J Clin Epidemiol Vol. 45, No. 11, pp. 1265–1287, 1992
Printed in Great Britain.

0895-4356/92 $5.00 + 0.00

OXYGEN AS A CAUSE OF BLINDNESS IN PREMATURE INFANTS: "AUTOPSY" OF A DECADE OF ERRORS IN CLINICAL EPIDEMIOLOGIC RESEARCH

ROBERT M. JACOBSON*‡ and ALVAN R. FEINSTEIN†
Clinical Epidemiology Unit, Yale University School of Medicine, New Haven, CT 06510, U.S.A.

(Received in revised form 4 June 1992)

Abstract—Several intellectual "autopsies" have recently reviewed errors in clinical epidemiologic studies of causation, such as the original claim that amyl nitrite "poppers" caused AIDS. The current autopsy was done to determine why it took more than a decade—1942 to 1954—to end an iatrogenic epidemic in which high-dose oxygen therapy led to retrolental fibroplasia (RLF) in premature infants, blinding about 10,000 of them. The autopsy revealed a museum of diverse intellectual pathology.

When first noted, RLF was regarded as neither a new disease nor a postnatal effect. In early investigations, the ophthalmologists did not establish explicit criteria for diagnosis and confused RLF with malformations previously seen in full-term infants. Because the patients were not referred until months after birth, the ophthalmologists assumed that the lesion, which resembled an embryologic structure, must have occurred prenatally. Other events suggesting a prenatal cause for RLF were its strong statistical associations with fetal anomalies, multiple gestations, and maternal infections. Although these events were also associated with prematurity, it was ignored when the RLF cases were compared with controls who were mainly full-term infants.

The postnatal timing of RLF was eventually recognized when investigators did cohort studies in premature infants and found that RLF could develop in eyes that were normal at birth. As the search for a cause turned to events occurring after birth, statistical associations were produced for agents such as light, vitamins, iron, vitamin E deficiency, and hypoadrenalism. Each study had its own methodologic flaws: controls were missing for light; co-maneuvers were ignored for vitamins and iron; objective diagnosis was not used for vitamin E deficiency; and the research on hypoadrenalism contained biases in susceptibility and detection as well as problems of a competing outcome event.

When the role of oxygen administration was first considered, the statistical association with RLF was stronger for vitamin- and iron-therapy than for oxygen. In addition, many investigators were dissuaded by contradictory evidence from institutions in which RLF was either absent despite high-dose oxygen or persistent despite reduced dosage. The contradictory evidence was later regarded as erroneous because of unsatisfactory delivery systems for the oxygen or failure to check the actual oxygen concentrations. An alternative explanatory hypothesis, rejecting the role of high-dose and long-duration oxygen, was the idea that RLF was due to "relative hypoxia", produced by overly rapid weaning from oxygen therapy rather than the duration of oxygen treatment itself. Because the occasional

*Work done during a research fellowship at the Robert Wood Johnson Clinical Scholars Program, Yale University School of Medicine. Current status: Instructor, Department of Pediatrics, Mayo Medical School, Rochester, MN. Senior Associate Consultant, Mayo Clinic and Foundation, Rochester, MN.
†Professor of Medicine and Epidemiology; Director, Clinical Epidemiology Unit and The Robert Wood Johnson Clinical Scholars Program, Yale University School of Medicine, New Haven, CT. Senior Biostatistician, Cooperative Studies Program Coordinating Center, Veterans Administration Medical Center, West Haven, CT.
‡All correspondence should be addressed to: Robert M. Jacobson, M.D., Mayo Clinic, 200 First Street, S.W., Rochester, MN 55905, U.S.A. [*Tel*: 507-284-4466; *Fax*: 507-284-0161].

spontaneous regression of RLF had not yet been recognized, some investigators even believed that RLF could be cured by a high concentration of oxygen.

Finally, despite complaints that deprivation of oxygen therapy would be unethical, its role was tested in a large, multicenter, randomized trial. The unequivocal results of that trial ended the epidemic.

Causation Clinical epidemiology Oxygen Prematurity Retinopathy Retrolental fibroplasia

INTRODUCTION

Epidemiologists do not have an opportunity to do the post-mortem examinations that have been used for more than two centuries to confirm or refute medical decisions in cause–effect reasoning. The decisions are expressed as clinicians' diagnoses for the cause of death in a particular person; and the necropsy (or autopsy) reveals what actually happened. The revelations have led to the fundamental progress that occurs whenever scientists can determine errors and analyze reasons for the errors. What is learned in autopsy analyses has allowed clinicians to improve the scientific quality of diagnostic reasoning for subsequent cause–effect decisions.

In the absence of post-mortem examinations, epidemiologists have had to seek alternative opportunities that may offer analogous progress by disclosing truth and revealing error. The alternative opportunities are not abundant. The "gold standard" research of a randomized controlled trial might confirm or refute the cause–effect decisions derived from non-randomized epidemiologic studies, but randomized trials are seldom ethically possible today for agents suspected of having noxious effects. Even when possible, the trials are seldom feasible if the effects would require protracted observation of a huge number of people. The problem of feasibility usually also precludes the scientific substitute that most closely resembles a randomized trial: a prospective non-randomized cohort study of exposed and unexposed persons.

Accordingly, epidemiologists generally have only two major opportunities to validate truth and appraise error. In one opportunity, an erroneous causal accusation is revealed when some other agent has been unequivocally demonstrated to be the correct cause of a disease. In the second opportunity, the evidence used for a previous accusation becomes thoroughly discredited by subsequent analyses and research. Because both types of opportunities are uncommon, the authors and editors can be vigorously applauded for the recent publication of an "intellectual autopsy" [1], analyzing an error that occurred in epidemiologic cause–effect reasoning when amyl nitrite "poppers" were falsely accused of causing AIDS. The publication of that analysis stimulated us to think about the possible enlightenment that might come from analogous intellectual autopsies for other well-acknowledged errors.

We chose a particularly striking example: the iatrogenic epidemic of retrolental fibroplasia (RLF) that lasted for more than a decade and that led to the blinding of about 10,000 premature babies [2] during the years 1942–1954. During that period of time, the provocative agent for the RLF epidemic was sought with many non-randomized studies and with a few randomized trials that led to a large sequential collection of erroneous accusations and decisions. Eventually, the role of therapeutic oxygen was suitably tested and unequivocally incriminated in a randomized trial that brought the epidemic to an end.

Although now often called *retinopathy of prematurity*, RLF continues to occur, and sometimes even appears in infants who received no oxygen therapy. Because RLF does not occur in all premature infants exposed to oxygen or in full-term infants receiving oxygen therapy, oxygen is neither a *necessary* nor a *sufficient* cause of RLF. Nevertheless, oxygen therapy was definitely a *contributory* cause in the epidemic.

Our goal in this study was to determine why the epidemic lasted so long and why the provocative role of oxygen eluded prompt discovery. Since the causes of the RLF epidemic received many investigations, we suspected that methodologic flaws might have been responsible for wrong conclusions and delayed recognition of oxygen's causal contributions. Just as the revelations found in conventional autopsies can help prevent repetition of pre-mortem diagnostic errors, we hoped that discernment of methodologic errors in the oxygen/RLF epidemic might be helpful for avoiding such errors in the many other causal relationships that are prominent investigative challenges today.

Because a "museum of pathology" emerged from our autopsy of the errors, we have organized this report in three main sections. The first contains a discussion of general sources of error in epidemiologic cause–effect reasoning. The second section describes seven main stages in the chronologic sequence of events investigating the relationship between RLF and oxygen therapy. The third section contains comments about what might be learned from the problems noted in the second section.

SOURCES OF ERROR IN EPIDEMIOLOGIC CAUSE–EFFECT REASONING

Although many scholarly discussions [3–14] have been devoted to principles of reasoning and sources of error in analyzing cause–effect relationships, most of the problems can be methodologically grouped into two simple categories [15]: "wrong editorials" and "wrong news".

The wrong editorials occur during the interpretation of the evidence accepted as news. If the news itself is correct, the editorial reaches a wrong causal conclusion because the writer has overlooked an alternative causal mechanism that can explain the same news equally well and that also happens to be correct. Two prominent old examples of epidemiologic error arose in this manner. One of them took place in 1852 when William Farr [16, 17] obtained correct statistical evidence that cholera occurred more commonly at sea level than in mountain regions. Using the then-fashionable miasma theory for etiology of disease, Farr ascribed the cause of cholera to atmospheric pressure rather than water pollution. Another prominent error happened in 1914 when the Thompson–McFadden Commission [18] obtained correct statistical evidence that pellagra was particularly common in residents of the same house and in neighboring houses. Using the contemporary focus on microbial transmission of disease, the Commission decided that pellagra was caused by infection, not by nutritional deficiency.

In the two instances just cited, the original research was itself impeccable, and the evidence presented as "news" was entirely correct. Repetition of the same studies on other localities would have repeatedly confirmed the original results. The errors arose as wrong editorial interpretations, because the investigators had overlooked a concomitant co-maneuver—something else that was happening beyond the role of the suspected causal agent. The errors were revealed by the discovery and demonstration of the co-maneuvers that were the correct etiologic agents.

A different, more complex, and probably more common problem arises when the news itself is wrong. In this situation, the interpretive editorials begin with erroneous evidence, but the editorial writers and readers may not know that the evidence is wrong. The errors can arise at every phase of the complex pathway that produces evidence when investigators assemble groups of compared people, collect raw data, and organize the results. Sometimes the error may be due to inaccurate data for the observed groups; sometimes the compared groups or data may be biased by unequal methods in the process of comparative collection and observation. In the absence of rigorous scientific rules for the methodologic quality of these activities, the acquired evidence can often have important distortions that are not immediately apparent. If strong and consistent, the wrong evidence may be accepted without suitable scientific attention to the possibility of error.

The erroneous editorial conclusion may then be particularly difficult to refute. Having been incorporated into the basic evidence, the errors may be suspected but cannot be demonstrated or refuted with the evidence itself. Unless an alternative, correct causal mechanism is unequivocally demonstrated with other research, the refutation of the original error will usually require contradictory evidence from additional studies of the same cause–effect relationship. The contradictory results will inevitably lead to controversy; and the disputing opponents may battle ineffectually until truth becomes established not through specific scientific standards for evidence and inference, but through withdrawal of arguments on one side or the other.

This type of problem occurred about 15 years ago when three case–control studies, each at a different international site, were published in the same issue of a prominent journal [19–21]. The studies showed a consistent statistical association that was interpreted as a causal relationship between reserpine and breast cancer. After creating extensive repercussions in the medical world, and after many contradictory studies, the relationship was eventually discredited [22]; and its advocacy was withdrawn by at least two of the initial proponents [23]. Intellectual autopsies of this problem have suggested that the original evidence was erroneous in allthree studies, with different types of error in each study [10, 22, 24]. The list of errors

included: unadjusted chance effects of multiple comparisons, biased exclusion in choice of control groups, unwarranted changes in control groups after analysis of the data, and ascertainment bias when information was assembled about exposure.

During the decade of research before the RLF–oxygen relationship was clarified, errors arose in both the editorials and the news. The causal trail was at first obscured because the investigators erroneously interpreted the basic nature of the disease. They then went looking in a wrong direction where any of the acquired evidence would be irrelevant. When the search was turned to the right direction, the investigators looked at the wrong targets and produced erroneous evidence indicting those targets. When these fallacies were corrected and the search finally went toward the right direction and right target, the investigators were initially misled either by more wrong evidence or by erroneous interpretation of correct evidence.

The next main section describes the recurrent sequence of error-and-correction that would ultimately be ended as error-and-trial.

CHRONOLOGY OF STAGES IN THE RLF RESEARCH

The sequential search for the cause of RLF can be divided into seven stages that began with identifying the disease's basic characteristics and timing and that ended with a randomized clinical trial. These stages refer to: (1) recognition and early concepts of RLF; (2) wrong concept of pathogenesis; (3) wrong evidence supporting wrong agents; (4) correcting the initial errors; (5) wrong searches in the right place; (6) distorted views of the right target; and (7) unequivocal demonstration of oxygen's role.

