American Journal of Medical Genetics 73:251–258 (1997)

The Return of Thalidomide: Are Birth Defects Surveillance Systems Ready?

Quanhe Yang,[1]* Muin J. Khoury,[2] Levy M. James,[3] Richard S. Olney,[3] Len J. Paulozzi,[3] and J. David Erickson[3]

[1]*Epidemic Intelligence Service, Epidemiology Program Office, and Division of Birth Defects and Developmental Disabilities, National Center for Environmental Health, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia*
[2]*Office of Genetics and Disease Prevention, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia*
[3]*Birth Defects and Genetic Diseases Branch, Division of Birth Defects and Developmental Disabilities, National Center for Environmental Health, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia*

In the 1960s, thalidomide caused limb deficiencies in thousands of infants worldwide. The limb deficiencies were frequently of the intercalary type. As a result, numerous countries started birth defect surveillance programs. In 1967, the Centers for Disease Control (CDC) started the Metropolitan Atlanta Congenital Defects Program (MACDP), a population-based surveillance system, to provide early warning against new teratogens. Recent studies have shown that thalidomide may be beneficial for a range of conditions, including cancer and AIDS, and it may once again become widely available. Here, we examine the ability of MACDP to detect an increase in the birth prevalence of limb deficiency as an early warning of fetal exposure to thalidomide. We calculated base rates for all limb deficiencies, for bilateral nonsyndromic intercalary or preaxial deficiencies, and for all nonsyndromic intercalary limb deficiencies among Atlanta infants born from 1968 through 1993. We used relative risk estimates from previous studies and a range of pregnancy exposure rates for thalidomide. We tested the statistical power of MACDP to detect subtle changes in the birth prevalence of these defects using Poisson and cumulative sum (CUSUM) techniques. The base rates for all limb deficiencies, for bilateral intercalary or preaxial deficiencies, and for all intercalary limb deficiencies, were 0.53, 0.035, and 0.022/1,000, and the estimated relative risks were 175, 4,570, and 8,180, respectively. We varied the assumed rate of exposure to thalidomide from 1/10,000 to 5/100. With a 1/1,000 exposure rate, both Poisson and CUSUM techniques will detect a rate change in intercalary limb deficiency in about 6 months of monitoring, and a rate change in bilateral intercalary or preaxial deficiencies in about 12 months of monitoring. When monitoring all limb deficiencies, a pregnancy exposure rate of 3.5% or less would go unnoticed by the Poisson method and would take more than 50 years for the CUSUM method to signal an alarm with a 1/1,000 exposure rate. However, for rates of exposure less than 1/1,000, a progressively longer period of time or larger sample are needed to detect a rate change by both methods. Our findings highlight the importance of enlarging the monitored population and correct case classification in birth defects surveillance. Am. J. Med. Genet. 73:251–258, 1997. © 1997 Wiley-Liss, Inc.†

KEY WORDS: thalidomide; birth defect; surveillance; epidemiologic method

INTRODUCTION

The thalidomide tragedy during 1958–1962 clearly showed that the classical probation of drugs before marketing gave no reliable protection to the human embryo. The tragedy sensitized manufacturers, governments, health professionals, and the public to the problem of birth defects and possible teratogenicity of drugs [Lenz, 1985]. Consequently, many countries established birth defects surveillance programs to serve as an early warning system should a new teratogen become widely spread in the environment [Källén and

*Correspondence to: Quanhe Yang, Birth Defects and Genetic Disease Branch, BDDD, NCEH, 4770 Buford Highway, MS F-45, Atlanta, GA 30341. E-mail: qay0@cdc.gov

Received 4 October 1996; Accepted 18 June 1997

Winberg, 1968; Källén et al., 1984a; Czeizel, 1973; Flynt and Hay, 1979a; Edmonds et al., 1981; Oakley, 1985; Holtzman and Khoury, 1986].

Although the birth defects surveillance systems in the world provide invaluable data for descriptive and analytical epidemiologic studies, their effectiveness in detecting subtle epidemics has been criticized [Chen, 1979, 1985; Klingberg et al., 1983; Källén et al., 1984a; Khoury and Holtzman, 1987]. The classification of malformations used in monitoring has been discussed [Bod and Czeizel, 1981; Källén et al., 1984b; Holtzman and Khoury, 1986]. Many recommendations have been made to improve the ability of birth defects monitoring to detect new teratogens [Källén et al., 1984a; Holtzman and Khoury, 1986; Khoury and Holtzman, 1987; Lynberg and Edmonds, 1992; Khoury and Edmonds, 1994].

Recent studies have shown that thalidomide may be beneficial for a range of conditions including cancer and AIDS [Burley, 1986; Maknokawkeyoon et al., 1993; D'Arcy and Griffin, 1994]. If thalidomide becomes widely available, stringent control measures must be taken to prevent the exposure of pregnant women, though the proportion of women at risk may be small [Jenkinson, 1993; Erickson, 1995; Castilla et al., 1997].

In this paper, we address two questions. First, if thalidomide becomes widely available, will birth defects surveillance systems be able to detect an increase in the rate of a specific limb deficiency and total limb deficiencies if only a small proportion of pregnant women are exposed? Second, what are the limitations of birth defects monitoring in detecting true subtle increases in the birth prevalence of specific birth defects?

To answer these questions, we used data from the Metropolitan Atlanta Congenital Defects Program (MACDP), a population-based surveillance system. The results of the study can also be used to evaluate the ability of birth defects monitoring to detect true subtle changes of birth defects under different circumstances of exposure frequency, teratogen potencies and etiologic heterogeneity of the outcome.

