2.1 Introduction

Pediatricians are by training the most prevention oriented of all the primary care physicians. Immunizations for various potentially life-threatening infectious diseases and early screening for inborn errors of metabolism are two shining examples of childhood disease prevention. Prior to 1970, no one had attempted to reduce the morbidity, or more importantly, the mortality, of any childhood cancer through preclinical detection, specifically by mass screening for this disease. Over the past 30 years there has been much effort put into better understanding the role of preclinical detection of neuroblastoma, and potentially lowering mortality from this most challenging of childhood solid tumors. This chapter addresses various aspects of screening for neuroblastoma in children.

2.2 The Rationale for Neuroblastoma Screening

Neuroblastoma has an incidence of about 10 per million children 0–14 per year throughout the developed world (Young et al. 1986; Bernstein et al. 1992). In North America neuroblastoma will develop in approximately one in 7000 children before the age of 5 years, and over 700 cases are expected to be diagnosed annually. The incidence of neuroblastoma is about twice that of phenylketonuria, almost tenfold higher than that of galactosemia, and slightly less common than neonatal hypothyroidism (Woods and Tuchman 1987), all diseases which are mandated by most neonatal screening programs throughout the U.S.

Neuroblastoma is a fascinating neoplasm because of several clinical and biologic characteristics. The
tumor is unique biochemically because it possesses metabolic pathways for catecholamine synthesis and metabolism. Homovanillic acid (HVA), the main metabolite of dopamine, and vanillylmandelic acid (VMA), the main metabolite of adrenalin and noradrenalin, are sensitive and convenient markers of neuroblastoma since they are excreted in excessive amounts in a patient’s urine (Hinterberger and Bartholomew 1969). Homovanillic acid and VMA have routinely been measured in patients with neuroblastoma for the past 30 years, and have been found to be invaluable aids in both neuroblastoma diagnosis and follow-up.

The treatment and outcome of neuroblastoma are highly age- and stage dependent. Children who are diagnosed with early-stage localized disease or under 1 year of age, irrespective of stage, can often be treated with limited therapy and have excellent survival (Bernstein et al. 1992; Matthay et al. 1989). In contrast, children over the age of 1 year who present with advanced-stage disease have a very poor survival despite aggressive chemotherapeutic treatment regimens, including bone marrow transplantation (Bernstein et al. 1992; Matthay et al. 1989). While some authors have hypothesized that neuroblastoma presents as at least two discreet clinical pathologic entities (Woods et al. 1992; Brodeur and Nakagawara 1992), others believe that malignant progression is a natural transition from benign-acting neuroblastoma in an infant to advanced-stage disease in a child. We presently know that favorable prognosis is strongly associated with specific tumor cellular characteristics (see Chaps. 4, 5, and 8) But in the 1980s, in the absence of this molecular genetic information, it was hypothesized that this natural transition may be interrupted by early detection to eradicate preclinical neuroblastomas.

2.3 Early Pioneering Studies Investigating Neuroblastoma Screening in Japan

The identification of elevated urinary catecholamines in infants with neuroblastoma was first made in 1957 (Mason et al. 1957). Over the next 15 years, methods for measuring the main urinary metabolic byproducts of dopamine and epinephrine, HVA, and VMA were refined. Twenty-four-hour collections were the rule, and elevated urinary catecholamines became extremely important in aiding in the diagnosis of children with “small round cell tumors” and subsequent follow-up of catecholamine-secreting neuroblastomas (Tuchman et al. 1987). A urinary VMA “spot test” based on the reaction of phenolic acids with diazotized p-nitroaniline became commonplace in pediatric oncology practice (LaBrosse 1968). In the early 1970s, Sawada and colleagues from the Kyoto Prefectural University of Medicine began pilot studies which led to implementing a mass screening program for 6-month-old children in eight cities and prefectures in Japan using the VMA spot test on random urine samples (Sawada et al. 1984). The annual incidence of neuroblastoma in Japan, 8 per million children, was similar to that reported in the U.S. at the time. Originally, 282,000 infants were screened by Sawada et al., representing 50–75% of all births in the areas studied (Sawada et al. 1984). Because of a positive test or logistic problems with the initial sample, almost 11,000 infants (3.8%) were retested. Among 264 infants (1 in 1000) who required clinical evaluation for neuroblastoma at a medical center because they had three consecutive positive urinary tests, 16 cases of neuroblastoma were subsequently identified, giving an incidence by screening of 1 in 17,600 infants. As opposed to the high expected incidence of metastatic disease at diagnosis, 5 patients were found with Evans stage-I tumor, 4 with stage II, 2 with stage III, 5 with stage IV-S, and none with stage IV. The 16 patients were treated with surgery and limited chemotherapy, 15 of whom were alive more than 5 years after diagnosis. The only death occurred 1 month after surgery in a patient with stage-II disease. Of the original screened cohort, an additional 6 children were found to have neuroblastoma 14–29 months after their urinary spot tests gave negative results. Hence, the false-negative rate of the Kyoto screening program was 6 of 22, or 27%, similar to what one would have expected using a VMA spot test (Sawada et al. 1984).