1. Recognition and early concepts of RLF

In 1942, Theodore L. Terry, a professor of ophthalmology at the Massachusetts Eye and Ear Infirmary in Boston, reported [25] an unusual cluster of five blind premature infants who each had an opaque membrane behind the lens of the eye. Although the first infant was initially thought to have cataracts [26], the diagnosis was later altered to persistent tunica vasculosa lentis after blood vessels were found in the membrane.

Terry initially regarded [25] the lesion as fibroplastic overgrowth of a vascular sheath behind the crystalline lens. Stating that the lesion had no precursor in normal embryologic structures, he believed that it was different from the rare, sporadic persistant tunica vasculosa lentis, which appeared usually in one eye of full-term infants. Terry suggested that if the lesion occurred regularly in premature infants, a new etiologic agent was at work.

In a later report [26], however, Terry changed his ideas about both pathology and etiology. After reviewing and comparing his seven cases with various other cases that had been brought to his attention, he decided that RLF was part of the spectrum of *persistent tunica vasculosa lentis*, although the latter disease had not been hitherto described in premature infants. Terry's stature as an ophthalmologist led to prompt acceptance of this new interpretation. The lesion that he called *retrolental fibroplasia* [27–29] would be initially perceived and investigated as neither a new disease nor a phenomenon of premature children but as a variant of an old disease in full-term infants.

Despite the change in interpretation, Terry demonstrated a dramatically increasing incidence for the lesion. The five cases he first reported [25] in 1942 later rose [26] to 20, and then [28] to 97 by 1944. In 1945, he noted [30] that large numbers of cases were occurring in Chicago, Boston, and Hartford. By 1946, Terry had personally and repeatedly examined 100 cases [31] and had learned of a total of 200 more [32]. In a presentation to ophthalmologists [28], Terry stated that he knew little of the care of premature infants. He assumed that the surprisingly large number of cases that he had seen in the previous few years could not be explained by any recent changes in care during the antecedent decade.

What Terry did not know was that therapeutic doses of oxygen had recently become used routinely in the premature newborn nursery. In previous treatment of cyanosis, the early delivery systems for oxygen were expensive and did not maintain a consistent concentration [2]. With improvement in technologic design, special new incubators could deliver relatively high, consistent concentrations of oxygen at relatively low cost [33–36]. In premature infants, oxygen administration had relieved cyanosis [37] and made respirations become regular [38]. In an era when blood oxygen content was not routinely measurable [39], premature infants were also thought to be frequently hypoxic without cyanosis. Consequently, as early as 1945, oxygen administration had become recommended [40, 41] for

all premature infants; and its prophylactic usage had been credited [42] in part for a decrease in neonatal mortality. While the practice of prophylactic oxygen was becoming widespread, newer incubator designs provided higher concentrations of oxygen more efficiently [43]. Terry did not know about this change in policy in the early 1940s, and that more and more pediatricians were giving oxygen to *all* premature infants in their nurseries.

2. Wrong concept of pathogenesis

Despite the belief that RLF was not a new disease, Terry [29, 31] maintained that it was most likely caused by a postnatal agent, because the eye lesion had never been noted until the premature infants were 4 or more months old. Terry said [31] that proponents of a prenatal cause "appear to give no serious consideration to the fact that both Clifford [Stewart H. Clifford, a pediatric colleague] and I have seen this ocular abnormality develop in apparently normal eyes 8 weeks or more after birth." The search for a postnatal etiologic factor was soon diverted, however, because the medical community became convinced that RLF was a congenital disease whose cause would be found in events that preceded birth.

In 1946, Algernon Reese and Frank Payne [44], professors at the Institute of Ophthalmology at Presbyterian Hospital in New York City, concluded from their own case series that the eye lesion resembled, and indeed was, a persisting embryologic structure. Arguing that the basic lesion of RLF was a "persistence in part or *in toto* of the primary vitreous" found in the embryo, Reese contended that the disease must arise prenatally. Two other prominent ophthalmologists [45, 46] also concluded that the disease occurred before birth.

These alternative interpretations relied on the same type of evidence available to Terry but were also supported by two other types of data. First, because specific diagnostic criteria had still not been established for RLF, the series of current and past cases assembled by different investigators included not only premature infants, but also some eye lesions noted at birth in full-term infants. To be present at birth, the ocular lesion had to arise before birth.

A second source of support for a prenatal cause came from observations [44, 45, 47] that an apparently high frequency of infants with RLF had congenital hemangiomas. Since the eye lesion histologically resembled a hemangioma, the ocular and the cutaneous lesions were regarded as having the same congenital pathogenesis.

3. Wrong evidence supporting wrong agents

With the belief that RLF was congenital, the search for its cause was aimed at prenatal events, and epidemiologists began to join the investigation. One of them, Theodore H. Ingalls of the Harvard School of Public Health, had just scored an apparent etiologic triumph in demonstrating [48] that fetal anoxia was an important mechanism in causing Down's syndrome (then called *mongolism*). Using the same approach he had applied to Down's Syndrome, Ingalls [47] combined three additional cases with those of Reese, Terry, and Arlington C. Krause. In this combination of cases, Ingalls ascertained the frequencies of various previous conditions and events.

Ingalls did not assemble a specific control group. Instead, he compared antecedent events for the RLF cases either with what he had previously found in his Down's syndrome series, or with events previously reported for general groups of infants, who today would be called "historical controls". He found that the RLF case group had an increase in placental problems, twinning and other multiple births, certain prenatal maternal infections, and a few manifestations attributed to fetal anoxia in singleton premature infants.

From this comparison, Ingalls concluded [47] that RLF was the result of anoxia during fetal development. In a later paper entitled, "The study of congenital anomalies by the epidemiologic method: with a consideration of retrolental fibroplasia as an acquired anomaly of the fetus" [49], he enthusiastically recommended that his methods become a model for future investigations of congenital anomalies.

Among several forms of evidence for Ingalls's hypothesis that fetal hypoxia led to RLF, one of the indicted factors was placental abnormalities and hemorrhage. When Ingalls compared the timing of hemorrhages during pregnancy, he found that the mothers of infants with RLF bled late in pregnancy, whereas the mothers of infants with Down's syndrome bled earlier. He decided that the higher frequency of late bleeding in the RLF group was a clinically significant link to its cause [50]. Ingalls also found that mothers of his combined case-series of infants with RLF had a higher frequency of placental disease, gestational hemorrhage, and

prolonged rupture of amniotic membranes than what would be expected from previously published occurrence values for these events in mothers of all infants [47].

Two other items of evidence were also offered for the fetal-hypoxia hypothesis. First, having found more twinning in infants with RLF than in the "historical-control" groups, Ingalls [50] decided that the twinning caused uterine-crowding and thereby fetal hypoxia. A second item of evidence [50] for the fetal-hypoxia hypothesis was that the mothers of RLF infants had an apparently increased rate of antepartum infections (although Ingalls offered no ordinarily expected comparative rate).

A different hypothesis for the congenital cause of RLF came from laboratory research rather than epidemiology. When compared with the offspring of normally fed rats, the offspring of rats given a vitamin A deficient diet had malformations in many soft and skeletal tissues. One of these malformations, called "open eye", was manifested by a fibrous retrolental membrane [51]. Although no supporting evidence was offered in humans, Terry [52, 53] concluded that human RLF might be due to maternal deficiency of vitamin A.

4. Correcting the initial errors

About 6–8 years had now elapsed since Terry first identified and christened RLF in Boston. The existence of the disease had been confirmed by Reese in New York, by Krause in Chicago, and by other ophthalmologists elsewhere. A rigorous uniform set of operational criteria, however, had not yet been proposed or established for the diagnosis.

The absence of such criteria was not unusual. Even today, many morphologic and clinical diagnoses are still made without the reproducibility that could be provided by the algorithmic rules and specific instructions of operational diagnostic criteria. Terry's first report, in 1942, occurred 2 years *before* his Boston colleague, T. Duckett Jones, became a pioneer in modern "clinimetrics" [54, 55] by establishing specific operational diagnostic criteria [56] in 1944 for a disease—in this instance, rheumatic fever—investigated at many different institutions.

The failure to develop standard diagnostic criteria was thus a customary feature of research practice in that era. The custom endured for many years thereafter and still sometimes persists today. In the absence of standardized, concordantly applied criteria, however, different ocular lesions might be designated as RLF by ophthalmologists in the different research regions where it was being studied. Frustrated by the observer variability and subsequent uncertainties about the contents of a collected case series, Krause said [57], "I have been wondering whether there are not three types of fibroplasia—the Boston, the Knickerbocker, and the Midwestern". The absence of standardized criteria helped produce the misleading wrong diagnoses of RLF at birth in full-term infants and inevitably led to the wrong interpretation that RLF had a congenital cause.

This wrong conclusion was abetted by wrong evidence produced by two types of methodologic errors. One of these errors involved improper longitudinal reasoning from cross-sectional studies. In all the case series studied by the ophthalmologists, the evidence was cross-sectional. Each ocular lesion had been noted at a single postnatal point in time, and no longitudinal cohort studies had been done to investigate the successive events as the RLF developed in each patient after birth. When each ocular lesion was discovered, the subsequent pathogenetic reasoning went backward from that date.

The erroneous cross-sectional conclusions about an embryologic or other congenital origin were ended, in 1949–1951, by a series of birth-cohort studies of premature infants [58–62]. The investigators found that premature chidren were born with normal eyes and that RLF, when it occurred, developed over the next few months. Although these observations became well accepted and helped demolish the congenital hypothesis, Reese and his colleagues remained unconvinced. Although Reese had previously admitted that he never immediately checked the eyes of newborn infants [63] and that his examinations began after the children were 2 months old, he continued to insist [64] as late as 1952 "that the disease can be prenatal" with "approximately 20 percent ... seen at birth ... if it can be assumed that skin hemangiomas have a common cause with retrolental fibroplasia".

The association with hemangiomas arose from a second type of methodologic error: neglect of bias in compared results for case and control groups. Initially the investigators [44, 45, 47] did not assemble specific control groups, being clinically impressed by what appeared to be a relatively high number of hemangiomas in the RLF cases. Later, when

the occurrence of hemangiomas in infants with RLF was directly contrasted [63] with a "historical-control" group consisting of previously reported results in both premature and full-term infants, hemangiomas were more frequent in the RLF infants. Since hemangiomas tend to be larger and more striking in the few months after birth, rather than just at birth, they were more likely to be apparent when the RLF was also first noted. Furthermore, the hemangiomas were sought with "active surveillance" in the RLF infants, whereas ordinary newborns may have received relatively "passive" searches. Consequently, the association of hemangioma with RLF may have arisen from an error now called "detection bias" [15]. The association was refuted when later checked in a case–control study of infants with RLF, where premature infants, carefully and contemporaneously examined [65], formed a more suitable control group.

We have not tried to determine how and where Ingalls went wrong in his conclusion that Down's syndrome was caused by fetal anoxia. (The many etiologic errors in studies of that syndrome would be an interesting topic for another autopsy analysis.) Ingalls's conclusion that fetal anoxia also caused RLF, however, was produced by another classical error. When he chose the control groups, he ignored the question of susceptibility to the outcome event. Convinced that RLF was congenital and examining a case-mixture comprising full-term infants with the questionable diagnosis and premature infants who had RLF, Ingalls used *full-term* infants for his control comparisons. One comparison was with his own series of cases of Down's syndrome. The other control group consisted of the "historical" results reported in the literature for normal full-term infants. Because of this error, all of the "etiologic" factors that Ingalls identified for RLF—placental problems and bleeding, twinning, and antenatal maternal infections—were causally related to prematurity, not to RLF. Ingalls's errors were later revealed when prematurity was suitably considered. In particular, when Harold Speert *et al.* [66] compared cases of RLF against controls who were born prematurely, no elevation was found for placental problems, bleeding, or antecedent maternal infections.

The hypothesis about vitamin A deficiency came from another type of classical error: an improper extrapolation from animal experiments to humans. Unaccompanied by any supporting data in humans, the vitamin A deficiency hypothesis became untenable when other investigators [67], measuring the degree of vitamin A deficiency needed to produce the eye lesion, determined that the required deficiency was so severe that it would prevent human mothers from being able successfully to deliver children. In subsequent human studies, the administration or withholding of vitamin A was found to have no effect on the incidence of RLF [53, 60, 68, 69].