METHODS

The Metropolitan Atlanta Congenital Defects Program (MACDP) is a population-based birth defects surveillance system that has been in operation since 1967 [Edmonds et al., 1981; Lynberg and Edmonds, 1992; Khoury and Edmonds, 1994]. It monitors all births occurring in the five-county metropolitan Atlanta area. The number of births monitored has increased from about 25,000 per year in 1968 to about 40,000 per year in 1994. One main objective of MACDP is to monitor regularly and systematically the birth of malformed infants in order to detect changes in rates or unusual patterns suggesting environmental influences.

MACDP includes information on all live-born and stillborn infants, with at least one major birth defect with onset during the infants' first year of life. All diagnoses must be ascertained within their first 5 years of life [Lynberg and Edmonds, 1992]. Case ascertainment includes the review of maternal and infant medical records from multiple sources, including birth hospitals, pediatric referral hospitals, and cytogenetic laboratories, and the review of vital statistics from the Georgia Department of Human Resources.

MACDP case records include basic demographic information, the case diagnosis, birth-related information, birth complications, prenatal data, pregnancy and family history, cytogenetic data, and information on other risk factors. Data on major birth defects are analyzed quarterly for changes in rates and any other unusual patterns. Until the early 1990s, the Poisson method was used to detect increases in birth defects rates [Khoury and Edmonds, 1994]. Since then, the cumulative sum (CUSUM) technique [Lucas, 1985] has been employed.

Limb deficiencies were defined according to the classification system of the International Clearinghouse for Birth Defects Monitoring Systems and EUROCAT [International Clearinghouse for Birth Defects Monitoring Systems, 1984]. Limb deficiencies associated with thalidomide have frequently been the intercalary type. Recently Castilla et al. (1997) have suggested that birth defects surveillance systems monitor a "thalidomide-like phenotype" defined as bilateral upper and/or lower limb deficiencies of the preaxial and/or intercalary types. In our study, infants were considered to have preaxial limb deficiencies if the radial or tibial parts of their limbs were absent or severely hypoplastic. Infants with absent thumbs and intact radii were included in the preaxial category. Infants were considered to have intercalary defects when proximal parts of limbs (e.g., humerus, radius, and ulna) were absent or severely hypoplastic. while distal structures (e.g., hand) were totally or partially present. In the MACDP, intercalary and preaxial limb deficiencies are subcategories of all limb deficiencies.

In the present study, we examined the ability of the program to detect a rate change of total limb deficiencies (TLD), bilateral nonsyndromic intercalary or preaxial deficiencies (BIPD), and intercalary limb deficiency (ILD). We excluded infants from the intercalary or preaxial categories when their limb deficiencies were not well-described anatomically, or when the etiologies of their defects were clearly chromosomal or single gene mutations. We also excluded infants from the category of "bilateral" limb deficiencies if laterality was vague or unspecified.

Statistical Analysis

As shown by Khoury and Holtzman [1987], the increase of the prevalence of a birth defect is a function of the frequency of exposure to the teratogen, the strength of the teratogen, and the etiological heterogeneity of the outcome. The formula is given as follows:

$$P_n = p[1 + fh\,(R - 1)], \quad (1)$$

where f is the frequency of exposure, h is the heterogeneity index, and R is the relative risk. If we define $\rho = p_n/p = [1 + fh(R-1)]$, then ρ indicates relative changes

in prevalence rates of p_n over p. We calculated ρ with the combinations of parameters h, f, and R shown in Table I. Regardless of baseline p, ρ reflects the magnitude of changes of p_n in relation to p.

Given the above conditions, to detect a statistically significant increase in p_n over p is to test H_0: ρ = 1 against H_1: ρ > 1. Several statistical methods are used in birth defects monitoring [Chen, 1979]. Among these methods, Poisson and CUSUM techniques are the two most commonly used [Hill et al., 1968; Chen, 1979; Lechat et al., 1985; Lechat, 1989; Lucas, 1985; Khoury and Holtzman, 1987; Khoury and Edmonds, 1994]. We examined both techniques and calculated the power curve for the Poisson method to test ρ = p_n/p = 1 against the one-sided alternative ρ > 1 for α = 0.05. We also simulated the CUSUM average run length (ARL) for an increase in birth prevalence of the birth defects with changes in the expected number of cases.

Poisson method. In birth defects monitoring, we observe a number of births over a period of years (N) and calculate the baseline birth prevalence rate p for a birth defect. The expected number infants with a particular birth defect = Np. p is related to p_n by Eqn. (1). For any given ρ, one can either calculate the sample size needed to test the hypothesis for a given level of α (type I error) or to determine the statistical power (type II error) for any given expected number of cases [Gail, 1974; Yang, 1978; Greenland, 1985; Lubin et al., 1988]. We used the statistical analysis package SABER [James, 1993] to calculate the power curves for testing ρ = 1 against ρ > 1 for α = 0.05.

Figure 1 shows the power to test ρ ≥ 1 for α = 0.05 plotted as a function of the expected number of cases for various values of ρ. It can also determine the sample size needed to detect significant changes of p_n over p. Suppose one wishes to detect a difference between two birth prevalence rates, p_n and p, with the baseline p = 0.001, and ρ ≥ 2. Reference to Figure 1 shows that the expected number of cases (9.2) gives a statistical power (1−β) of 0.8; thus 9,200 (9.2/0.001) is the number of births needed to detect the true increase in birth prevalence. Suppose a birth defect surveillance program registers 40,000 births a year and a baseline rate of a specific birth defect of 0.0005. The expected number of annual cases (Np) is 20. As shown in Figure 1, the statistical power would then be less than 0.8 for all values of ρ less than 1.6. It is clear that the smaller the values of ρ, the less statistical power one has or the longer the period of time that is needed to have enough births to detect a significant change in the rate of a birth defect.