This encouraging early trial was reconfirmed by Sawada in longer-term follow-up of the screened population (Sawada 1986). Subsequently, many other screening trials were initiated in Japan, increasingly
using quantitative assays for measuring both VMA and HVA. In an important trial from Sapporo City, Nishi, Takeda and colleagues demonstrated markedly improved survival in children in that city offered screening compared with neighboring rural areas in Hokkaido Prefecture, in which no screening was available (Nishi et al. 1987). Several childhood cancer experts throughout Europe and North America called for the institution of neuroblastoma screening programs on their continents. Many other Japanese investigators began trials in their own prefectures, and by 1986, screening for neuroblastoma was mandated by law in Japan.

A more careful analysis of the Japanese neuroblastoma screening studies revealed many methodologic limitations (Tuchman et al. 1990). Firstly, there was no utilization of a population-based cohort of infants: studies were generally performed in prefectures which did not have the ability to guarantee ascertainment of all neuroblastoma cases occurring in that region, either detected by screening, or missed and subsequently clinically found. Secondly, the data were all based on survival rather than mortality. To the untrained observer, one would surmise that mortality is the reverse of survival (or “one minus survival”). In fact, mortality represents the number of deaths in a given population, and is not affected by the incidence of a disease in that given population. This difference from “survival” becomes most important in evaluating neuroblastoma screening trials. For example, as the Japanese increasingly used more sensitive and specific quantitative assays for measuring VMA and HVA in their trials, there were increasing data suggesting a rise in neuroblastoma incidence (Yamamoto et al. 2002). As can be seen graphically in Fig. 2.1, if one has an incidence in a disease of 1X, with a survival of 50%, of 100 children, 50% will die. If one artificially raises the incidence to 2X, or 200 individuals in this case, and one maintains the same mortality (50 deaths), there is an artificial increase in the survival to 75% (150 of 200). Hence, looking at survival only, when the actual relevant end point is death rate, can greatly mislead an investigator. In addition, the early Japanese studies utilized no control groups other than historical controls; therefore, potential declines in neuroblastoma mortality could have been attributed to improvements in therapy, rather than preclinical detection. Finally, without the utilization of a population-based cohort trial, several other classic methodologic issues, such as lead time or length bias, could produce falsely optimistic results.

### 2.4 Initial North American and European Neuroblastoma Screening Trials

In the context of the potentially exciting results coming out of Japan mixed with the realities of those studies’ limitations, several groups throughout Europe and North America began early pilot studies looking at the potential effectiveness of neuroblastoma screening. Small exploratory studies were initiated in Quebec (Scrimer et al. 1987), Minnesota (Tuchman et al. 1989), northern England (Craft et al. 1989), Germany (Schilling et al. 1991), France (Mathieu et al. 1996), Austria (Kerbl et al. 1997), and elsewhere (Bergeron et al. 1998). Newcastle hosted the first International Symposium on Neuroblastoma Screening in 1988, where investigators had the opportunity to share logistical challenges and early results. Several important methodologic aspects of
neuroblastoma were revealed. These aspects deserve some comment, given that they represent challenges of any population-based screening approaches for any diseases:

- **Sample collection.** In a series of important studies, Tuchman and colleagues from Minnesota demonstrated that measuring spot urines to determine HVA and VMA were as valid as 24-hour sample collections, thus obviating long collections for children in whom neuroblastoma was suspected clinically (Tuchman et al. 1985). In Japan, urine was squeezed out of diapers into plastic soy sauce bottles which held 5–10 ml of urine. North American and European investigators began collecting urine on diapers blotted against a 10×10 cm piece of filter paper which, when dried, could be mailed by regular mail to the screening laboratory for accurate determination of VMA and HVA, with urinary creatinine as the internal standard.

- **Assays.** Multiple laboratory assays for measuring catecholamines were debated, from the totally qualitative VMA spot test and the semi-quantitative thin layer chromatographic approach, through high performance liquid chromatography (HPLC), ELISA immuno-assays, and ultimately gas chromatography/mass spectroscopy (GC-MS) as the gold standard.

- **Compliance.** No adequate population-based screening trial can be done without a high compliance rate among the participants. Early studies in Minnesota (Tuchman et al. 1989), Texas (Ater et al. 1998), and Austria (Kerbl et al. 1997) found compliance rates of returning filter papers by parents of 6 month olds to be as low as 9 percent, pointing out the need for a massive public health infrastructure to support adequate compliance, even for something as simple as collecting urine from a diaper. Because of such issues, investigators in Minnesota joined forces with those in Quebec, combining clinical trials expertise (Bernstein et al. 1992), an infrastructure already in place for collecting urine in a large majority of 3-week olds, as part of a urinary metabolic screening program for various inborn errors of metabolism (Scrivener et al. 1987); and a rapid GC-MS assay for VMA and HVA determination (Tuchman et al. 1983).

- **Sample sizes.** It became rapidly clear that to adequately study neuroblastoma screening, one might need a trial studying up to a million children or more to get meaningful results (Esteve et al. 1995). This sobering reality led to major modifications in many trials, some of which were abandoned due to the cost, and others that waited years for adequate funding before they proceeded.

- **Case and control ascertainment.** In-place state and country-wide tumor registries collecting incidence and mortality data with greater than 90–95% ascertainment are an important requirement for an adequate prevention study.

---

**Table 2.1.** Principles to be considered for a cancer screening program (Adapted from Prorok and Connor 1986)

1. The disease should be a “common” serious health problem, with substantial morbidity and mortality
2. The target population should be clearly defined and have a reasonable disease prevalence
3. The target population should be accessible, with reasonable compliance to screening expected
4. The screening test should be acceptable in its performance (sensitivity, specificity) and acceptable to those screened
5. Effective treatment should exist for the disease to be detected by screening
6. There should be a reasonable expectation that patients with positive screening will comply with recommended work-up, diagnosis, therapy, and follow-up
7. Sufficient resources should be available to perform the screening
8. Develop policies for early recall of patients testing positive and follow-up of those testing negative
9. Quality control procedures to maintain sensitivity and specificity of the screening test should be in place.
Table 2.1, from Prorok and Connor (1986), lists principles to be considered for a cancer screening program. One could argue that any childhood disease is not a “common” serious health problem with substantial morbidity and mortality, to warrant the expense of a screening program; however, based on past precedent and the fact that children represent the future of the world, should screening of any childhood disease lower mortality, implementation would be seriously considered.

### 2.5 Follow-up Studies from Japan and Europe

Since 1986, when screening for neuroblastoma in Japan was mandated by law, compliance with the Japanese screening program has been much greater than 80% nationwide (Sawada and Takeda 2000). Although highly successful in recruiting parents to participate in this program, such widespread mass screening also led to less ability to measure the efficacy of this approach, for example, by comparing mortality from neuroblastoma in a population offered screening versus that not offered screening; however, subsequent attempts to document screening efficacy were performed in Japan on relatively small populations. Investigators in general found no diminution in the incidence of late-stage disease, an early marker of potential screening success, in the incidence of the disease with unfavorable biologic features, or in mortality (Yamamoto et al. 2002; Bessho et al. 1991; Yamamoto et al. 1995; Kaneko et al. 1990; Suita et al. 1998).