When all these corrections had been accomplished, the research was ready to move away from the erroneous focus and erroneous agents of investigations aimed at the idea that RLF was a prenatal congenital disease. Most researchers began to agree with the statement of William C. and Ella U. Owens [59] that "the most fruitful investigation is to be found in a study of factors which are active in postnatal life." By this time, even the problems of diagnosis itself had also improved. In 1949, 3 years after his original study, still maintaining that RLF had a prenatal etiology, Reese [63] divided his case series into two diseases. He described one lesion, "persistence and hyperplasia of the primary vitreous," which was found at birth typically in one eye of full-term infants. He described a second lesion, "retrolental fibroplasia," which he found typically in both eyes of premature infants manifesting months after birth.

After about 7 or 8 misdirected years, the search for cause could therefore go in the right direction toward a postnatal etiology in premature infants. In an editorial in the **Journal of Pediatrics** in 1950, Lawrence T. Post [70] said:

> "If one believes with Reese that all cases of this disease are congenital in origin, the cause must be sought in the mother. If, however, the Owenses are correct in contending that the disease develops after birth, one would look for something in the treatment of the child. . . . It may be, therefore, that some foods, drugs, or treatment that are given to the premature child may be a factor."

Nevertheless, many investigators did not promptly accept the idea of a postnatal cause. In a conference of prominent RLF researchers [71] in April 1951, the summary contained the remark that the etiologic agent "for RLF is not known, nor is it known whether the disease results from prenatal or postnatal causes." Thus, the impact of the early confusion had not yet been fully removed.

5. Wrong searches in the right places

As other investigators began to seek postnatal rather than prenatal agents, the trail toward oxygen was diverted by several wrong turns that produced erroneous evidence or interpretations indicting four other possible causes. They were precocious exposure to light, water-miscible vitamins and iron, vitamin E deficiency, and hypoadrenalism. The production and eventual correction of these errors are discussed in the next few subsections.

Precocious exposure to light. Before the medical community became convinced that RLF was congenital, Terry [29–31, 72] had proposed precocious postnatal exposure to light as a possible cause. Postnatal light was thought to stimulate myelination of the optic nerve [29], and Terry's animal data showed that the variation of postnatal light exposure affected the timing of the regression of the hyaloid arterial system in animals [30]. Furthermore, Terry [72] thought that the variation in nursery lighting might explain the diverse incidence of RLF from nursery to nursery. Terry actually proposed that mydriatics or miotics be used to prevent RLF, not knowing whether the pupils should be dilated to resemble their natural state in the presumably dark uterus or constricted to prevent excessive exposure to environmental light. Although generally dormant while RLF was considered congenital, the environmental light hypothesis was refuted in 1949 when RLF was shown [73] to develop in four of five premature infants who had been blindfolded almost immediately after birth for about 1–2 months. Another investigation [74] reached the same conclusion after an unusual control arrangement in which one eye was patched in each of 22 premature infants. Early changes of RLF had occurred in both patched and unpatched eyes.

The original accusation that environmental light caused RLF arose from a lack of controls. All premature infants, as well as any infant who would later get RLF, would have been exposed.

Water-miscible vitamins and iron. In 1949, V. Everett Kinsey and Leona Zacharias reported a secular association [75] study of changes in the annual incidence rates of RLF at a series of hospitals. The changing rates were related to the annual levels of 47 variables used in premature care at the same institutions. Kinsey and Zacharias found a close correlation between the increasing annual incidence of RLF and the postnatal use of water-miscible vitamins and iron. All three annual rates—the incidence of RLF, the average actual amount of water-miscible vitamins in drops administered per infant per hospital stay, and the average administration of iron in grams—seemed to rise, peak, and fall together. These findings were "confirmed" in the same report by a separate statistical association between the percent incidences of RLF in the 3–4 lb birthweight group and the differing quantities of iron and water-miscible vitamins used in six hospitals. The results showed higher rates of RLF in hospitals with higher use of vitamins and iron.

The proposed hypothesis was then tested in a prospectively assembled cohort of infants receiving vitamins and iron. When RLF in this cohort developed at the same rate as in untreated historical controls, Kinsey [53] withdrew his previous hypothesis and concluded that RLF was not caused by water-miscible vitamins or iron.

The original error was probably a faulty interpretation arising from failure to consider a concomitant co-maneuver. The same hospitals that liberally used oxygen therapy may also have been using the minerals and vitamins. Another possible explanation was erroneous data due to incomplete detection of RLF. The apparent fall in RLF incidence near the end of the observation period, when iron and vitamin dosage were simultaneously reduced, may have been an artifact. RLF was often not detected until long after the neonatal hospitalization, and the cases in the last few years of data for the secular groups may have been underenumerated. Another possible explanation for the erroneous positive secular association is that it arose by chance as a statistical "fluke" of the "multiple-comparison" process when 47 different variables were explored individually. The error in the "corroborating" statistical association is more difficult to explain. One possibility is that zero usage was assumed when data were missing for iron and vitamin quantities in the hospitals under study. (In one of the eight hospitals, RLF occurred in 4% of patients despite no use of vitamins and low use of iron).

Vitamin E deficiency. In a clinical trial begun in a premature nursery in 1948, infants admitted with a birth weight of 1360 g or less were alternately (rather than randomly) given synthetic vitamin E. The Owenses [59] found no cases of RLF in the 11 premature infants who

received vitamin E and 5 cases in the 15 who did not. Although not "statistically significant" (two-tailed $p = 0.053$, by Fisher test), this encouraging result made the investigators end the trial after 10 months and propose that vitamin E be routinely administered. Without explaining the disparity in numbers for the "alternately" assigned patients, the investigators concluded that RLF was caused by a theoretically plausible deficiency of vitamin E and that RLF could be prevented by suitable treatment. After the trial ended, only one of the next 12 infants routinely treated with vitamin E developed RLF—a baby so ill with apnea and cyanosis that the oral vitamin E could not be given until the eleventh day of life.

Pooling results obtained before or during the trial, the investigators compared attack rates of 1/23 (4.3%) in infants supplemented with vitamin E and 17/95 (17.8%) in those who were not. The authors claimed the result was clinically significant, but it did not reach statistical significance (two-tailed $p = 0.10$, by Fisher test). The vitamin E hypothesis became refuted, however, when the initial "success" failed to repeat itself in a series of studies [53, 76, 77] where the vitamin was given prophylactically to premature infants and remedially to infants with "early" RLF.

The reasons for the original vitamin E error are not clear. In fact, some investigators today still do not acknowledge it as an error; and the possible therapeutic role of vitamin E in retinopathy of prematurity (the modern name for RLF) is still being investigated and disputed [78, 79]. If an error occurred, however, it probably had one or both of two sources. The first was that investigators during that era were not aware of the need for objective ("masked" or "blind") evaluation of outcome events. Since the examiners of each infant's outcome knew what treatment had been received, the identification of RLF may have been biased. A second possibility, suggested by the unequal numbers of treatments that should have been assigned alternatingly (and thus equally), was that clinicians who regarded vitamin E as dangerous might reserve it for the healthier looking infants who were also less likely to be receiving supplemental oxygen. With this judgmental form of allocation, oxygen would be given more frequently to patients denied than to those who received vitamin E. The effects of the concomitant presence or absence of oxygen-therapy would then be mistakenly ascribed to the absence or presence of vitamin E.

Hypoadrenalism. In the era when adrenocorticotrophic hormone (ACTH) had just been isolated [80–82], it was inevitable that hypoadrenalism would be proposed as a cause of RLF [77]. The proposed mechanism was that a premature infant misses the apparent hyperadrenalism of the third trimester of pregnancy and thus develops a relative hypoadrenalism that might lead to RLF.

Investigators [77] began to test this hypothesis in 1950 and found that progress of the eye lesion either arrested or regressed in all of 14 infants who received ACTH for rapidly progressing RLF. No infants developed a partial or complete retrolental membrane. Some fundi actually returned to normal, although others had abnormal masses of tissue remaining. In a "control" group of 8 infants with rapidly progressing RLF, who for unstated reasons did not receive ACTH, 3 developed complete bilateral membranes, 3 partial membranes, and 2 had abnormal masses of tissue in the fundi. These results encouraged the investigators to proceed with further studies.

In a larger case series, which included those 14 infants and 17 additional infants treated with ACTH for acute RLF, the investigators [83] reported similar success except in one child who had no response and in two children who had incomplete responses, resulting in one blind eye each. The infant who failed to respond had actually died during treatment of an exacerbation of RLF. Three other infants also died despite an apparently excellent response to ACTH. All of the deaths occurred when ACTH was being given at the first signs of RLF, whereas formerly no deaths had occurred when ACTH was reserved for instances of progressive RLF.

When the earlier therapeutic policy was followed [83], progressive RLF developed in 16 of 24 infants thought to be at high risk for RLF. All 16 patients received ACTH, and none died, although two had one blind eye each. This result seemed to compare favorably with the occurrence of total blindness in six of seven infants, at a second hospital in the same city, where RLF was not treated with ACTH.

A more extensive comparison of results was then conducted for the two hospitals [83]. At the first hospital where ACTH was used for progressive RLF, only one (0.7%) infant developed blindness out of 135 infants born in 1950

and weighing less than 2000 g. At the second hospital, using no ACTH, 6 (8%) of 72 such infants were blind. The investigators concluded their comparison by calling for a randomized trial of ACTH treatment. Simultaneously, however, other investigators described their uncontrolled experiences with ACTH given in various doses for various stages of disease [84–86]. The results, representing 26 or 27 infants, did not support the use of ACTH.

Infants recognized as having progressive RLF were then entered into a randomized trial [87] at the two hospitals. Of the 30 eyes treated with ACTH, 16 (53%) retained vision, 10 (33%) did not, and 4 (13%) were indeterminate. Of 36 untreated eyes, 25 (69%) retained vision, 7 (19%) did not, and 4 (11%) were indeterminate. After tabulating the total results of both uncontrolled and controlled studies, the investigators concluded that ACTH treatment neither improved nor worsened vision. The mortality data, however, showed a 6/36 (17%) death rate in the ACTH group, compared with 1/49 (2%) in the untreated group. The rates of significant infectious disease morbidity showed corresponding results. The investigators therefore recommended that ACTH be abandoned as a treatment for RLF.

The faulty conclusion in the original cohort study was probably due to methodologic errors in both susceptibility and detection bias. The compared groups of ACTH-treated and untreated babies came from two different hospitals and had different initial severity in their RLF lesions. In addition, the physicians who assessed the outcome event (improvement of RLF) were aware that ACTH had been used when they made the subjective ophthalmoscopic observations of its effects.

A separate problem in the original cohort study was the higher mortality with ACTH, so that some patients died before their RLF had a chance to become apparent or more severe. This methodologic problem still occurs in randomized trials today and has not been solved by the use of life-tables or intention-to-treat analyses. The problem arises when a "competing event" removes patients from follow-up before they have had a chance to develop the main event. In many analyses today, the adverse-competing-event patients are withdrawn as "censored" without the event being attributed to the associated treatment. Consequently, certain adverse effects of treatment may be unrecognized unless all adverse events (analogous to "total mortality" rather than "cause-specific mortality") are attributed to the corresponding treatment. The controversy about which policy to use—count all adverse events or only the main event—is currently unresolved [15].

6. Distorted views of the right target

The time and effort occupied by forming and refuting wrong postnatal hypotheses led to substantial delay in thoroughly evaluating the etiologic role of oxygen. The possible impact of oxygen had actually been explored on one relatively early occasion and on several later occasions during the RLF epidemic. On each occasion, however, the role of oxygen was minimized and dismissed. The dismissals came from three sources: (1) a weak association, showing that RLF was more strongly related to water-miscible vitamins and iron than to oxygen; (2) contradictory results, such as the complete absence of RLF in certain hospitals using high-dose oxygen; and (3) the interpretation that RLF was caused not by hyperoxia, but by relative hypoxia when patients were "weaned" too rapidly during reduction of oxygen therapy.