CUSUM technique. Cumulative sum (CUSUM) quality-control schemes are widely used in industry [Lucas, 1982, 1985]. A CUSUM quality-control scheme accumulates the difference between an observed value of Y_i and a reference value k. If this sum equals or exceeds the decision interval value h, an out-of-control signal is given [Lucas, 1985]. Kenett and Pollak [1978] showed that a CUSUM technique was better than non-CUSUM techniques in detecting changes in the birth prevalence rate of congenital malformations. CUSUM is usually evaluated by calculating the average run length (ARL). The ARL is the average number of samples taken before an out-of-control signal is obtained.

In the MACDP, the CUSUM technique is used to compare numbers of babies with defects in each category each month with the numbers counted during the previous 15 months. If the count exceeds a previously chosen decision boundary, an alarm is signaled and the cases are marked for further investigation. We simulated the ARL for different expected numbers of cases and different ρ values. The null ARL is set to be 300 months, which is similar to α < 0.01 in the Poisson method.

Figure 2 gives the ARL for different values of ρ by the number of expected cases. For example, it shows that if the expected number of cases = 10 and ρ = 2, it would take about 2 months for CUSUM to signal an alarm. As with the Poisson method, as the expected number of cases approaches zero and ρ approaches 1, the ARL increases rapidly.

TABLE I. Values of ρ = [1 + $fh(R - 1)$] by Frequency of Exposure, Relative Risk, and Heterogeneity Index

Relative risk (R)	h	Frequency of exposure 0.001	0.005	0.01	0.05	0.10	0.25	0.50	0.75	0.90	1.00
Mild (R = 2)	0.05	1.000	1.000	1.001	1.003	1.005	1.013	1.025	1.038	1.045	1.050
	0.2	1.000	1.001	1.002	1.010	1.020	1.050	1.100	1.150	1.180	1.200
	0.5	1.001	1.003	1.005	1.025	1.050	1.125	1.250	1.375	1.450	1.500
	1	1.001	1.005	1.010	1.050	1.100	1.250	1.500	1.750	1.900	2.000
Moderate (R = 5)	0.05	1.000	1.001	1.002	1.010	1.020	1.050	1.100	1.150	1.180	1.200
	0.2	1.001	1.004	1.008	1.040	1.080	1.200	1.400	1.600	1.720	1.800
	0.5	1.002	1.010	1.020	1.100	1.200	1.500	2.000	2.500	2.800	3.000
	1	1.004	1.020	1.040	1.200	1.400	2.000	3.000	4.000	4.600	5.000
Strong (R = 10)	0.05	1.000	1.002	1.005	1.023	1.045	1.113	1.225	1.338	1.405	1.450
	0.2	1.002	1.009	1.018	1.090	1.180	1.450	1.900	2.350	2.620	2.800
	0.5	1.005	1.023	1.045	1.225	1.450	2.125	3.250	4.375	5.050	5.500
	1	1.009	1.045	1.090	1.450	1.900	3.250	5.500	7.750	9.100	10.000
Potent (R = 100)	0.05	1.005	1.025	1.050	1.248	1.495	2.238	3.475	4.713	5.455	5.950
	0.2	1.020	1.099	1.198	1.990	2.980	5.950	10.900	15.850	18.820	20.800
	0.5	1.050	1.248	1.495	3.475	5.950	13.375	25.750	38.125	45.550	50.500
	1	1.099	1.495	1.990	5.950	10.900	25.750	50.500	75.250	90.100	100.000

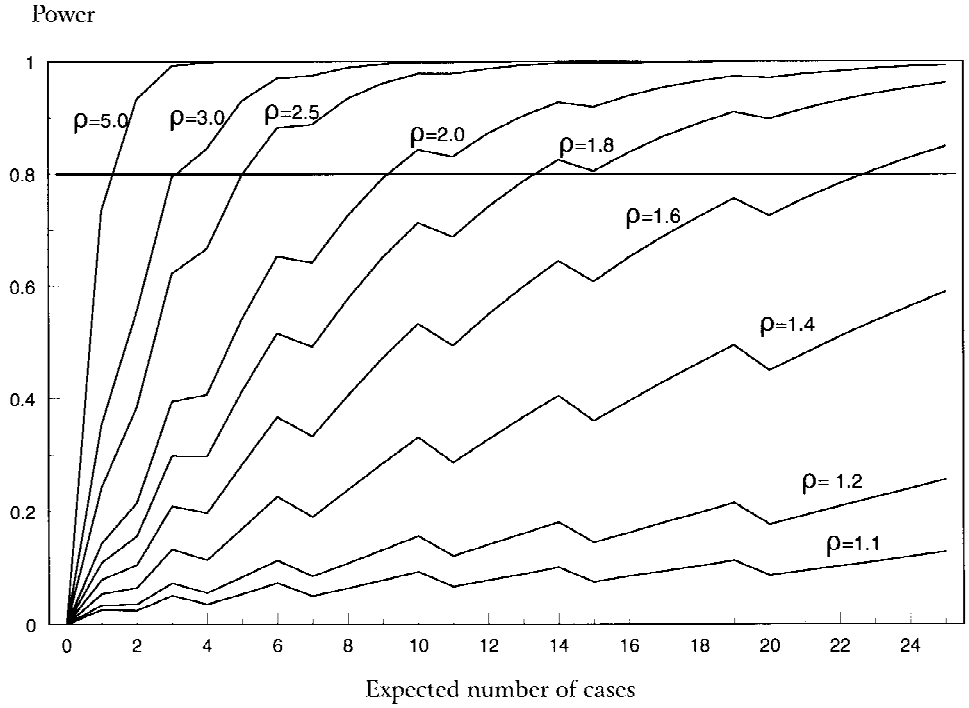

Fig. 1. Power curves by expected number of cases and ρ (Poisson method).

