms, the probability of the suspected disease being present is reasonably high. When a diagnostic test is conducted in screening a population of asymptomatic individuals, however, the probability that the target disease is present is usually relatively low. Low disease prevalence, coupled with the false-positive rate expected for a typical diagnostic test, has a catastrophic effect on test reliability.

126. *See generally* JOHN C. CHICKEN & TAMAR POSNER, THE PHILOSOPHY OF RISK (1998).

127. Sheila Sherlock et al., *Liver Biopsy Today: The Royal Free Hospital Experience*, 1 J. HEPATOLOGY 75, 76 (1984).

of two risks: the background risk (*pre-exposure incidence*) plus the risk associated with the chemical exposure (*magnitude of increased risk*):[128]

```
┌─────────────────────────────────────────────────────────────┐
│                   PRE-EXPOSURE INCIDENCE                      │
│                                                               │
│  PREVALENCE =                  +                              │
│                                                               │
│                   MAGNITUDE OF INCREASED RISK                 │
│                   (Toxic Exposure)                            │
└─────────────────────────────────────────────────────────────┘
```

Most legal opinions categorically deny claims for any medical monitoring other than "special" monitoring, that is, monitoring that could be justified on the basis of specific individuals' background risk (pre-exposure incidence) alone.[129] Indeed, for some diseases, the pre-exposure incidence may be high enough that the same monitoring protocol is recommended for certain groups, regardless of the toxic exposure (e.g., for women age sixty, the incidence of life-threatening breast cancer is 19.5 per 1000 per year).[130] In such instances, unless the increase in risk caused by the toxic exposure was sufficiently high as to warrant additional or alternative diagnostic testing or follow-up treatment, courts will not award monitoring.

There are other circumstances, however, when the background risk for other diseases is too low for a physician to justify monitoring as a general practice (e.g., ovarian cancer mortality incidence is 8.5 per 100,000 per year).[131] Assume that a certain community suffers a toxic exposure known to increase the risk of a disease with such a low background risk. To conclude that monitoring is appropriate requires that the risk increase associated with the exposure (magnitude of increased risk) is high enough that it causes the aggregate prevalence of disease for a given population to rise appreciably and, in turn, causes a sufficiently large increase in the associated PPV to justify exposure-prompted medical monitoring.[132] It follows,

---

128. David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L. REV. 849, 855-56 (1984) (referencing epidemiological studies to demonstrate that disease risk is attributable to both specific toxic exposures and background risk).

129. Note that the background risk for any given individual reflects general health statistics (e.g., the rate of lung cancer in the general population) for the population *as well as* any risk factors that the particular individual or group may also have (e.g., the incidence of lung cancer for a group of regular smokers will be higher as a result of that activity).

130. Russell Harris & Linda Leininger, *Clinical Strategies for Breast Cancer Screening: Weighing and Using the Evidence*, 122 ANNALS INTERNAL MED. 539, 540 (1995).

131. Marilyn M. Schapira et al., *supra* note 41, at 838.

132. Recall court-recognized monitoring applies only to a population (or individual) allegedly at a "*significantly* higher risk" for specified disease(s) as a result of chemical exposure. *See, e.g.,* O'Neal v. Dept. of the Army of the United States, 852 F. Supp. 327, 336 (M.D. Pa. 1994). This phraseology implies that not all increased risks, but only those that are "significant," permit medical monitoring. In the previous part, we demonstrated that the significance of an increase in risk is typically well-proxied quantitatively by the PPV, which in turn is dependent on the disease's prevalence.

then, that a toxic exposure to a chemical agent will make medical monitoring reasonably necessary only when the type, amount, and duration of such toxic exposure is sufficient to raise the predicted prevalence of the target condition in the person(s) to be tested to a sufficiently high incidence that "special" medical monitoring, unwarranted before the exposure, now becomes necessary as a result of that exposure.