Weak association of oxygen with RLF. Supplemental oxygen had been included among the 47 variables studied in Kinsey's 1949 exploration [75] of secular associations for RLF. The supplemental oxygen was regarded as only weakly associated, however, because the rising annual rates of RLF had paralleled the increasing rates of oxygen use in the first few years but had fallen in the last 2 years despite a plateau in use of oxygen. In contrast, the use of water-miscible vitamins and iron had directly paralleled both the rise and apparent fall of RLF. The stronger association with the vitamins and iron thus eliminated attention to the weaker association with oxygen.

The weakness of this association was challenged and rejected several years later when similar studies [88–90] of secular associations showed a direct parallel time trend for an increased incidence of RLF and an increased use of oxygen. The source of error in the initial secular association [75] is not clear. One possible explanation (discussed earlier) is a delayed detection of RLF during the last 2 years of the secular data, when the discordant drop occurred in the incidence of RLF. The consequence would have been a loss in the parallel secular trends and the erroneous conclusion that RLF was only weakly associated with

oxygen. The Committee on Fetus and Newborn of the American Academy of Pediatrics [91] suggested another possible explanation: difficulties in maintaining a constant and known concentration of oxygen in the incubators. The incubators had to be frequently opened for infant care and often leaked even when closed. Furthermore, the concentration of oxygen was usually not measured, and its delivery was determined only by the flow-rate of oxygen into the incubator. Until Beckman paramagnetic oxygen analyzers became widely available in 1953, oxygen concentrations could not be measured reliably or frequently (W. A. Silverman, personal communication).

Contradictory dismissal of correct evidence. From outside the U.S., evidence incriminating postnatal oxygen was beginning to accumulate. In 1951 V. Mary Crosse [92] reported that 12 of 14 infants with RLF in Birmingham, England, had been given oxygen continuously for more than 2 weeks via an oxygen tent or incubator rather than with a face mask used in previous years. That same year Fergus W. Campbell [88], in Melbourne, Australia, retrospectively contrasted infants at one institution, who freely received high concentrations [40–60%] of oxygen therapy by oxygen cot, with infants at two other institutions, who had received moderate concentrations of oxygen (often in a restricted manner) via funnel or catheter. When the respective rates of RLF were 23/123 (19%) and 4/58 (7%) in the two groups, Campbell concluded that oxygen probably caused RLF and urged that high-dose oxygen be avoided. Also in 1951, A. C. L. Houlton [93] noted that RLF in Oxford, England, had suddenly increased in close association with the use of oxygen delivered by tent.

In 1952, Crosse and Philip J. Evans [90] examined the temporal association of RLF with periods of oxygen availability in Birmingham, England. In one special premature-baby unit, the results were as follows:

Time	*Availability of oxygen*	*Occurrence of RLF*
1931–45	None	0/105 (0%)
1946–48	Some	1/27 (4%)
1949–50	Freely used in amount and duration	5/26 (19%)
1950–51	Restricted	0/24 (0%)

After noting that special premature-baby units had appeared in Europe, England and Australia, Crosse and Evans suggested that the growing use of these units had fueled the spread of RLF. They argued that the well-intentioned subsidy of the units by National Health insurance in England helped foster the RLF epidemic, and urged [90] that the oxygen-policy be revised.

All of these arguments from outside the U.S., however, were disregarded because of the evidence regarded as particularly compelling that oxygen did *not* cause RLF. This evidence was the absence of any RLF cases at the New Orleans Charity Hospital, where high-dose oxygen had been used abundantly [94]. Whenever oxygen was accused of being the causal agent, the absence of RLF at Charity Hospital would be offered as a cogent counter-argument. The counter-argument was supported by reports from hospitals in Oxford [93] and Paris [95] where the increasing appearance of RLF had failed to subside when the routine use of oxygen was restricted.

The amount of decrease in oxygen at Oxford was never specified [93], but the Paris study [95] offered more detail. In the Paris study, 479 infants were enrolled in a trial and allocated in an unspecified manner into 2 groups. In 135 infants, a restricted regimen of 50% oxygen was given when clinically indicated and only for the first 8 days of life. The other 344 infants received a liberal regimen of 60% oxygen for 3–8 weeks. The occurrence rate of "regressive" RLF was four times higher in the restricted group (22.2% contrasted to 5.8%), although the more severe "membranous" RLF occurred at the same rates in both groups (6–8%). The authors concluded that the hypoxia probably caused RLF. No explanation is available for the peculiar results in the Paris study. Since the pattern of allocation of patients is unknown, one possibility is that the group receiving the relatively high dose of "restricted" oxygen was more premature and at higher risk for RLF than the group receiving the "liberal" therapy. Another possibility is that the RLF was not identified with similar diagnostic procedures and criteria in the two treatment groups. If oxygen therapy confined infants to closed incubators and if access to patients was restricted in order to maintain desired oxygen levels, the fleeting, spontaneous regressing forms of RLF would have been difficult to detect.

The belief that oxygen did not cause RLF persisted until it was displaced by the overwhelming evidence of randomized clinical trials (to be discussed later) showing the causal role of high-dose oxygen. After oxygen had been un-

equivocally inculpated, however, an attractive explanation was offered [96] for the contradictory results at the municipal Charity Hospital: The nursery was large, with an average daily census of 100 infants, and the incubators had to be opened widely to care for the infants. Additionally, the oxygen concentration was not directly measured. The flow at the source, set at 4 l. a minute, was only *assumed* to give 50% oxygen to the infant. Thus the infants probably failed to receive significant oxygen supplementation. The interventional maneuver under study—oxygen supplementation—was not given with suitable proficiency. The failure to consider inadequate performance of the maneuver [15] thus led to the wrong conclusion that oxygen supplementation did not lead to RLF.

Contrary explanations: hypoxia and rapid weaning. By 1952, two types of well-accepted evidence had been assembled to show a strong relationship between oxygen and RLF. One set of evidence came from the secular association studies discussed previously. Entirely new evidence, however, came from a clinical trial led by Arnall Patz [97], showing that RLF occurred more often in premature infants receiving liberal oxygen supplementation than in the control group receiving restricted oxygen.

Funded by a small National Institutes of Health (NIH) grant, Patz, then a young ophthalmology resident, and his colleagues alternatively assigned one of two treatments for infants with birth weights less than 3.5 lb at a Washington DC, municipal hospital. The first group received 65–70% oxygen for 4–7 weeks and was weaned over 1 week. The second group received less than 40% oxygen for 1 day–2 weeks (the reason for the range was not specified) and the dosage was weaned over 1–3 days. The two groups otherwise received similar treatment. The patients' eyes were examined regularly in the hospital and after discharge for 6 months.

The original proposal by Patz *et al.* for a controlled trial of oxygen therapy evoked the vigorous opposition that could be anticipated in an era when the idea of controlled trials was not yet well accepted. Holding strong beliefs about the virtues of oxygen therapy, one of the clinical reviewers of Patz's proposal for NIH support said, "... these guys are going to kill a lot of babies by anoxia to test a wild idea" [98]. Although Patz *et al.* agreed to keep all the babies in the low oxygen group at "a healthy pink color," the attending nurses feared the possible harm of restricting oxygen therapy and would turn up the oxygen surreptitiously [98].

Patz *et al.* [97] enrolled 76 infants but later excluded 11: 8 because their oxygen levels fluctuated and 3 who failed to return for follow-up. The investigators reported that 17 (61%) cases of RLF occurred in the 28 infants receiving liberal oxygen and 6 (16%) in 37 receiving restricted oxygen. Furthermore, 7 (25%) of the liberal-oxygen group had RLF changes beyond Grade II (detached retina, retinal hemorrhages, opacities, retinal membrane), but no such changes were found in the restricted-oxygen group. The investigators did not include any data about mortality in their report, thereby implying that the restricted use of oxygen had no adverse effects. While concluding that the liberal use of oxygen was no longer advisable, Patz *et al.* were puzzled by the unexplained variation in the reported incidences of RLF and called for more controlled studies.

One criticism of the trial [99] was the high drop-out rate. Nevertheless, in a hypothetical examination of the opposing possibility, if all 10 of the liberal-oxygen dropouts had no RLF and the 1 dropout of the restricted-oxygen group had RLF, the incidence results for RLF would still have been persuasive: 17/38 (45%) vs 7/38 (18%) respectively. Both the original and hypothetical results are statistically significant at $p < 0.05$ as well as dramatically different for the 2 groups. Thus, the effect of the dropouts would not have importantly altered the conclusions.

A second criticism [99] arose from the effect produced when the nurses gave oxygen surreptitiously late at night to the restricted-oxygen group. The extra oxygen probably increased the group's morbidity from RLF. Thus the dramatic difference in RLF in the two groups would have been even larger without the unscheduled extra oxygen.

A third criticism referred to the allocation of treatment by alternate rather than randomized assignment. In a trial where allocation is not done "blindly", physicians can monitor the pattern of alternating assignments and arrange the patients in suitable order to receive the desired treatment. This type of allocation bias can be suspected when the compared groups have substantial differences in baseline characteristics such as birth weights. Since these characteristics were not reported by Patz *et al.*, the possibility of allocation bias could not be examined.

A fourth criticism concerned the small sample size in the Patz trial. A total of 770 infants would have been needed to have an 80% power for detecting a 10% decrease in RLF at a level of $p < 0.05$. (Nevertheless, the 65 patients in the trial provided greater than 95% power to detect the difference that was found.)

A fifth criticism was aimed at the "non-blinded" observation and diagnosis of RLF. Patz *et al.* did not indicate whether the observers were aware of the treatment groups, so that detection bias could have occurred either toward or away from the hypothesized association of oxygen with RLF. Although "masked" or "blinded" schemes of observation were first [15] introduced in 1931, the strategy was still relatively new in the early 1950s and did not become commonplace until a decade later.

Although the evidence reported by Patz *et al.* and from Birmingham, Oxford, and Melbourne was correct, some clinicians preferred an alternative interpretation. They were reluctant to believe that harmful effects could come from oxygen, which had an almost mystical reputation for doing good. The alternative interpretation was that the observed RLF effects were attributable not to hyperoxia, but to relative hypoxia occurring when the oxygen concentration was dropped too rapidly [100–104] during an abrupt wean from high concentrations. Thaddeus S. Szewczyk, the leading proponent of the rapid-wean theory, expressed it as follows [103]:

> "On superficial examination of these reports, it would indeed appear that oxygen is toxic and that its use does cause the eventual development of retrolental fibroplasia. As I will show later, however, the true cause of retrolental fibroplasia is hypoxia and this, in the greatest majority of the cases is produced by improper use of oxygen and by its too rapid withdrawal from a child who has become acclimated to it."

Three associations were used to support this claim: (1) in a retrospectively assembled case-series [100, 101, 104] RLF was mainly diagnosed only after the oxygen was discontinued; (2) evidence from case-series [100], secular associations [101–103], and an alternating-assignment controlled trial [104] appeared to support the idea that RLF occurs after fast but not slow weans from oxygen; and (3) prospective cohorts [100–104] seemed to show that oxygen could reverse RLF. The investigators concluded that drastically rapid weans should be avoided by avoiding excessively high concentrations of oxygen and by giving oxygen only to those infants who needed it [101, 104]. The Szewczyk hypothesis [103] even incorporated Ingalls's belief that *prenatal* hypoxia was causing RLF—a theory still taken seriously [71] in the early 1950s. Szewczyk [103] suggested that prenatal hypoxia probably created a "hyper-reactor" state that made infants particularly susceptible to postnatal hypoxia, thus developing RLF more easily.

Certain studies [100–104] indirectly supported the relative hypoxia hypothesis by showing the presumptive benefit of hyperoxia in the *remedial* treatment of RLF. When high-dose oxygen was given to infants with manifestations of early RLF, the ocular lesion stabilized or regressed. These results were contradicted by a finding in Patz's trial [97] that sudden removal of oxygen did not worsen the existing RLF in two infants. In fact, both infants studied by Patz underwent a spontaneous regression of the lesions. Spontaneous regression of RLF was just being recognized [53, 105] at that time and later became the explanation for many of the therapeutic effects seen in the early, uncontrolled studies of various treatments for RLF.