RESULTS

We used data from MACDP to examine rates of TLD, BIPD, and ILD and the ability of the program to detect true changes in the birth prevalence of each for a range of rates of in utero exposure to thalidomide. Table II shows the number of cases and the birth prevalence rates of TLD, BIPD, and ILD for 1968–1993 from MACDP.

The birth prevalence rates of total limb deficiencies showed a slight decline since 1968, from about 0.7/1,000 in the early 1970s to about 0.5/1,000 in the early 1990s. There were a total of 411 cases of limb deficiency from 1968 to 1993. We used cases and births that oc-

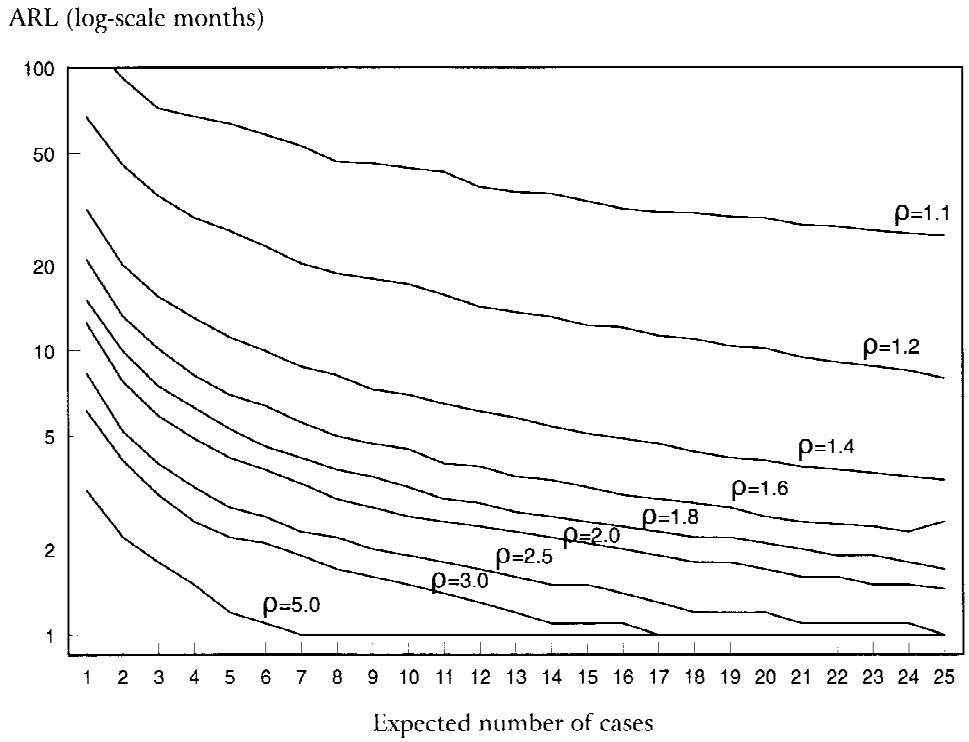

Fig. 2. Simulated CUSUM average run length (ARL) for increases by expected cases and ρ (Null ARL = 300).

TABLE II. Number and Prevalence Rates (Per 1,000 Live Births) of Limb Deficiencies, Bilateral Intercalary or Preaxial Deficiencies, and Intercalary Limb Deficiency, Metropolitan Atlanta Congenital Defects Program (MACDP), 1968–1993

	Limb deficiency		BIPD[a]		%	ILD[b]		%
Year	Cases	Rate	Cases	Rate	BIPD	Cases	Rate	ILD
1968–1972	91	0.66	7	0.05	7.7	5	0.04	5.5
1973–1977	76	0.64	6	0.05	7.9	2	0.02	2.6
1978–1982	66	0.50	4	0.03	6.1	4	0.03	6.1
1983–1987	57	0.37	4	0.03	7.0	4	0.03	7.0
1988–1993	110	0.48	6	0.03	5.5	2	0.01	1.8
Total	411	0.53	27	0.035	6.6	17	0.022	4.1

[a]BIPD represents bilateral intercalary or preaxial deficiencies.
[b]ILD represents intercalary limb deficiency.

curred from 1968–1993 to calculate the estimated baseline of TLD, $P_{limb} = 0.53/1{,}000$. There were 27 cases (including three ILD cases) of BIPD from 1968–1993. BIPD cases accounted for about 6.6% of all limb deficiency cases with an estimated baseline $P_{bipd} = 0.035/1{,}000$. ILD cases (among 17 ILD cases; there were three cases also classified as BIPD) accounted for about 4.1% of all limb deficiency cases with an estimated baseline $P_{ild} = 0.022/1{,}000$. There were a total of 41 BIPD and ILD cases from MACDP for 1968–1993. The estimated relative risk (R) for thalidomide was about 175 for all limb deficiencies [Leck, 1979]. By using MACDP data, we estimated the relative risk (R) for BIPD and ILD to be about 4,570 and 8,180, respectively. We assumed that the heterogeneity index (h) equals 0.10 for total limb deficiencies (41 BIPD and ILD cases divided by 411 TLD cases), $h = 0.8$ for BIPD, and $h = 0.9$ for ILD (Leck, 1979; Lenz, 1985; Castilla et al., 1997). In our model, we varied the frequency of exposure to thalidomide from 1/10,000 to 5/100 among live born infants and demonstrated the ability of MACDP to detect significant changes in incidence rates of TLD, BIPD, and ILD by the Poisson and CUSUM techniques.