One important question remains: from whence should the data be derived to calculate the risk increase associated with the exposure for a particular community? That is, how does one go about calculating the magnitude of increased risk? If both the type and the amount of the chemical exposure at issue have been the subject of previous scientific study, then calculating the future risk of the studied disease in the population to be monitored is relatively straightforward. For example, it has been reported that persons exposed to 200 parts-per-million per year ("ppm-yrs") of benzene by inhalation have about a six-fold increase in risk of developing leukemia as compared to the background risk.[133] Assuming a lifetime background risk for adults of about 50/100,000,[134] this risk can be expressed also as about 300 benzene-caused leukemias per 100,000 exposed persons. Such a scientifically derived value might be regarded as a "best estimate" of the expected risk for a population under consideration for medical monitoring for leukemia exposed to this amount of benzene in the air because both the *nature* of the risk—leukemia is an adverse health effect known to occur[135]—as well as the *frequency* of its occurrence are based on observed outcomes in scientific studies.

What if there are no equivalent studies of people exposed to a toxic chemical in some lower amounts, or by a different pathway, or both? How then should one render a value for the increased risk as a result of the exposure? Returning to the immediate benzene example, if a community has been exposed to 20 parts-per-*billion* per year ("ppb-yrs"), this represents a significantly smaller amount than those exposures that have been investigated in formal scientific studies.

To deal with such gaps in data, regulatory agencies use risk-assessment methods rather than toxicologic observations to generate hypothetical risk estimates.[136] These risk estimates are used for such regulatory purposes as clean-up of waste-contaminated sites, establishing priorities for addressing chemicals of concern, establishing guidelines for exposure that protect public health, etc. While the risk assessment process may employ scientific reasoning and inferences,[137] it is

---

133. R.A. Rinsky et al., *Benzene Exposure and Hematopoietic Mortality: A Long-Term Epidemiologic Risk Assessment*, 42 AM. J. OF INDUS. MED. 474, 474 (2002).

134. National Cancer Institute, *Surveillance, Epidemiology, and End Results Database 12 Registries Incidence and Mortality*, at http://canques.seer.cancer.gov/cgi-bin/cq_submit?dir=devcan2000&db=1&rpt=TAB&sel=4%5E9%5E1%5E%5E23%5E%5E&x=Starting%20Age%5E0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19&y=Ending%20Age%5E1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20&z=Sex%5E1,2,3&dec=3 (2002) (last visited Nov. 27, 2003).

135. This fulfills criteria for scientific causation, and satisfies medical monitoring element four. *See supra* text accompanying note 9.

136. *See* Robert C. Barnard, *Scientific Risk Assessment and the Regulation of Human Cancer Risks: Background and New Directions*, 48 AM. INDUS. HYGIENE ASS'N J. 798 (1987); Robert C. Barnard, *Scientific Method and Risk Assessment*, 19 REG. TOXICOLOGY & PHARMACOLOGY 211 (1994).

137. NATIONAL RESEARCH COUNCIL, RISK ASSESSMENT IN THE FEDERAL GOVERNMENT: MANAGING THE PROCESS 28 (1983).

strongly biased toward protecting the public.[138] Hence, risk assessment estimates, unlike scientific studies, are emphatically not a "best estimate" of the risk resulting from an exposure.[139] Rather, risk assessment traditionally provides, as a matter of public policy, an *upper bound* of what the risk might be. The U.S. Environmental Protection Agency ("USEPA") has justified why it applies this conservative approach, but simultaneously has emphasized caution in directly applying exposure-risk values derived from risk-assessment techniques for other purposes:

> A legitimate use of worst-case scenarios is to determine if the exposure or risk is low enough even at this extreme so as to dismiss concern for this scenario. It is not legitimate to use a worst-case scenario to prove that there in fact exists a concern in a real population . . . . It is critical that the results of a worst-case individual scenario are not immediately applied to an entire population, since in almost all cases this will result in a substantial overestimate of a potential problem.[140]

To understand the important distinction between an "upper bound" estimate and a "best" estimate, consider the following: the USEPA states that the lifetime cancer risk from drinking water containing between 100 to 1000 ppb of benzene is