As the rapid-weaning-relative-hypoxia hypothesis became accepted, the recommendation in textbooks [106–108] began to call for less use of oxygen than in the past and for slow weaning of infants who received it. Nevertheless, persistent doubts remained about the role and the causal mechanism of the oxygen therapy [109–110]. These doubts led to a randomized clinical trial called the National Cooperative Study [111], whose results (as discussed in the next section) showed unequivocally that RLF depended on the duration of oxygen, not the rate of wean.

The slow-wean hypothesis had been supported by accurate data that received an erroneous interpretation, which ignored the concurrent variation of two maneuvers. As described by Robert H. Bedrossian *et al.* [104], the slow-wean procedure consisted of 1 day at 50% oxygen, 5 days at 40% oxygen, and 5 days at 30% oxygen. The rapid wean consisted of 11 days at 60% oxygen. In both procedures, room air began on the twelfth day. Thus, as subsequently pointed out [91], the rapidly weaned infants actually received a longer exposure to a high concentration of oxygen, whereas infants weaned slowly received much less oxygen.

The comparative data may also have been biased by the lack of simultaneous controls and objective observation. For example, the investi-

gators [101] initially evaluated the proficiency of the oxygen delivery only if RLF developed in an infant who was supposed to receive the slow wean. If the incubator gave less oxygen than originally intended, the infant was considered to have suffered a rapid wean.

The peculiar results of oxygen-therapy as a medical remedial treatment for RLF could also have resulted from the lack of controls and of objectivity. Objective observation of a control group would have shown [105] that after a certain age RLF does not worsen with more oxygen and that many cases in the RLF spectrum become better or stable.

7. Unequivocal demonstration of oxygen's role

By 1953, despite the publication of Patz's controlled trial incriminating high-dose oxygen, the situation was still in a state of vigorous controversy [111]. Many pediatricians and ophthalmologists continued to doubt the etiologic role of oxygen, which was being used liberally at some institutions and restrictedly at others. The results of the relatively small trial by Patz *et al.* failed to overcome the opposing beliefs produced from New Orleans that RLF was absent despite liberal use of oxygen and from Paris that RLF was abundant despite restricted usage.

The arguments, offered by Szewczyk and others, that RLF occurs only after oxygen is stopped, were countered by other reports [97, 112] that RLF appeared *during* oxygen treatment. In addition, Patz's clinical trial [97] was discounted by clinical observers [91, 113] who had become skeptical after the early enthusiasm and subsequent rejection of previous clinical trials (with ACTH and other agents). These clinicians sought confirmation before accepting the hypothesis that oxygen-therapy could cause RLF. They feared that uncritical acceptance of this theory would lead to unnecessary hypoxic consequences in premature infants.

Furthermore, animal studies [114–122] continued to provide contradictory data showing that eyes were damaged by both hypoxia and hyperoxia. Although each group of investigators found that the damage was homologous to RLF, no animal model ever showed the full stages of retrolental fibroplasia. In retrospect, it seems reasonable that both hypoxia and hyperoxia can change retinal vessel size and growth. The later animal models [121, 122] showed that hyperoxia transiently altered retinal vessel growth in newborn kittens but not in mature cats and not elsewhere in the kittens' vasculature. The lack of permanently scarring lesions, however, militated against accepting the belief that hyperoxia could cause a lesion like RLF [91].

A final problem [98] that argued against prompt acceptance of the Patz trial was its small size. The results could not offer assurance that the restricted-oxygen group had avoided excess mortality—an adverse effect that had been overlooked in the early studies of ACTH.

Because of the persistent doubts, a group of investigators representing 18 hospitals began a randomized clinical trial—called the National Cooperative Study—sponsored by the U.S. Public Health Service National Institute of Neurological Diseases and Blindness. Planned [113] in 1952 and begun on 1 July 1953, the trial compared two regimens that were each denounced [2]. One group of opponents feared that randomized allocation would deprive many infants of the oxygen that had become routinely recommended as life-saving. The other group believed that infants randomly assigned to receive high doses of oxygen would be unnecessarily exposed to the dangerous risk of RLF [99]. The controversy was summarized by Kinsey, the leader of the trial, and F. M. Hemphill [113] as "One could anticipate as provocative a discussion from the statement that oxygen was probably a factor in production of retrolental fibroplasia as from one suggesting that it was unrelated to the disease."

The multicenter trial [111] was aimed at determing whether liberal oxygen therapy produced RLF and whether restricted use of oxygen would increase mortality in premature infants weighing $\leqslant 1500$ g who had survived 48 hr. One group was randomly assigned to receive 50% oxygen for 28 days. The other group was to receive either no oxygen or up to 50% if "the clinical condition of the infant demanded it."

The existing controversy led to various compromises [113] in the design of the trial. Because of the contentions that mortality would be increased by low oxygen and RLF increased by high oxygen, infants were randomized in a 1:2 ratio. During the first 3 months of the study, for every infant assigned to the routine liberal use of oxygen, two were placed in the restricted oxygen group. Thereafter, if mortality was the same in the two groups, *all* infants would receive only restricted oxygen for the rest of the trial. The investigators believed that the sample size would be sufficient to provide statistically convincing distinctions in both RLF and mortality.

Another feature of the compromised design [113] was that infants would be enrolled only after the first 48 hours beyond birth, thereby allowing the use of oxygen during the first 2 days of life. In addition, infants assigned to the restricted-oxygen group could receive supplemental oxygen if needed for clinical reasons—reasons that were not stated before the trial or tabulated afterward. Thus, despite a median of 2 days, the durations of oxygen treatment ranged from 0 to 50 days; and some infants in the restricted-oxygen group received long periods of supplemental oxygen. Finally, in implicit deference to the Szewczyk hypothesis, the group [113] receiving liberal oxygen was weaned over 3 days.

Because the importance of objective observations (via "double masking" or "double blinding" of the observer) was not fully appreciated in that era, no such precautions were taken. The ocular examiners were able to determine the antecedent treatment for each patient.

The preliminary results of the trial were announced on 19 September 1954 and published [113] in 1955. In the first 3 months, 68 infants were enrolled in the liberal oxygen group and, as planned for a 2:1 allocation ratio, 144 received the curtailed treatment. The rate of mortality (deaths before 40 days of life) was similar, 22 and 25%, respectively in the two groups. In the next 3 months, therefore, 179 more patients were enrolled and, as planned, were assigned only to the curtailed group. The final mortality rates were 22% (15/68) in infants receiving liberal oxygen and 20% (65/323) in those receiving curtailed oxygen.

The preliminary results also reported the incidence of RLF in all infants followed at least $2\frac{1}{2}$ months after birth. This requirement affected only the later entrants assigned solely to the curtailed oxygen group. Thus, the RLF occurrence rate was determined from all 53 survivors among patients receiving liberal oxygen and in 245 of the 258 survivors of those receiving curtailed oxygen. Among the surviving infants, RLF developed in 72% (38/53) of the liberal-oxygen group and in 30% (73/245) in the curtailed-oxygen group [113].

Although a later participant in the multicenter cooperative study, Bellevue Hospital in New York City was also the site of a second trial, conducted before the cooperative study and concurrent with the Patz trial [91]. In 1954, two investigators in the National Cooperative Study, Jonathan T. Lanman and Loren P. Guy, together with Joseph Dancis, published results of a randomized trial [123] containing 86 infants enrolled when less than 12 hours old and weighing between 1000 and 1850 grams. Two compared regimens were randomly allocated in a factorial design that produced four groups treated with liberal or restricted oxygen and with or without oral estrogen. The estrogen regimen was investigated because of the hypothesis that RLF might result from the premature infants' early removal from exposure to maternal estrogens. The Lanman group defined the liberal-oxygen therapy as a high-oxygen concentration (a retrospective average of 69%) given for at least two weeks or until the infant weighed 1500 g. The restricted-oxygen therapy provided low concentrations (a retrospective average of 38%) only when the infant was cyanotic. The latter regimen was stopped daily and restarted only if the infant became cyanotic again.

The trial [123] by Lanman *et al.* thus had a design quite different from that of the multicenter study. The multicenter study did not test estrogen, enrolled no infants until 48 hours after birth, and allowed diverse dosages of oxygen in the immediate postnatal period. Lanman *et al.* justified their trial by stating that the previous study by Patz *et al.* was not suitably designed, having assigned treatment by alternation rather than randomization. Like all other trials in that era, however, the Lanman trial did not use "blinded" examinations of outcome events.

In the results of the Lanman trial [123], one infant in the restricted-oxygen group was lost to follow-up after discharge, at which time the eyes were normal. Among the other infants, the mortality rates were 20% (9/45) in the liberal-oxygen and 30% (12/40) in the restricted-oxygen group—a difference regarded as not statistically significant ($p = 0.29$, by chi-square test). The corresponding rates of permanently scarring RLF in the survivors, however, were 22% (8/36) in the liberal group and 0% (0/28) in the restricted group (two-tailed $p = 0.0068$, by Fisher test).

The estrogen treatment in the trial had no effect upon mortality: 11 deaths occurred in the 45 (24%) infants treated with estrogen and 10 in 40 (25%) of those not treated. Permanently scarring RLF occurred in 3 of 34 (8.8%) survivors in the estrogen treated group and in 5 of 30 (17%) not so treated. The difference, however, was not statistically significant (two-tailed $p = 0.71$, by Fisher test).

In 1954, the year when the Lanman trial was published and preliminary results announced for the National Cooperative Study, Harry H. Gordon *et al.* [96] published a retrospective cohort study contrasting results at their institution for infants treated from 1947 to 1950 with liberal oxygen ($\geqslant$60% on average) vs infants treated from 1951 to 1953 with low concentrations of oxygen, carefully maintained at 30–40% for shorter durations. The investigators found residua of RLF present in 45% (9/20) of the earlier group and in 8% (8/97) of the later group (two-tailed $p = 0.0002$, by Fisher test). The group who received restricted oxygen in the later years had no reduction in survival.

In 1956, the final report of the National Cooperative Study [111] confirmed the preliminary results: liberal-oxygen therapy produced a striking increase in RLF and a slight but statistically non-significant increase in mortality. Death before 40 days of life occurred in 15 (22%) of 68 infants enrolled in the liberal-oxygen group and in 151 (21%) of 718 infants in the curtailed-oxygen group. Some active form of RLF developed in 38 (72%) of the 53 survivors of the liberal-oxygen group and in 178 (33%) of the 533 survivors in the curtailed-oxygen group ($p \ll 0.001$ by chi square).

In the final report [111], the investigators noted that RLF did appear while oxygen was being received and did not immediately occur when oxygen was withdrawn. This finding refuted the previous beliefs [100–104] that RLF occurred only after withdrawal from therapeutic oxygen, resulted from rapid weaning of oxygen, and improved with oxygen therapy.

By the time of the final report, the RLF epidemic had begun to wane. The erroneous rapid-wean hypothesis had fortunately led to a policy of avoiding oxygen therapy [106–108] although some academic medical centers were still giving it routinely [124] to all premature infants as late as 1954. By 1956, however, the National Cooperative Study had brought the consensus—expressed in journal editorials and pediatric textbooks [125–129]—that using therapeutic oxygen can lead to RLF and that oxygen should be given only if clinically necessary.

The iatrogenic epidemic of retrolental fibroplasia was over.

Subsequent uncertainties. The results of the National Cooperative Study impressively demonstrated [111] that prolonged exposure to therapeutic oxygen was particularly likely to result in RLF in presumably premature infants of low birth weight who survived the first 48 hours. Having found no large quantitative difference in death rates, the investigators also concluded that the restriction of oxygen would not affect mortality. In later statistical analyses [91, 130], however, the sample size of the study was shown to be too small to detect a sizable difference in mortality.

The full report [111] emphasized that no differences in the rates of cicatricial, or scarring, RLF were found when oxygen concentration was between 30–39% and 40–49%. The corresponding rates of RLF were 6.2% (8/130) and 5.6% (7/124) respectively. The investigators concluded that "cicatricial RLF for the most part was not dependent upon the *concentration* of *oxygen* administered." This conclusion conflicted with the Bellevue-study [123] recommendation that oxygen concentrations be kept below 40%. Instead, the National Cooperative Study [111] stated "... for all practical purposes there is no concentration of oxygen in excess of that in air that is not associated with risk of developing RLF." Many pediatricians, however, decided [131] that the risk of RLF was lower with lower concentrations of oxygen, despite the absence of experimental data separating the effects of concentration and duration.