Figure 3 shows the estimated sample size needed for various values of f for $\alpha = 0.05$ and $(1 - \beta) = 0.8$ by the Poisson method. The X axis represents the frequency of fetal exposure to thalidomide, ranging from 1/10,000 to 5/100. The Y axis is the sample size needed (number of births needed) in log-scale to detect a significant change in birth prevalence for given values of f, h, and R. The reference line is 40,000 births, the annual number of births that MACDP monitors. If we monitor ILD, MACDP would detect a thalidomide fetal exposure rate of as little as 7/10,000. About 39,000 births would be needed to detect this exposure rate, which equal about 12 months of MACDP monitoring. When monitoring BIPD, it would need 62,700 births to detect a significant change in rate at the same exposure rate (7/10,000). If we were monitoring TLD, an exposure rate of 3.5% or less would go unnoticed by MACDP. At the

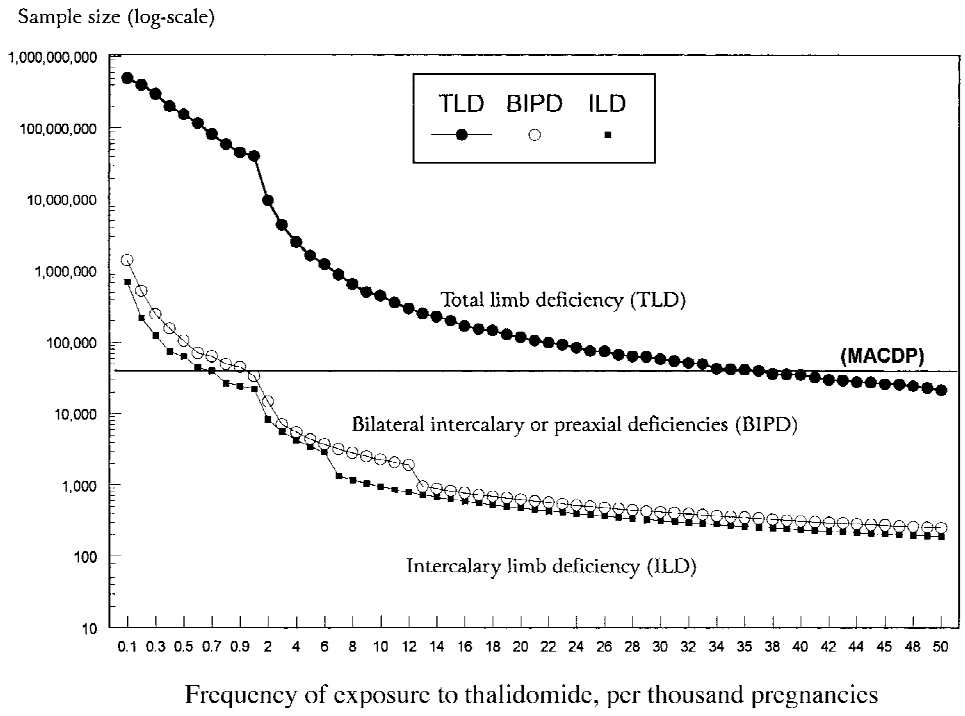

Fig. 3. Sample size needed to detect a significant increase in the birth prevalence of total limb deficiency, bilateral intercalary or preaxial deficiencies, and intercalary limb deficiency for various frequencies of exposure to thalidomide.

level of 1/1,000 exposed, some 40 million births would be needed to detect a significant change in birth prevalence, far beyond the ability of the MACDP or any operating nationwide birth defects monitoring program.

Figure 4 shows the CUSUM ARL (in months) needed to detect the effect of thalidomide on the incidence of TLD, BIPD, and ILD for various frequencies of exposure. The CUSUM results were similar to those by the Poisson method. If ILD is monitored, the ARL would be about 12 months given a 7/10,000 exposure rate. For monitoring BIPD, it would take about 18 months, whereas if TLD was monitored, the ARL would be about 20 years. In the MACDP, the statistical analysis has been done semiannually on cases from the proceeding 6 months. By monitoring ILD, we would be able to detect any significant changes in incidence at exposure rates of 1/1,000 or more in a timely fashion. By monitoring BIPD, we would be able to detect any significant change in incidence at exposure rate 2/1,000 or more, but by monitoring TLD, we would not be able to detect changes in incidence unless the exposure rate reached 5%.

DISCUSSION

Although the therapeutic utility of thalidomide is still under investigation, an evaluation of the sensitivity of birth defects monitoring seems to be necessary. Because thalidomide is a well-known teratogen, if it should become widely available, restrictive measures must be taken to prevent in utero exposure. However, recent experience with isotretinoin, another known teratogen, suggests that no foolproof method exists to prevent completely in utero exposure [Mitchell et al., 1995].