---

138. Robert C. James et al., *Risk Assessment, in* PRINCIPLES OF TOXICOLOGY: ENVIRONMENTAL AND INDUSTRIAL APPLICATIONS 437, 452-53 (Phillip L. Williams et al. eds., 2000) ("Regulatory agencies address uncertainty in risk assessments by using conservative approaches and assumptions. That is, in the face of scientific uncertainty, they will select models and assumptions that tend to overestimate, rather than underestimate, risk so as to be health protective. Since most of the use of risk assessment is by, or for, regulatory agencies, this conservatism is a dominant theme in risk assessments and a continuous source of controversy. Some view the conservatism employed by regulatory agencies as excessive, resulting in gross overestimation of risks and unwarranted regulations that waste billions of dollars. Others question whether regulatory agencies are conservative enough, and suggest that the public (particularly more sensitive individuals such as children) may not be adequately protected by contemporary risk assessment approaches."). Cernon N. Hawk, *The Risk of Risk Assessment*, 9 REG. TOXICOLOGY & PHARMACOLOGY 257 (1989) ("As currently done, the final product of risk assessment generally produces numbers that have the illusion of precision, but, in fact, this process involves many uncertainties. Risk assessment includes judgmental decisions, and each decision is usually made on the conservative side. Therefore, the end result is a product with many conservative judgments and is therefore an estimate of risk that provides maximum protections to the public health."). *Id.*
    139. As stated by the U.S. Environmental Protection Agency:
> In general, risk values, such as those on IRIS, cannot be used to predict the actual incidence of human disease or the type of effects chemical exposures may have on humans. This is due to the numerous UNCERTAINTIES involved in risk assessment, including those associated with extrapolations from animal data to humans and from high experimental doses to lower environmental exposures. The organs affected and the types of adverse effects resulting from chemical exposure may differ between study animals and humans. In addition, many factors besides exposure to a chemical influence the occurrence and extent of human disease.

Integrated Risk Information System (IRIS); Health Risk Assessment; Guidelines, etc., 53 Fed. Reg. 20,162, (June 2, 1988) (emphasis in original).
    140. Proposed Guidelines for Exposure-Related Measurements, 53 Fed. Reg. 48,829, 48,846 (Dec. 2, 1988).

[1 per 10,000 people (1 x 10^{-4})].[141] Yet this statement is not equivalent to an assertion that 1/10,000 is the actual risk derived from historical events observed scientifically. The "upper bound" estimate instead indicates that the increase in lifetime cancer risk from drinking such contaminated water is *no more than* 1 per 10,000. Indeed, the USEPA cautions that its risk assessment estimates do not "necessarily give a realistic prediction of the risk. The true value of the risk is unknown, and may be as low as zero."[142] If, to calculate the prevalence of disease as part of the process of deciding about medical monitoring, a USEPA-style risk assessment is used to estimate the number of exposure-related cancers expected to occur (that is, the magnitude of increased risk) in a population drinking water containing benzene at 100-1000 ppb, it must be understood that the actual risk is not *known* to be 1/10,000; the actual risk from exposure could be nil.

Hypothetical estimates of risk are simply not intended for use as "best" estimates. Consequently, such figures are not very useful to physicians needing to have an accurate estimate of risk when trying to determine how effective medical monitoring of a given population for a given condition will be. The USEPA, which relies on risk assessment for manifold purposes, has itself noted, "an established procedure does not yet exist for making 'most likely' or 'best' estimates of risk within the range of uncertainty defined by the upper and lower estimates. If data and procedures become available, the Agency [USEPA] will also provide 'most likely' or 'best' estimates of risk."[143] Neither the USEPA nor, to the authors' knowledge, any other group has yet found another method to provide a short-cut to a "best" estimate of risk. Controlled scientific and epidemiologic studies remain the only known methods for providing an accurate measure of increases in risk from toxic exposure.