Another limitation of the National Cooperative Study was that it could not address the use of oxygen in the first 48 hours of life, because infants were enrolled only after that period. In fact, among 1420 potentially eligible infants at the cooperating institutions, 634 (45%) died before 48 hours, so that the trial included only the surviving half of the population at risk. Nevertheless, the results of the trial were being used for conclusions that all use of oxygen should be curtailed, even in the first 48 hours of life [132].

In the years that followed, when many clinicians concluded that the restriction of oxygen would completely eliminate RLF [126, 127, 133], its rates sharply decreased, but mortality increased [134]. A statistical association was soon demonstrated [135] between restriction of oxygen and increased mortality from the respiratory distress syndrome. Furthermore, RLF was not consistently associated with supplemental oxygen and was repeatedly found in premature infants who had received none [136]. In other infants, who received liberal oxygen, RLF did not occur at all or affected only one eye [111]. Another item of inconsistent evidence came from a second multicenter study directed

by Kinsey [137] and published in 1977. The investigators found that RLF had no distinct association with *arterial* oxygen concentrations and no correlation with administered oxygen concentration, except at very high concentrations (≧80%) and in infants weighing <1201 g.

Finally, despite the originally proclaimed victory over RLF, including Reese's statement that RLF "has now bowed out" [126], a second and still unabated epidemic of RLF [138–141] began in the mid-1960s. The second epidemic was eventually attributed to the improved survival of very low birth-weight infants [139–141] who would not have been included in the previous clinical observations and research. Some of the increase in RLF, however, might have arisen from a better diagnostic detection of mild, transient RLF [91].

Randomized clinical trials had thus unequivocally demonstrated the role of oxygen therapy as a contributory cause of RLF. As with other randomized trials, however, the subsequent therapeutic actions did not produce unequivocally successful results. Although RLF became sharply reduced, infants with respiratory distress may sometimes have been deprived of life-sustaining oxygen. Furthermore, the decreased use of oxygen therapy did not eliminate RLF. In fact, proliferative retinopathy—the end stage of RLF—has sometimes occurred in infants who never received oxygen therapy [136]. Thus the hyperoxigenic epidemic of RLF has ended, but the disease still occurs.

LESSONS TO BE LEARNED

The long saga of the RLF–oxygen epidemic has three striking phenomena that probably would not occur in an analogous situation today.

1. Identification of epidemic

The first phenomenon was the relative absence of early attention to the epidemic itself. Without good national surveillance mechanisms, pediatricians and ophthalmologists were probably unaware of how many children were being blinded as the epidemic began to increase in magnitude and scope. The type of national surveillance now offered by agencies such as the Centers for Disease Control would probably bring much earlier emphasis to an analogous epidemic occurring today.

2. Paucity of direct research

The second phenomenon was the paucity of direct clinical research in the first few years of the epidemic. Without ample funding from national sources, direct clinical research was not a common activity in that era. The relatively few clinical investigators who did research usually aimed at specifically defined projects whose goals could not be readily altered. This problem would probably not recur today because suitable *ad hoc* research funds would be made available from both federal and private sources.

3. Absence of pediatric investigators

A third phenomenon was the relative absence of pediatricians from the research activities. Although the infants with RLF were receiving direct observation and care from pediatricians, almost all of the research reports during the first half of the epidemic were prepared by nonpediatricians. Ophthalmologists were investigating the morphologic characteristics of the ocular lesion, and public-health epidemiologists were looking for possible causative factors, but pediatricians were seldom included among the investigative *dramatis personae*. This problem was recognized in an early comment by Terry [31]:

> "To have a really close relationship between the pediatrician and the ophthalmologist, not only in routine clinical studies, but also in experimental and investigative research, would speed the solution in a mutually advantageous manner."

The pediatric clinical investigators of the 1942–1954 era worked mainly in laboratories, with animals or inanimate substances, exploring experimental issues in pathophysiology. With a laboratory focus in research, the investigators might not have felt comfortable doing the patient-centered studies required for RLF.

The absence of pediatricians, however, deprived the ophthalmologists and other nonpediatric investigators of clinical consultants or collaborators who were intimately aware of changes in the use of oxygen for premature infants. The possible etiologic role of oxygen seemed to receive no serious investigative attention until it appeared in 1949 (7 years after the first cases were reported) on the list of 47 variables explored by Kinsey and Zacharias [75]. This problem would not occur today because so many epidemiologists have received post-doctoral clinical training and because so many pediatricians have become specialized for research as clinical epidemiologists. Such pedia-

tricians would doubtlessly become promptly involved in exploring an analogous epidemic today.

The other investigative problems, however, were methodologic difficulties that can occur in any era. These problems are discussed in the next few sections.

4. Diagnostic criteria

The early research activities suffered from observer variability in diagnostic criteria—a problem that has occurred many times before and since the RLF epidemic. Whenever a disease must be identified clinically, without an easily applied objective test, substantial problems can occur until specific operational criteria have been firmly established and checked for their consistent application. In the instance of RLF, different ophthalmologists used different visual observations and criteria for their diagnostic identifications, but made no apparent effort to standardize their activities. Reese [63] acted admirably in 1949 when he admitted his earlier error in admixing two diseases and when he proposed new criteria for distinguishing RLF. By that time, however, the epidemic had been in progress for 7 years and many research efforts had been wasted by being aimed in the wrong direction of a congenital etiology.

5. Other methodologic problems

Several other methodologic problems in the RLF epidemic have been recurrent scientific difficulties in cause–effect research at every era in medical history. The available "evidence" was wrong because the raw data were inaccurate or the comparisons were biased. These problems can be catalogued [15] in the categories of *susceptibility*, *performance*, *detection*, and *transfer*.

Problems in susceptibility. (i) Because premature infants were particularly susceptible to RLF, the use of full-term infants in control groups of case–control studies led to the erroneous conclusion that RLF was caused by conditions that led to prematurity. This problem could have been avoided with better criteria [15] for selection of control groups in case–control studies. (ii) In clinical trials where the comparisons were arranged by alternation or other non-randomized methods, the contrasted groups contained unequal numbers of patients and probable inequalities in susceptibility to RLF. Thus, the patients treated "successfully" with agents such as ACTH or vitamin E may have entered the trial at different times after birth and may have received different preceding exposures to oxygen. (iii) In the absence of cohort studies, the initial investigators did not recognize the "natural history" in which RLF could stop progressing or even regress after its initial development. Because of this error, late treatment with oxygen was given false credit for the apparent improvements.

Problems in performance. (i) Because the actual performance of the oxygen treatment was not carefully monitored, the New Orleans data were erroneous in contentions that oxygen therapy, which was actually administered ineffectually, had not produced RLF. Conversely, the "restricted use" of oxygen that seemed to produce RLF in the Paris data was actually quite liberal. (ii) When vitamins and iron were regarded as causal maneuvers, investigators did not give adequate attention to the concomitant performance of co-maneuvers, such as oxygen therapy. (iii) When data were unknown regarding dosage of iron and vitamins, the investigators assumed, probably erroneously, that the treatment was not given. (iv) When rapid-weaning and slow-weaning of oxygen were contrasted, the investigators inadvertently gave the rapidly weaned group a higher concentration and longer duration of oxygen whose magnitude was analytically ignored.

Problems in detection. (i) In the absence of objective examination procedures, the diagnosis of RLF in certain studies may have been biased by the observer's awareness of antecedent treatment. (ii) Reluctant to discontinue oxygen treatment so that the eyes could be carefully examined, the investigators were delayed in discovering that RLF was absent at birth and developed only later. (iii) In certain secular comparisons of incidence of RLF, the data for previous years contained diagnoses made during and after hospitalization, whereas the data for recent years contained only diagnoses made during hospitalization.

Problems in transfer. (i) The use of cross-sectional rather than cohort studies led to diagnostic detection problems and the erroneous belief that RLF was present at birth rather than a postnatal development. (ii) Because death and RLF were "competing events", the occurrence and analysis of the two events became confused, particularly when the true occurrence of RLF was reduced by the premature deaths provoked by ACTH treatment. (This problem still occurs

and still has not been fully resolved in randomized trials.)

6. Mathematical issues

Two additional problems arose from mathematical issues in statistics rather than clinical methods in getting data and comparing groups.

Inadequate sample sizes. One mathematical problem was the relatively small sample size in the National Cooperative Study. Although the results showed unequivocally that a restricted use of oxygen would prevent RLF, the numbers were not large enough to demonstrate definitively that respiratory deaths could be avoided as an adverse side effect of the restricted use.

Multiple comparisons. A second problem was the statistical possibility that a "significant" result—such as the presumptive protective value of iron and vitamins—might occur by chance during the multiple comparisons tested when the investigators checked relationships between RLF and about 50 candidate causes. The best way to manage this problem—either mathematically, by changing the α level, or scientifically, by developing better specifications for the research [15]—is still unresolved.

7. Role of animal research

Another still unresolved problem is the pertinence of animal research for etiologic decisions in non-infectious human disease. Although a suitable animal model of RLF has never been developed, animal studies provided support for contradictory etiologic theories about both hypoxia and hyperoxia, and animal research produced erroneous ideas about the etiologic role of vitamin A.

8. Wrong interpretations of correct data

Finally, the last methodologic problem to be cited has also occurred in cause–effect research at every era in medical history: failure to consider alternative interpretations of accurate data. The error of implicating postnatal relative hypoxia due to rapid weaning from oxygen would have been difficult to avoid even today. The secular association data, the early randomized-trial data, and the lower rate of RLF found in slow-wean groups were all accurate and consistent with both the hyperoxia and the relative-hypoxia hypotheses. Probably the only way to solve this problem would have been to test both hypotheses simultaneously (as was done in the definitive randomized trial) or to persuade proponents of the rapid-wean hypothesis that slow weaning was successful because it also reduced the total exposure of oxygen.

CONCLUSIONS

Some of the problems that prolonged the RLF epidemic would not occur today because of changes in personnel and research support. The epidemic itself would be noted more promptly and would be investigated by researchers who were more oriented to the complete clinical environment in which the epidemic occurred. Other problems, however, such as the lack of standardized distinctive diagnostic criteria, can teach lessons that do not yet seem to have been fully learned. In particular, rigorous criteria have not been established or applied for the death certificate diagnoses used to monitor the incidence of cancer and atherosclerosis that are major non-infectious "epidemics" receiving research today; and standardized, distinctive criteria for the diagnosis of AIDS have been recently given a deliberately reduced precision.

The other methodologic problems of accuracy or biased comparison refer to difficulties or inequalities in susceptibility, performance, detection, and transfer of compared groups. These problems remain unresolved, and various proposed solutions have not evoked consensus. Although the neglect of susceptibility in case–control comparisons supported the erroneous idea that RLF was congenital, many epidemiologists today continue to reject suitable attention to susceptibility, claiming incorrectly that the adjustments would produce "overmatching". The erroneous RLF data from New Orleans and Paris arose because the actual performance of the oxygen therapy was not carefully assessed. Nevertheless, despite the known frequency with which patients fail to comply with therapeutic recommendations, many prescribed agents, in both epidemiologic studies and clinical trials, are analyzed as though each agent had been received as intended. The equal detection of outcome events, although carefully planned with "double-blinding" in randomized trials, continues to be generally overlooked as a source of bias in non-randomized comparisons, particularly when the outcome event is the diagnostic detection of often "silent" diseases, such as cancer and atherosclerosis. The problems of transfer bias in selected groups of people continue to receive relatively little attention in case–control studies or in "convenience cohort" research.

The role of these unresolved methodologic problems will be interesting to examine when analogous "autopsies" eventually become possible for the persisting current epidemics of cancer, atherosclerosis, and other non-infectious diseases.

Statistical problems in sample size now receive much more attention than in the 1950s. Many studies remain undersized not because of neglect, however, but because of pragmatic difficulty in getting large groups assembled for the research. The meta-analyses that produce large sample sizes by pooling results from disparate studies have become popular among statisticians and in media reports, but the scientific validity of the process remains uncertain [142]. The problem of multiple statistical comparisons has not received a consensus solution about when the adjustments should be made or what mechanism should be used.