Our study used data collected from MACDP and examined the ability of birth defects monitoring to detect subtle increases in the birth prevalence of limb deficiency assuming a range of rates of in utero exposure to thalidomide. Our results demonstrated that we can detect, within a year of monitoring, in utero thalidomide exposure rates of 7/10,000 if monitoring ILD and 1/1,000 exposure rate if monitoring BIPD in Atlanta. Monitoring for all limb deficiencies will not detect a rate increase quickly. However, as the rates of the exposure less than 1/1,000, a considerably longer time or larger sample are needed to detect a rate change. Our findings suggest that the appropriate classification of thalidomide-induced birth defects as ILD or BIPD rather than unspecified limb defects increases our ability to detect subtle changes. But when the rates of the exposure are sufficiently lower (less than 7/10,000), MACDP will not detect a rate increase in a timely fashion even by monitoring ILD. Our findings also suggest that enlarging the monitored population would play an important role in increasing the sensitivity of our monitoring program. For example, if the monitored population increased from 40,000 to 400,000 births annually, we would be able to detect in utero thalidomide exposure rates as low as 1.5/10,000 if monitoring ILD and 3/10,000 if monitoring BIPD within 6 months of monitoring.

As pointed out by other studies, there is no clear definition of thalidomide-induced limb defects [Calzolari et al., 1990; Smithells and Newman, 1992; Lin et al., 1993; Castilla et al., 1995]. The clinical delineation of thalidomide embryopathy syndrome is difficult and considerable differences in case definition exist [Smithells and Newman, 1992; Castilla et al., 1997]. The term phocomelia is neither sensitive nor specific enough to identify thalidomide-induced limb deficiency and is often ill-defined in epidemiologic studies [Castilla et al., 1997]. In the present study, we used two

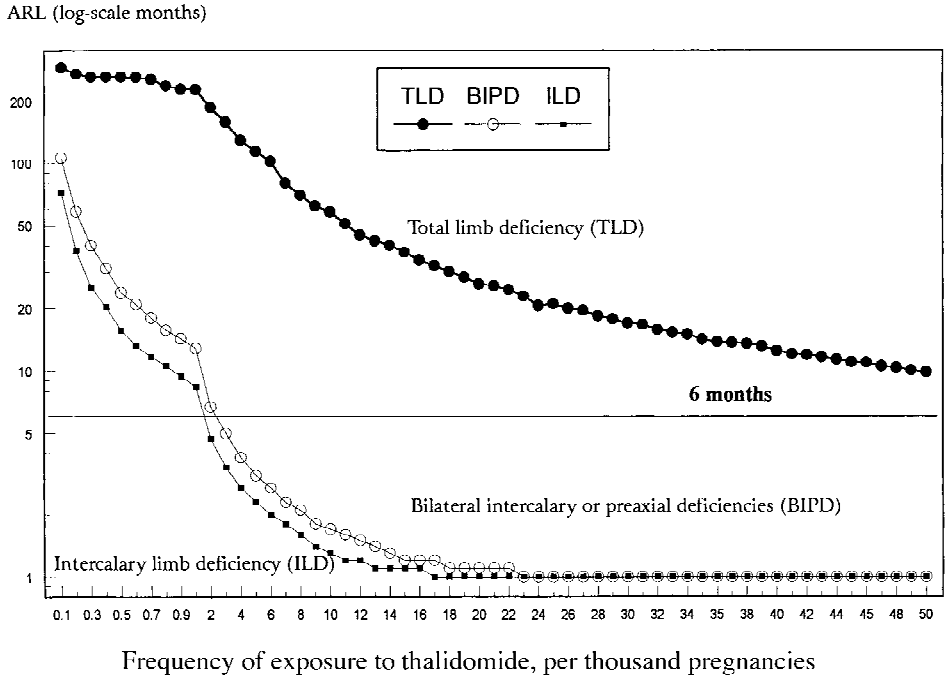

Fig. 4. CUSUM ARL (in months) to detect an increase in the birth prevalence of total limb deficiency, bilateral intercalary or preaxial deficiencies, and intercalary limb deficiency by frequency of exposure to thalidomide.

categories of thalidomide-induced limb deficiency: ILD and BIPD. Our results suggested that by monitoring ILD, one would have more power to detect a change in rate of in utero exposure to thalidomide. For example, for exposure rate less than 1/1,000, we would need on average 50% less of births by monitoring ILD than by monitoring BIPD to detect a significant change in rate. Under the same circumstances for CUSUM, we would need about 35% less of ARL for monitoring ILD than for monitoring BIPD. The differences in timing of monitoring between ILD and BIPD reduces as the exposure rates increases. For example, for exposure rates between 1–3%, we would need 30% less births for monitoring ILD using Poisson method, and about 20% less of ARL using CUSUM. In our study, we assumed that about 90% of thalidomide-induced limb deficiency would be ILD ($h = 0.9$) and about 80% would be BIPD ($h = 0.8$). Therefore, ILD represents a more homogeneous group. In addition, the baseline for BIPD (0.035/1,000) was higher than ILD (0.022/1,000), which indicated that BIPD were more likely to be caused by other genetic or environmental risk factors. These two factors (homogeneity and low baseline rate for ILD) contributed to the more power of monitoring ILD to detect a change of rates of in utero exposure to thalidomide.

There are discussions about which technique to use in birth defects monitoring, how to evaluate the baseline rates, and where to set the level of statistical significance to signal an alarm [Källén et al., 1984a; Chen, 1985]. Furthermore, statistically significant detected changes in rates can be due to factors other than true changes in prevalence rates. These factors include registration artifacts caused by improved case ascertainment (either the result of better clinical identification or better reporting); the incomplete or inaccurate reporting of defects; and delays in reporting defects, processing data, and conducting statistical analysis.