One conclusion that could be drawn from the preceding discussion is that a plaintiff who calculates a PPV using prevalence data derived from risk assessments has simply not met his burden of showing a substantial increase in risk as a result of chemical exposure; as the USEPA has noted, "it is not legitimate to use a worst-case scenario to prove that there in fact exists a concern in a real population."[144] We are reluctant to be as categorically dismissive of calculating an exposure-specific PPV that incorporates risk assessment data as is the USEPA; although the risk is unknown and may indeed be zero, it may also be as high as the risk derived from risk assessment projections. However, our rule of thumb—a minimum of a one to five percent PPV to justify monitoring—is predicated on a review of the medical literature, which uses "best" estimates of risk in assessing the usefulness of particular monitoring routines. A prevalence calculated with risk assessment data, which present a "worst-case" picture of the risk increase from exposure, could allow a monitoring protocol apparently to survive our proposed PPV cutoff, while in actuality "best" estimates of the risk from exposure, if they were undertaken, could reveal a far less dramatic risk increase and show the monitoring scheme to be imprudent.

---

141. U.S. Environmental Protection Agency, *Benzene (CASRN 71-43-2), at* www.epa.gov/iris/subst/0276.htm#topBenzene (last modified Apr. 17, 2003).

142. Guidelines for Carcinogen Risk Assessment, 51 Fed. Reg. 33,991, 33,998 (Sept. 24, 1986).

143. *Id.* at 33,998.

144. Proposed Guidelines for Exposure-Related Measurements, 53 Fed. Reg. at 48,830.

What we can state with certainty is that if risk assessment data are used to calculate the PPV, rather than "best" estimates from scientific studies, and this "upper bound" PPV fails to meet our one to five percent criteria, it means that even the worst-case increase in risk is almost surely "insignificant" from a medical standpoint, and the medical monitoring claim should be denied.

IV. DECISION ANALYSIS FOR MEDICAL MONITORING

Courts seem, appropriately, to have sensed the absence of a bright line for monitoring efficacy; of late, many medical monitoring cases cite the expression, "[n]o particular level of quantification is necessary to satisfy [the] requirement of significantly increased risk."[145] This statement—whether intentionally or fortuitously—is technically correct. But one must take care not to confuse the assertion that "no particular level of quantification is necessary" to medically monitor with the conclusion that there is no need to quantify risk assumed from exposure. Indeed, there may well be communities that suffer large risk increases as a result of exposure, yet monitoring may still not be recommendable whenever diagnostic tests available to screen for the disease in question have low specificity or sensitivity, or both. Thus, one cannot, as the above quote confirms, simply judge the propriety of monitoring on a "particular level" of quantified increase in risk; available test accuracy must be considered in tandem as well. This is achieved when the PPV, a singular statistic, incorporates both the risk level and the accuracy of available tests to provide a rapid estimation of a proposed monitoring protocol's effectiveness.

We have suggested that if the natural history of the disease at issue is not amenable to early intervention, a court finds a proposed monitoring program to have a PPV of less than 1-5%, or if no PPV is calculable, the monitoring should be rejected without further inquiry.[146] In this sense, this Article has linked an objective, quantifiable value with the qualitative expression, a "significant increase" in risk. In exceedingly rare cases, however, it is conceivable that a monitoring program that fails our proposed 1-5% PPV rule of thumb may be nonetheless advisable. This must be because the court perceives the potential costs of the monitoring regime to be exceptionally low (despite, by definition, the necessarily high error rates of monitoring), and the benefits to be exceptionally high. Again, however, these terms are qualitative and demand objective answers: how low are "exceptionally low" costs and how great are benefits that are "exceptionally high"?

Without quantified estimates of health outcomes, no solid basis exists for comparison or justification of a monitoring program.[147] Therefore, this Part introduces decision analysis: the common, quantitative methodology by which a court can—and should—objectively determine whether the monitoring program

---

145. The genesis of this statement came in *Hansen v. Mountain Fuel Supply Co.*, 858 P.2d 970, 979 (Utah 1993).

146. If the PPV is not presently calculable, but it is believed that later a reliable estimate can be made through an appropriate epidemiological or controlled scientific study, then the court or agency might consider dismissing the monitoring claim pending the outcome of that study.