Animal research continues to be conducted in search of causes for human disease, and etiologic suspicions continue to arise and to produce controversy when diseases are produced by unrealistically huge doses of agents, like those that indicted vitamin A as a cause of RLF.

Finally, the erroneous interpretation of accurate data remains as much of a scientific hazard today as it was 140 years ago when high atmospheric pressure was deemed a cause of cholera, 78 years ago when pellagra was regarded as infectious, 39 years ago when RLF was attributed to rapid weaning from oxygen, or a few years ago, when AIDS was ascribed to amyl nitrite "poppers".

Thus, the epidemic of RLF would probably be recognized more rapidly today and would receive more prompt investigative attention. The randomized trial that unequivocally indicted the oxygen therapy would doubtlessly have been conducted much sooner in the course of the epidemic. The other methodologic problems that helped contribute to the persistence of the epidemic, however, continue to offer lessons that have been unrecognized, unlearned, or unresolved.

Acknowledgements—We are grateful to Drs George B. Logan, Thomas K. Oliver and William A. Silverman for valuable comments and suggestions, and to Mary Lyman for typing the manuscript. Dr Feinstein also thanks the Council for Tobacco Research for unrestricted support given to the Clinical Epidemiology Unit at Yale.

REFERENCES

1. Vandenbroucke JP, Pardoel VPAM. An autopsy of epidemiologic methods: the case of 'poppers' in the early epidemic of the acquired immunodeficiency syndrome (AIDS). **Am J Epidemiol** 1989; 129: 455–457.
2. Silverman WA. **Retrolental Fibroplasia: a Modern Parable**. New York: Grune and Stratton; 1980.
3. Lilienfield AM. Epidemiologic methods and inferences in studies of noninfectious diseases. **PAR Rev** 1957; 72: 51–60.
4. Yerushalmy J, Palmer CE. On the methodology of investigations of etiologic factors in chronic diseases. **J Chron Dis** 1959; 10: 27–40.
5. Sartwell PE. On the methodology of investigations of etiologic factors in chronic diseases—further comments. **J Chron Dis** 1960; 11: 61–63.
6. Susser M. **Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology**. New York: Oxford University Press; 1973.
7. Sackett DL. The diagnosis of causation. In **Strategies in Clinical Research. Mead Johnson Symposium on Perinatal and Development Medicine**. No. 9. Evansville, IN: Mead Johnson; 1977.
8. Bollet AJ. On seeking the cause of disease. **Clin Res** 1964; 12: 305–310.
9. Feinstein AR. Scientific standards in epidemiologic studies of the menace of daily life. **Science** 1988; 242: 1257–1263.
10. Evans AS. Causation and disease: the Henle–Koch postulates revisited. **Yale J Biol Med** 1976; 49: 175–195.
11. Popper K. **Conjectures and Refutations**. New York: Basic Books; 1963.
12. Buck C. Popper's philosophy for epidemiologists. **Int J Epidemiol** 1975; 4: 159–168.
13. Rothman KJ, Ed. **Causal Inference**. Chestnut Hill, MA: Epidemiology Resources, Inc.; 1988.
14. Weed DL. On the logic of causal inference. **Am J Epidemiol** 1986; 123: 965–979.
15. Feinstein AR. **Clinical Epidemiology: the Architecture of Clinical Research**. Philadelphia: Saunders; 1985.
16. Farr W. Influence of elevation on the fatality of cholera. **J Stat Soc** 1852; 15: 155–183.
17. Eyler JM. William Farr on the cholera: the sanitarian's disease theory and the statistician's method. **J Hist Med Allied Sci** 1973; 28: 79–100.
18. Siler JF, Garrison PE, MacNeal WJ. A statistical study of the relation of pellagra to use of certain foods and to location of domicile in six selected industrial communities. **Arch Intern Med** 1914; 14: 293–373.
19. Boston Collaborative Drug Surveillance Program. Reserpine and breast cancer. **Lancet** 1974; 2: 669–671.
20. Armstrong B, Stevens N, Doll R. Retrospective study of the association between use of rauwolfia derivatives and breast cancer in English women. **Lancet** 1974; 2: 672–675.
21. Heinonen OP, Shapiro S, Tuominen L, Turunen MI. Reserpine use in relation to breast cancer. **Lancet** 1974; 2: 675–677.
22. Labarthe DR. Methodologic variation in case–control studies of reserpine and cancer. **J Chron Dis** 1979; 32: 95–104.
23. Shapiro S, Slone D. Comment (in discussion of case-control studies). **J Chron Dis** 1979; 32: 105–107.
24. Horwitz RI, Feinstein AR. Exclusion bias and the false relationship of reserpine and breast cancer. **Arch Intern Med** 1985; 145: 1873–1875.
25. Terry TL. Extreme prematurity and fibroblastic overgrowth of persistent vascular sheath behind each

crystalline lens: i. Preliminary report. **Am J Ophthalmol** 1942; 25: 203–204.
26. Terry TL. Fibroblastic overgrowth of persistent tunica vasculosa lentis in premature infants: ii. Report of cases—clinical aspects. **Arch Ophthalmol** 1943; 29: 36–53.
27. Terry TL. Fibroblastic overgrowth of persistent tunica vasculosa lentis in infants born prematurely: iii. Studies in development and regression of hyaloid artery and tunica vasculosa lentis. **Am J Ophthalmol** 1942; 25: 1409–1421.
28. Terry TL. Retrolental fibroplasia. **Am J Dis Child** 1944; 68: 75–76.
29. Terry TL. Fibroblastic overgrowth of persistent tunica vasculosa lentis in premature infants: iv. Etiologic factors. **Arch Ophthalmol** 1943; 29: 54–68.
30. Terry TL. Retrolental fibroplasia in premature infants: v. Further studies in fibroblastic overgrowth of persistent tunica vasculosa lentis. **Arch Ophthalmol** 1945; 33: 203–208.
31. Terry TL. Retrolental fibroplasia. **J Pediatr** 1946; 29:770–773.
32. Sitchevska O. Retrolental fibroplasia in premature infants: Terry's syndrome. **Arch Ophthalmol** 1946; 36: 362–363.
33. Hess JH. Oxygen unit for premature and very young infants. **Am J Dis Child** 1934; 47: 916–917.
34. Boothby WM. Miniature oxygen chamber for infants; a modification of the Hess incubator. **Proc Staff Meetings Mayo Clinic** 1934; 9: 129–131.
35. Chapple CC. An incubator for infants. **Am J Obstet Gynecol** 1938; 35: 1062–1065.
36. Dunham EC, Dickinson HC, Gowens GJ, Witters J. Incubators for premature infants. **Am J Public Health** 1940; 30: 1415–1421.
37. Bakwin H. Oxygen therapy in premature babies with anoxemia. **Am J Dis Child** 1923; 25: 157–162.
38. Wilson JL, Long SB, Howard PJ. Respiration of premature infants: response to variations of oxygen and to increased carbon dioxide in inspired air. **Am J Dis Child** 1942; 63: 1080–1085.
39. Smith CA, Kaplan E. Adjustment of blood oxygen levels in neonatal life. **Am J Dis Child** 1942; 64: 843–859.
40. Anderson NA. Care of the premature infant. In: Nelson WE, Ed. **Mitchell–Nelson's Textbook of Pediatrics**. Philadelphia: WB Saunders; 1945.
41. Bloxsom A. A study of premature infants observed in a private hospital before and after the establishment of a premature nursery. **J Pediatr** 1946; 28: 418–421.
42. Clifford SH. Care of the newborn. **N Engl J Med** 1949; 240: 61–67.
43. Dancis J, Cardullo HM. Incubator care of the premature infant: discussion of optimal conditions and the current methods for their provision. **Pediatrics** 1950; 6: 432–440.
44. Reese AB, Payne F. Persistence and hyperplasia of the primary vitreous (tunica vasculosa lentis or retrolental fibroplasia). **Am J Ophthalmol** 1946; 29: 1–24.
45. Krause AC. Congenital encephalo-ophthalmic dysplasia. **Arch Ophthalmol** 1946; 36: 387–444.
46. Klien BA. Histopathologic aspects of retrolental fibroplasia. **Arch Ophthalmol** 1949; 41: 553–561.
47. Ingalls TH. Congenital encephalo-ophthalmic dysplasia: epidemiologic implications. **Pediatrics** 1948; 1: 315–326.
48. Ingalls TH. Etiology of mongolism: epidemiologic and teratologic implications. **Am J Dis Child** 1947; 74: 147–165.
49. Ingalls TH. The study of congenital anomalies by the epidemiologic method: with a consideration of retrolental fibroplasia as an acquired anomaly of the fetus. **N Engl J Med** 1950; 243: 67–74.
50. Ingalls TH. Epidemiology of encephalo-ophthalmic dysplasia. **JAMA** 1948; 138: 261–264.
51. Warkany J, Schraffenberger E. Congenital malformations induced in rats by maternal vitamin A deficiency: i. Defects of the eye. **Arch Ophthalmol** 1946; 35: 150–169.
52. Terry TL. Retrolental fibroplasia. **Adv Pediatr** 1948; 3: 55–67.
53. Kinsey VE, Chisholm JF. Retrolental fibroplasia: evaluation of several changes in dietary supplements of premature infants with respect to the incidence of the disease. **Am J Ophthalmol** 1951; 34: 1259–1268.
54. Feinstein AR. The Jones criteria and the challenges of clinimetrics. **Circulation** 1982; 66: 1–5.
55. Feinstein AR. **Clinimetrics**. New Haven: Yale University Press; 1987.
56. Jones TD. The diagnosis of rheumatic fever. **JAMA** 1944; 126: 481.
57. Krause AC. Retrolental fibroplasia. **Arch Ophthalmol** 1946; 36: 364–368.
58. Owens WC, Owens EU. Retrolental fibroplasia in premature infants. **Am J Ophthalmol** 1949; 32: 1–21.
59. Owens WC, Owens EU. Retrolental fibroplasia in prematures: ii. Studies on the prophylaxis of the disease: the use of alpha tocopheryl acetate. **Am J Ophthalmol** 1949; 32: 1631–1637.
60. Gilger AP. Retrolental fibroplasia in prematurely born children. **Am J Ophthalmol** 1949; 32: 917–929.
61. La Motte WO, Tynor GS. Observations on retrolental fibroplasia. **Am J Ophthalmol** 1950; 33: 1925–1928.
62. Laupus WE, Bousquet FP Jr. Retrolental fibroplasia: the role of hemorrhage in its pathogenesis. **Am J Dis Child** 1951; 81: 617–626.
63. Reese AB. Persistence and hyperplasia of primary vitreous; retrolental fibroplasia—two entities. **Arch Ophthalmol** 1949; 41: 527–552.
64. Reese AB, Blodi FC, Locke JC. The pathology of early retrolental fibroplasia: with an analysis of the histologic findings in the eyes of newborn and stillborn infants. **Am J Ophthalmol** 1952; 35: 1407–1426.
65. Silverman WA, Blodi FC, Locke JC, Day RL, Reese AB. Incidence of retrolental fibroplasia in a New York nursery. **Arch Ophthalmol** 1952; 48: 698–711.
66. Speert H, Blodi FC, Reese AB. Retrolental fibroplasia—a hazard of premature birth. **Am J Obstet Gynecol** 1950; 59: 246–258.
67. Jackson B, Kinsey VE. The relation between maternal vitamin A intake, blood level, and ocular abnormalities in the offspring of the rat. **Am J Ophthalmol** 1946; 29: 1234–1242.
68. Clifford SH, Weller KF. The absorption of vitamin A in prematurely born infants: with experience in the use of absorbable vitamin A in the prophylaxis of retrolental fibroplasia. **Pediatrics** 1; 1948: 505–511.
69. Dancis J, Lewis JM, Guy LP. Retrolental fibroplasia: a statistical study with particular reference to the effect of vitamin A. **N Engl J Med** 1951; 245: 402–406.
70. Post LT. Retrolental fibroplasia. **J Pediatr** 1950; 36: 833–836.
71. Holt LE. **Report of the Second M & R Pediatric Research Conference: Subject Retrolental Fibroplasia**. Columbus, OH: M &R Laboratories; 1951.
72. Terry TL. Ocular maldevelopment in extremely premature infants; retrolental fibroplasia: vi. General consideration. **JAMA** 1945; 128: 582–585.
73. Hepner WR, Krause AC, Davis ME. Retrolental fibroplasia and light. **Pediatrics** 1949; 3: 824–828.
74. Locke JC, Reese AB. Retrolental fibroplasia: the negative role of light, mydriatrics, and the ophthalmoscopic examination in its etiology. **Arch Ophthalmol** 1952; 48: 44–47.
75. Kinsey VE, Zacharias L. Retrolental fibroplasia: incidence in different localities in recent years and a