Our analysis emphasized the fact that the changes of birth prevalence of the specific birth defects are a function of the frequency of exposure to the teratogen (f), the relative risk associated with the exposure to the teratogen (R), and the etiologic heterogeneity of a measured defect (h) [Khoury and Holtzman, 1987]. As shown by Figures 1 and 2, as ρ approximates one, which could result from many combinations of f, h, and R, one loses power (or needs longer ARL for CUSUM) to detect significant changes for a given expected number of cases. Most birth defects surveillance systems monitor from 10,000 to 250,000 births annually [International Clearinghouse for Birth Defects Monitoring Systems, 1984; Holtzman and Khoury, 1986; Khoury and Holtzman, 1987]. For any rare birth defects, if the frequency of exposure is sufficiently low or the heterogeneity index is sufficient high, even for a potent teratogen like thalidomide ($R > 175$), the relative risk (ρ) could be close to one (Table I). If this occurs, either larger samples are needed or a longer period of time is needed for a given surveillance system to detect the true increase of birth prevalence rate of a birth defect.

Our methods may serve as a basic tool to evaluate the ability of birth defects monitoring to detect subtle increases in the birth prevalence of birth defects. Our results show that monitoring for BIPD or ILD could detect a resurgence of thalidomide in a surveillance system similar to MACDP only at levels of in utero exposure that would be unacceptably high. They also indicate that monitoring all limb deficiencies would be even more unacceptably inefficient in detecting a rate increase. These findings highlight the importance of a large monitored population, focused surveillance, and appropriate case classification. As suggested by Khoury and Holtzman [1987], the ability of birth defects monitoring to detect human teratogens can be improved by increasing the number of births monitored (for example, by pooling data from several programs as is routinely done by the International Clearinghouse for Birth Defects Monitoring Systems (ICBDMS) [Flynt and Hay, 1979b; Erickson, 1991], European Registration of Congenital Anomalies (EUROCAT) [A EUROCAT Working Group, 1993], or state-based surveillance systems); by monitoring highest-risk populations (i.e., by setting up drug registries for those drugs that are more likely to be teratogens and follow-up of all female patients of childbearing potential to actively monitor for pregnancy during teratogen use and the outcome of such pregnancies); and by classifying birth defects into more homogeneous subgroups.

Patient and physician educational campaigns and public awareness of the teratogenic effects of the thalidomide would no doubt play a crucial role in minimizing the teratogenic impact of thalidomide if it becomes widely available again. Under current regulations of testing drugs for reproductive adverse effects, a tragedy on the scale of thalidomide during the early 1960s seems unlikely. What seems more likely is the introduction either of a weak teratogen or of a potent teratogen like thalidomide with limited use by pregnant women. Either of these situations can result in true subtle changes in the birth prevalence of specific birth defects, changes that may go unnoticed by many birth defects surveillance systems. Therefore, it is important that birth defects surveillance systems focus on high-risk populations and classify birth defects as precisely as possible in order to detect possible subtle epidemics of birth defects.

ACKNOWLEDGMENTS

We are grateful to the Metropolitan Atlanta Congenital Defects Program (MACDP) abstractors, Charlie Mae Peters, Connie Thompson, Debbie Nurmi, Joan Garcia, Joann Donaldson, and Jo Anne Croghan, whose constant data collection efforts provide the foundation upon which MACDP research is built. We thank Dr. Cynthia Moore for her help in reviewing and clarifying the cases of limb deficiencies from MACDP. We thank Michael Atkinson and Yecai Liu for their technical assistance. We also thank two anonymous reviewers for their helpful suggestions.

REFERENCES

A EUROCAT Working Group (1993): EUROCAT Report 5. Surveillance of congenital anomalies, 1980–1990. Brussels: EUROCAT Central Registry, Institute of Hygiene and Epidemiology, Ministry of Public Health and Environment.

Bod M, Czeizel A (1981): Congenital malformation surveillance. Teratology 24:277–283.

Burley D (1986): The decline and fall of thalidomide. In: Scheinberg IH, Walsh JM (eds): "Orphan Diseases and Orphan Drugs." London: Manchester University Press in association with the Fulbright Commission, pp 195–201.

Calzolari ED, Maservigi GP, Garani G, Cocchi C, Magnani, Milan M (1990): Limb reduction defects in Emillia Romagna, Italy: Epidemiologic and genetic study in 173,109 consecutive births. J Med Genet 27:353–357.

Castilla EE, Cavalcanti DP, Dutra MG, Lopes-Camelo JS, Paz JE, Gadow EC (1995): Limb reduction defects in South America. Br J Obstet Gynecol 102:393–400.

Castilla EE, Ashton-Prolla P, Barreda-Mejia E, Brunoni D, Cavalcanti DP, Correa-Neto J, Delgadillo JL, Dutra MG, Felix T, Giraldo A, Juarez N, Lopez-Camelo JS, Nazer J, Orioli IM, Paz JE, Pessoto MA, Pina-Neto JM, Quadrelli R, Rittler M, Rueda S, Saltos M, Sánchez O, Schüler L (1997): Thalidomide, a current teratogen in South America. Teratology 54:273–277.

Chen R (1979): Statistical techniques in birth defects surveillance systems. In: Klingberg MA, Weatherall JAC (eds): "Epidemiologic Methods for Detection of Teratogens." Basel/Munchen: Karger, pp 184–189.

Chen R (1985): Detection and investigation of subtle epidemics. Prog Clin Biol Res 163B:39–44.