147. This is no less true if the 5% PPV requirement *is met* in a given instance. There, too, the court should subject a proposed monitoring program to the decision analysis protocol presented in this Part.

meets these exceptional criteria, and the likely outcomes of a screening effort.[148] Put differently: *failing or passing the 1-5% PPV rule is not an excuse to permit monitoring in these exceptional circumstances without undergoing decision analysis.* The decision to recommend a monitoring program simply must be based on an explicit comparison of the benefits, harms, and costs of each component of the program.[149] Making such a comparison obviously requires estimates of the *magnitudes* of the benefits, harms, and costs. As one expert has noted,

because screening is rarely free of cost and often carries risk, either from the test itself or from the subsequent workup it can induce, we must decide when and for whom screening is appropriate. A formal decision analysis can help structure the problem, organize data, elucidate tradeoffs, and estimate benefits and costs.[150]

Such models include the sensitivity and specificity of the test or intervention, the probability of disease, the probability of complications, and the relative change in the risk factor and disease under consideration, delineated by decision analysis of the factors graphically represented as decision trees.[151] Decision analysis has been used in screening assessment for a host of diseases, including breast, cervical, and colon cancers, as well as for coronary artery disease and osteoporosis.[152] The

---

148. *See* Peter Doubilet & Barbara J. McNeil, *Clinical Decisionmaking*, 23 MED. CARE 648 (1985); Jerome P. Kassirer et al., *Decision Analysis: A Progress Report*, 106 ANNALS INTERNAL MED. 275 (1987); Barbara J. McNeil & Stephen G. Pauker, *Decision Analysis for Public Health: Principles and Illustrations*, 5 ANN. REV. PUB. HEALTH 135 (1984); Pauker & Kassirer, *supra* note 70, at 250; Harold C. Sox, Jr., *Decision Analysis: A Basic Clinical Skill?* 316 NEW ENG. J. MED. 271 (1987). Decision analysis is an alternative to conducting a gold-standard controlled clinical trial of a testing program. Experts in monitoring/screening use decision analysis across a set of criteria to simulate the outcome of a controlled clinical trial were it to be conducted. Doubilet & McNeil, *supra* at 648 ("Decision analysis is most applicable to clinical questions that cannot be answered by appealing directly to the results of clinical trial or to a large database. They can occur because no trial has been carried out or because the patient in question differs substantially from the populations in existing sources of data.").

149. Pauker & Kassirer, *supra* note 70, at 250 ("Excellent clinical judgment requires optimal decision making."). Compare a statement made by the Utah Supreme Court:

> [I]f a reasonable physician would not prescribe it for a particular plaintiff because the benefits of the monitoring would be outweighed by the costs, which may include, among other things, the burdensome frequency of the monitoring procedure, its excessive price, or its risk of harm to the patient, then recovery would not be allowed.

Hansen, 858 P.2d at 980, with a ruling by the West Virginia Supreme Court:

> Diagnostic testing must be 'reasonably necessary' in the sense that it must be something that a qualified physician would prescribe based upon the demonstrated exposure to a particular toxic agent. This Court is not entirely in accord with the statement . . . that if a reasonable physician would not prescribe . . . [medical monitoring] for a particular plaintiff because the benefits of monitoring would be outweighed by the costs, which may include, among other things, the burdensome frequency of the monitoring procedure, its excessive price, or its risk of harm to the patient, then recovery would not be allowed.

Bower v. Westinghouse Elec. Corp., 522 S.E. 2d 424, 433 (W. Va. 1999).