correlation of the incidence with treatment given the infants. **JAMA** 1949; 139: 572–578.
76. La Motte WO, Tynor GS. Observations on retrolental fibroplasia. **Am J Ophthalmol** 1950; 33: 1925–1928.
77. Reese AB, Blodi FC. Retrolental fibroplasia: the fifth Francis I Proctor lecture. **Am J Ophthalmol** 1951; 34: 1–24.
78. Oliver TK (Chairman, Institute of Medicine Special Committee), Avery ME, Feinstein AR, Kalina RE, Kensler TW, Kertchmer N, Oski F, Patz A, Perrin EB, Silverman WA, Warshaw J. **Vitamin E and Retinopathy of Prematurity**. Washington, DC: National Academy Press; 1986.
79. Ehrenkranz RA. Vitamin E and retinopathy of prematurity: still controversial. **J Pediatr** 1989; 114: 801–803.
80. Collip JB, Anderson EM, Thomson DL. The adrenotropic hormone of the anterior pituitary lobe. **Lancet** 1933; 2: 347–348.
81. Li CH, Evans HM, Simpson ME. Adrenocorticotropic hormone. **J Biol Chem** 1943; 149: 413–424.
82. Sayers G, White A, Long CNH. Preparation and proporties of the pituitary adrenotropic hormone. **J Biol Chem** 1943; 149: 425–436.
83. Blodi FC, Silverman WA, Day R, Reese AB. Experiences with corticotropin (ACTH) in the acute stage of retrolental fibroplasia. **Am J Dis Child** 1951; 82: 242–243.
84. Laupus WE. Experiences with corticotropin (ACTH) in the acute stage of retrolental fibroplasia. **Am J Dis Child** 1951; 82: 243.
85. Pratt EL. Experiences with corticotropin (ACTH) in the acute stage of retrolental fibroplasia. **Am J Dis Child** 1951; 82: 243–244.
86. Smith CA. Experiences with corticotropin (ACTH) in the acute stage of retrolental fibroplasia. **Am J Dis Child** 1951; 82: 244–245.
87. Reese AB, Blodi FC, Locke JC, Silverman WA, Day RL. Results of use of corticotropin (ACTH) in treatment of retrolental fibroplasia. **Arch Ophthalmol** 1952; 47: 551–555.
88. Campbell K. Intensive oxygen therapy as a possible cause of retrolental fibroplasia: a clinical approach. **Med J Aust** 1951; 11: 48–50.
89. Ryan H. Retrolental fibroplasia: a clinicopathologic study. **Am J Ophthalmol** 1952; 35: 329–342.
90. Crosse VM, Evans PJ. Prevention of retrolental fibroplasia. **Arch Ophthalmol** 1952; 48: 83–87.
91. American Academy of Pediatrics, Committee on Fetus and Newborn. History of oxygen therapy and retrolental fibroplasia. **Pediatrics** 1976; 57 (Suppl.): 591–642.
92. Crosse VM. The problem of retrolental fibroplasia in the city of Birmingham. **Trans Ophthalmol Soc UK** 1951; 71: 609–612.
93. Houlton ACL. A study of cases of retrolental fibroplasia seen in Oxford. **Trans Ophthalmol Soc UK** 1951; 71: 583–590.
94. Exline AL, Harrington MR. Retrolental fibroplasia: clinical statistics from the premature center of the Charity Hospital of Louisiana at New Orleans. *J Pediatr* 1951; 38: 1–7.
95. Lelong M, Rossier A, Fontaine M, LeMasson C, Michelin et Audibert J. Sur la retinopathie des prematures (fibroplasie retrolentale). **Arch Fr Pediatr** 1952; 9: 897–914.
96. Gordon HH, Lubchenco L, Hix I. Observations on the etiology of retrolental fibroplasia. **Bull Johns Hopkins Hosp** 1954; 94: 34–44.
97. Patz A, Hoeck LE, De La Cruz E. Studies on the effect of high oxygen administration in retrolental fibroplasia: i. Nursery observations. **Am J Ophthalmol** 1952; 35: 1248–1253.
98. Silverman WA. Experimentation in children: American style. Unpublished transcript of the first Richard L. Day lectureship. New York, NY: Columbia University College of Physicians and Surgeons; 7–8 April, 1975.
99. Silverman WA. **Human Experimentation: a Guided Step into the Unknown**. Oxford: Oxford University Press; 1985.
100. Szewczyk TS. Retrolental fibroplasia: etiology and prophylaxis. **Am J Ophthalmol** 1951; 34: 1649–1650.
101. Szewczyk TS. Retrolental fibroplasia: etiology and prophylaxis. **Am J Ophthalmol** 1952; 35: 301–311.
102. Jefferson E. Retrolental fibroplasia. **Arch Dis Child** 1952; 27: 329–336.
103. Szewczyk TS. Retrolental fibroplasia and related ocular diseases: classification, etiology, and prophylaxis. **Am J Ophthalmol** 1953; 36: 1336–1361.
104. Bedrossian RH, Carmichael P, Ritter J. Retinopathy of prematurity (retrolental fibroplasia) and oxygen; part i. Clinical study; part ii. Further observations on the disease. **Am J Ophthalmol** 1954; 37: 78–86.
105. Owens WC. Spontaneous regression in retrolental fibroplasia. **Trans Am Ophthalmol Soc** 1953; 51: 555–579.
106. Crosse VM. Prevention of retrolental fibroplasia. In: Parson L, Barling S, Eds. **Diseases of Infancy and Childhood**. London: Oxford University Press; 1954; 2: 1662–1707.
107. McGavic JS. Retinopathy of prematurity (retrolental fibroplasia). In: Nelson WE, Ed. **Textbook of Pediatrics**, edn 6. Philadelphia, PA: WB Saunders; 1954: 1412–1414.
108. Simpson DG. Retina. In: Paterson D, McCreary JF, Eds. **Pediatrics**. Philadelphia: Lippincott; 1956: 498.
109. Editor's column: Anoxia and retrolental fibroplasia. **J Pediatr** 1952; 40: 684–685.
110. Editor's column: Retrolental fibroplasia and oxygen. **J Pediatr** 1954; 44: 122–123.
111. Kinsey VE, Jacobus JT, Hemphill FM. Retrolental fibroplasia: cooperative study of retrolental fibroplasia and the use of oxygen. **Arch Ophthalmol** 1956; 56: 481–543.
112. Bousquet FP, Laupus WE. Studies on the pathogenesis of retrolental fibroplasia. **Am J Ophthalmol** 1952; 35: 64–68.
113. Kinsey VE, Hemphill FM. Etiology of retrolental fibroplasia; preliminary report of a cooperative study of retrolental fibroplasia. **Trans Am Acad Ophthalmol Otolaryngol** 1955; 55: 15–24.
114. Ingalls TH, Curley FJ, Prindle RA. Anoxia as a cause of fetal death and congenital defect in the mouse. **Am J Dis Child** 1950; 80: 34–45.
115. Campbell FW. The influence of a low atmospheric pressure on the development of the retinal vessels in the rat. **Trans Ophthalmol Soc UK** 1951; 71: 287–300.
116. Gyllensten LJ, Hellstrom BE. Retrolental fibroplasia—animal studies: the effect of intermittingly administered oxygen on the postnatal development of the eyes of full term mice. A preliminary report. **Acta Pediatr** 1952; 41: 577–582.
117. Hepner WR. Retrolental fibroplasia in premature infants and in kittens. **Am J Dis Child** 1952; 84: 748–749.
118. Hepner WR. Retrolental fibroplasia: animal studies; induced vascular engorgement and hyperplasia of the iris, tunica vasculosa lentis and retina in kittens. **Pediatrics** 1952; 9: 602–605.
119. Ingalls TH, Tedeschi CG, Helpern NM. Congenital malformations of the eye induced in mice by maternal anoxia; with particular reference to the problem of retrolental fiborplasia in man. **Am J Ophthalmol** 1952; 35: 311–329.

120. Patz A, Eastham A, Higginbotham DH, Kleh T. Oxygen studies in retrolental fibroplasia: ii. The production of the microscopic changes of retrolental fibroplasia in experimental animals. **Am J Ophthalmol** 1953; 36: 1511–1522.
121. Ashton N, Ward B, Serpell G. Role of oxygen in the genesis of retrolental fibroplasia: a preliminary report. **Br J Ophthalmol** 1953; 513–520.
122. Ashton N, Ward B, Serpell G. Effect of oxygen on developing retinal vessels with particular reference to the problem of retrolental fibroplasia. **Br J Ophthalmol** 1954; 38: 397–432.
123. Lanman JT, Guy LP, Dancis J. Retrolental fibroplasia and oxygen therapy. **JAMA** 1954; 155: 223–226.
124. Gordon HH. Oxygen administration and retrolental fibroplasia; editorial comment. **Pediatrics** 1954; 14: 543–546.
125. Spekter L. **The Pediatric Years; a Guide in Pediatrics for Workers in Health, Education, and Welfare**. Springfield: Thomas; 1955: 483.
126. Reese AB. An epitaph for retrolental fibroplasia. **Am J Ophthalmol** 1955; 38: 267–269.
127. Foote FM. Editorial. A new opportunity—preventing blindness from retrolental fibroplasia. **Am J Public Health** 1955; 45: 795–797.
128. Kinsey VE. Retrolental fibroplasia. **Am J Ophthalmol** 1955; 39: 105–107.
129. Hardyment AF. The premature baby. In: Paterson D, McCreary JF, Eds. **Pediatrics**. Philadelphia: Lippincott; 1956: 64–67.
130. Day RL, Dell RB, Darlington RB, Fleiss JL. Interpretation of statistical nonsignificance and retrolental fibroplasia. **Pediatrics** 1979; 63: 342–344.
131. Kinsey VE. Caution in oxygen therapy to prevent retrolental fibroplasia. **Pediatrics** 1956; 18: 511.
132. Gordon HH. Oxygen therapy and survival rate in prematures. **Pediatrics** 1957; 19: 967.
133. Guy LP, Lanman JT, Dancis J. The possibility of total elimination of retrolental fibroplasia by oxygen restriction. **Pediatrics** 1956; 17: 247–249.
134. Cross KW. Cost of preventing retrolental fibroplasia? **Lancet** 1973; 2: 954–956.
135. Avery ME, Oppenheimer EH. Recent increase in mortality from hyaline membrane disease. **J Pediatr** 1960; 57: 553–559.
136. Editorial: Retrolental fibroplasia (RLF) unrelated to oxygen therapy. **Br J Ophthalmol** 1974; 58: 487–489.
137. Kinsey VE, Arnold HJ, Kalina RE *et al.* PaO2 levels and retrolental fibroplasia: a report of the cooperative study. **Pediatrics** 1977; 60: 655–668.
138. Akeson N, Bonapart B. What is the current incidence of RLF? **Pediatrics** 1976; 58: 627.
139. Mushin AS. Retinopathy of prematurity—a disease of increasing incidence? **Trans Ophthalmol Soc UK** 1974; 94: 251–257.
140. Gibson DL, Sheps SB, Schechter MT, Wiggins S, McCormick AQ. Retinopathy of prematurity: a new epidemic? **Pediatrics** 1989; 83: 486–492.
141. Valentine PH, Jackson JC, Kalina RE, Woodrum DE. Increased survival of low birth weight infants: impact on the incidence of retinopathy of prematurity. **Pediatrics** 1989; 84: 442–45.
142. Gerbarg Z, Horwitz RI. Resolving conflicting clinical trials: a new approach to meta-analyses. **J Clin Epidemiol** 1988; 41: 503–509.