Czeizel A (1973): The Hungarian congenital malformation register. Acta Univ Carol Med Monogr (Praha) 56:53–70.

D'Arcy PF, Griffin JP (1994): Thalidomide revisited. Adverse Drug React Toxicol Rev 13:65–76.

Edmonds LD, Layde PM, James LM, Flynt JD, Erickson JD, Oakley GP Jr (1981): Congenital malformations surveillance: Two American systems. Int J Epidemiol 10:247–252.

Erickson JD (1991): International Clearinghouse for Birth Defects Monitoring System: Past, present and future. Int J Risk Safety Med 2:239–248.

Erickson JD (1995): Thalidomide use on the rise. Teratol Soc News 5:5–6.

Flynt JW, Hay S (1979a): International clearinghouse for birth defects monitoring. Contr Epidemiol Biostat 1:44–52.

Flynt JW, Hay S (1979b): International clearinghouse for birth defects monitoring. In: Klingberg MA, Weatherall JAC (eds): "Epidemiologic Methods for Detection of Teratogens." Basel/Munchen: Karger, pp 44–52.

Gail M (1974): Power computations for designing comparative Poisson trials. Biometrics 30:231–237.

Hill GB, Spicer CC, Weatherall JAC (1968): The computer surveillance of congenital malformations. Br Med Bull 24:215–218.

Holtzman NA, Khoury MJ (1986): Monitoring for congenital malformations. Annu Rev Public Health 7:237–266.

International Clearinghouse for Birth Defects Monitoring Systems (1984) Limb reduction defects. In: "1982 Annual Report." Stockholm: Garnisonstryckeriet, pp 8–9.

James LM (1993): "Statistical Analysis Battery for Epidemiologic Research (SABER)." Atlanta: Centers for Disease Control and Prevention.

Jenkinson N (1993): Prescribing disaster. Belfast Telegraph. June 5, 1993, p. 9.

Källén B, Winberg J (1968): A Swedish register of congenital malformations. Pediatrics 41:765–776.

Källén B, Hay S, Klingberg M (1984a): Birth defect monitoring systems. In: Kalter H (eds): "Issues and Reviews in Teratology." New York: Plenum Publishing Corp., pp 1–22.

Källén B, Rahmani TMZ, Winberg J (1984b): Infants with congenital limb reduction registered in the Swedish register of congenital malformations. Teratology 29:73–85.

Kenett R, Pollak M (1983): On sequential detection of a shift in the probability of a rare event. J Am Stat Assoc 78:389–395.

Khoury MJ, Holtzman NA (1987): On the ability of birth defects monitoring to detect new teratogen. Am J Epidemiol 126:136–143.

Khoury MJ, Edmonds LD (1994): Metropolitan Atlanta Congenital Defects Program: Twenty-five years of birth defects surveillance at the Centers for Disease Control. Proceedings of III ASM International Symposium.

Klingberg MA, Papier CM, Hart J (1983): Birth defects monitoring. Am J Ind Med 4:309–328.

Lechat MF, De Wals P, Weatherall JAC (1985): European community's concerted action on congenital anomalies. Prog Clin Biol Res 163B:11–16.

Lechat MF, ed. (1989): "Eurocat Report: Surveillance of Congenital Anomalies, Years 1980–86. Brussels: Department of Epidemiology, Catholic University of Louvain.

Leck I (1979): Teratogenic risks of disease and therapy. Contr Epidemiol Biost 1:23–43.

Lenz W (1985): Thalidomide embryopathy in Germany, 1959–1961. In: Marois M (eds): "Prevention of Physical and Mental Congenital Defects. Part C: Basic and Medical Science, Education, and Future Strategies." New York: Liss, pp 77–84.

Lin S, Marshall EG, Davidson GK, Roth GB, Druschel CM (1993): Evaluation of congenital limb reduction defects in Upstate New York. Teratology 47:127–135.

Lubin JH, Gail M, Ershow AG (1988): Sample size and power for case-control studies when exposures are continuous. Stat Med 7:363–376.

Greenland S (1985): Power, sample size and smallest detectable effect determination for multivariate studies. Stat Med 4:117–127.

Lucas JM (1976): The design and use of cumulative sum quality control schemes. J Qual Technol 8:1–12.

Lucas JM (1982): Combined Shewhart-CUSUM quality control schemes. J Qual Technol 14:51–59.

Lucas JM (1985): Counted data CUSUM's. Technometrics 27:129–144.

Lynberg MC, Edmonds LD (1992): Surveillance of birth defects. In: Halperin W, Baker EL, Monson RR (eds): "Public Health Surveillance." New York: Van Nostrand Reinhold, pp 157–177.

Makonkawkeyoon S, Limson-Pobre RN, Moreira AL, Schauf V, Kaplan G (1993): Thalidomide inhibits the replication of human immunodeficiency virus type 1. Proc Natl Acad Sci USA 90:5974–5978.

Mitchell AA, Van Bennekom CM, Louik C (1995): A pregnancy-prevention program in women of childbearing age receiving isotretinion. N Engl J Med 333:101–106.

Oakley GP (1985): Birth defect epidemiology and surveillance. In: Marois M (eds): Prevention of physical and mental congenital defects. Part A: The scope of the problem. New York: Liss, pp 71–90.

Smithells RW, Newman CGH (1992): Recognition of thalidomide defects (syndrome of the month). J Med Genet 29:716–723.

Yang MCK (1978): Determination of sample size for testing the relation between an incidence and a set of random variables in a sample survey. Biometrika 65:635–640.