150. Pauker, *supra* note 117, at 901.

151. Pauker & Kassirer, *supra* note 70.

152. David M. Eddy, *Screening for Breast Cancer*, 111 ANNALS INTERNAL MED. 389 (1989); David M. Eddy, *Screening for Colorectal Cancer*, 113 ANNALS INTERNAL MED. 373

U.S. Preventative Services Task Force summarized this consensus on an approach to decisionmaking, stating, "[t]he underlying philosophy of the Task Force fits the times perfectly: health professionals should recommend only those interventions for which there is convincing evidence that the benefits will outweigh the potential harms."[153]

A decision analysis enables the physician to work through the potential benefits and harms of each of a series of decisions and outcomes that may be fraught with significance for the patient. To perform a formal decision analysis, a decision tree is used to represent the logical and temporal sequence of a clinical problem between the starting point and the final outcome. Numerical values are then assigned to each component. When all the necessary data are included, the results are combined, or the tree is "folded back," to calculate the contribution of each component to the decision. This process is demonstrated briefly in the following adapted case study.

A. Case Study: Use of Decision Analysis in Ovarian Cancer Screening

The availability of two new tests (transvaginal sonography ["TVS"] and CA 125 [test for ovarian cancer cell specific surface protein]) has raised the question of whether screening of women for ovarian cancer should be recommended. Per principle one, it is true that ovarian cancer has a natural history that permits early intervention.[154] The results of a recently published decision analysis found that average life expectancy in the screened population would increase only trivially (approximately 3/4 of a day of life for women over the age of 65, and 1/3 of a day for women generally).[155] This led the authors to conclude that: "[m]ass screening for ovarian cancer will not improve average life expectancy in the population by a meaningful amount of time and cannot be recommended as an effective health policy."[156]

The basis for this negative recommendation is evident from the results of the decision tree summarized in the Table below. For every cancer detected early as a result of screening a 100,000-person cohort of forty-year-old asymptomatic women, about three women will be falsely alarmed and will undergo needless surgeries, and approximately one woman will be falsely reassured she is healthy when, in fact, she has malignant ovarian cancer. And for every eighteen detected cancers, one woman will die as a direct result of unnecessary surgery.

Assume, as an illustration, that early detection of ovarian cancer has been established to improve survival, that accurate screening tests are available, and that a community has been exposed to waterborne contaminants that are established to cause a 10% increase in the rates of ovarian cancer above the background rate. This is the sort of situation that might well result in a presumptively convincing claim for "special" medical monitoring. Yet in the spirit of the decision analysis by

---

(1990); Eddy, *supra* note 82, at 232; Alan M. Garber et al., *Screening Asymptomatic Adults for Cardiac Risk Factors: The Serum Cholesterol Level*, 110 ANNALS INTERNAL MED. 622 (1989); L. Jospeh Melton III et al., *Screening for Osteoporosis*, 112 ANNALS INTERNAL MED. 516 (1990).

153. USPSTF, *supra* note 18, at 187-91, 209-18.

154. Schapira et al., *supra* note 41, at 838 ("The most important prognostic factor for ovarian cancer is stage of disease at diagnosis.").

155. *Id.* at 841-42.

156. *Id.* at 842.

Schapira et al.,[157] we find that even this arguably "significant" increase in chemically-induced risk, an increase clearly unacceptable from a public health viewpoint, is too small to change the expected outcome of a medical screening effort and the recommendation not to screen.

RESULTS OF DECISION ANALYSIS FOR SCREENING WOMEN FOR OVARIAN CANCER[158]		
EXPOSURE STATUS	BACKGROUND	ADDED RISK BECAUSE OF CHEMICAL EXPOSURE (+10%)
Women without cancer	99,921	99,918
False Positives	50	50
Cancers Detected	18	20
Cancers Missed	11	12
Exploratory Surgeries	68	70
Needless Surgeries	50	50
Deaths from Surgery	1	1

From this exercise, it can be seen that the magnitude of the risk must be known to make a proper decision about medical monitoring. Equally important, as a substitute for such qualitative terms for increased risk as "significant" or "clinically significant" or "meaningful," it is clear that decision analysis can provide the mathematical means to objectively decide if the risk is high enough so that medical monitoring that was not necessary and appropriate prior to the exposure becomes so as a result of the exposure.

CONCLUSION

Medical practitioners must make careful, quantitative considerations when generating recommendations for "special" medical monitoring of asymptomatic populations who have been exposed to elevated amounts of potentially toxic chemicals. To elucidate these critical characteristics for judges and other jurists who are called upon to authorize—at least preliminarily—a medical monitoring program for detecting specific diseases in such persons, this Article assembles and reviews the fundamental reasons for the need to assess and quantify the projected benefits and costs of any medical monitoring request.

---

157. *Id.*

158. *Assumptions*: Annual incidence is 13.8/100,000; Annual mortality is 8.5/100,000; Risk factors are increasing age, family history, residing in two western countries and nulliparity. Most deaths due to ovarian cancer occur within five years of diagnosis.
Probabilities for the Decision Tree: One-time screen in healthy, forty-year-old women; Prevalence of ovarian cancer in forty-year-old women is 28.6/100,000; Two tests in screen: TVS and CA 125 (test for ovarian cancer cell specific surface protein); Sensitivity for CA 125 is 50%-90% depending on early or late stage disease, respectively. Specificity for CA 125 is 97.6%; Sensitivity for TVS is 90%. Specificity of TVS is 98.1%; For combined tests (TVS and CA 125), sensitivity is 45% and 81% depending on early or late stage disease, respectively. Specificity is 99.95%; Mortality from diagnostic laparotomy is 0.23% (2.3/1,000).

Contrary to popular thinking and to some court and agency decisions, the protocol for deciding the medical propriety of monitoring is not haphazard. Nor is it warranted by the egregiousness of a defendant's conduct.[159] More importantly, monitoring is not costless. Indeed, as this Article demonstrates, reliance on the often incorrect belief that "an ounce of prevention is better than a pound of cure" can compel a potentially lethal cascade of events. Subjecting asymptomatic individuals to medical procedures carries a risk of "punching the tar baby"—of injuring healthy individuals needlessly and causing damaging psychological harm from "labeling" healthy individuals as diseased. Ill-considered monitoring can also deter diseased individuals who are erroneously proclaimed healthy from returning promptly when symptoms do present, and can lead to severe psychological harm. In addition, the economic, manpower, and time costs for such programs are usually substantial.

In light of these considerations, we propose jurists adopt a method—already commonly employed by medical professionals—that provides objective means to determine the advisability of any individualized "special" medical monitoring program. This protocol has two steps: First, as a threshold inquiry, whenever a diagnostic protocol has a PPV for a given population for a given disease of less than between 1-5%, or whenever reliable data are lacking to construct a PPV, the proposed monitoring program should ordinarily be dismissed. Calculation of the PPV requires calculation of the estimated increase in prevalence of disease among the exposed population. This prevalence should, where possible, be calculated using data derived from "best" estimates in the medical literature, not from published risk assessments. In exceptional circumstances when it is evident that benefits from monitoring would be exceptionally high (that is, life-saving), and the potential costs of monitoring are atypically small as a result of unique, non-invasive treatment opportunities, a monitoring regime failing to meet the threshold requirements survives this first step of the inquiry.

Second, any monitoring program that passes the threshold inquiry must be subjected to a decision analysis. The interpretation of this secondary inquiry is less capable of being fitted to a "bright line" than is the threshold analysis. However, decision analysis offers significant utility to determining the efficacy of a monitoring program, in that it quantifies the costs and benefits of monitoring. To interpret the results of a decision tree and to conclude whether a given program is net-beneficial, we suggest that the jurist rely on the recommendations of scientific advisors or the testimony of medical experts.

Medical monitoring is a burgeoning element of toxic tort law that continues to grow. It is our hope that the methodology outlined in this article will offer jurists some confidence that the rigorous constructs physicians place on medical monitoring can be properly reflected in the law.

---

159. Henderson, Jr. & Twerski, *supra* note 13, at 837 ("Another reason for the intuitive appeal of medical monitoring claims is that . . . toxic substances have come to epitomize the evils of ruthless industrial technology in the public eye, and the plaintiffs are quintessentially innocent victims of wrongdoing.").