During the 1990s, as noted above, several smaller studies were also performed in Europe, usually without controls, with preliminary results suggesting that screening increased the incidence of the disease (Mathieu et al. 1996; Bergeron et al. 1998). Ultimately, only two prospective population-based controlled trials examining the role of neuroblastoma screening in reducing mortality from this disease were implemented that had adequate funding to guarantee a high screening compliance rate; uniform neuroblastoma evaluation, staging, treatment, and follow-up; and optimum ascertainment procedures for determining incidence and mortality. These were the Quebec Neuroblastoma Screening Project (Woods et al. 1996, 2002) and the German Project on Neuroblastoma Screening (Schilling et al. 1998, 2002). Both of these studies deserve special mention, noting similarities and differences.

### 2.6 Definitive Controlled Trials from Quebec and Germany

#### 2.6.1 Studies, Designs, and Logistics

The greatest strength of both the Quebec and German trials was that they were prospective, population-based controlled studies in which neuroblastoma mortality was the definitive end point, rather than survival (vide supra). Both studies had considered a randomized trial approach, but “randomized controlled trials in population-based intervention studies are not always feasible” (Woods et al. 1999) as pointed out by the Quebec researchers. To clarify, the North American group had to decide what they were actually studying by introducing a new screening procedure in an infant population. Were they going to evaluate the screening test itself (urine sampling of 6-month-old babies by parents at home), or were they going to study the entire public health intervention which included introducing a new screening test? They decided that the latter question was much more relevant to improving scientific knowledge, and that to achieve a reasonable compliance rate multiple population-based education methods would be necessary, as noted below. If these measures led to a “halo effect,” with an increased incidence in the non-screened population, the study results would have been viewed with skepticism. There were also practical matters including the fact that there were no other infant urinary screening programs in place in North America with a high compliance rate. Hence, control populations were picked throughout North America where no public health interventions were performed, and where had such been attempted they would have required millions of dollars in resources to be successful. These areas included the states of Minnesota and Florida, the Greater Delaware Valley, and the Province of Ontario (Woods et al. 1996). German investigators faced a similar problem. They im-
plemented screening in six German states selected on the basis of the “feasibility of implementing the screening program” (Schilling et al. 2002).

The Quebec Neuroblastoma Screening Project was a joint collaboration of 31 investigators throughout North America. The Quebec trial was designed specifically to answer the question of whether screening infants at or before 6 months of age (and the public health interventions associated with it) would lower mortality from this disease. After appropriate sample-size estimates were performed, geared at lowering overall mortality by 40%, it was decided to offer screening to a 5-year birth cohort in the Province beginning 1 May 1989, once NIH funding was secured. The only screening data available around the world at that time was for infants screened at 6 months of age in Japan. Investigators hence decided that they would screen at the same age, to be able to confirm or refute results from Japan. It was furthermore decided that infants would be screened at two ages: once at 3 weeks to take advantage of the urinary screening infrastructure which had been in place for well over 10 years (Scriver et al. 1987), and again at 6 months of age with a new public health intervention. Parents were given a “screening kit” at the birth of their child. The kit included filter paper collection instructions and a bilingual consent form with a “passive” informed consent process specifically explained, approved by an NIH-certified review board. Parents knew that if they did not want to screen their infants for either inborn errors or neuroblastoma, they did not need to return the filter paper. On the other hand, if they wanted their infants screened for the already-in-place program for metabolic abnormalities but not for neuroblastoma, they simply needed to check a box indicating refusal to participate in the “cancer test,” mailing the consent form with the filter paper. Greater than 90% compliance was expected with the 3-week test, as compliance had consistently been above that level for several years for the metabolic screen (Scriver et al. 1987). Because the 6-month screen represented a new public health measure, multiple mechanisms were put in place to achieve compliance of about 75%. Some of these mechanisms included radio/television appearances and public service announcements, newspaper and magazine articles, posters in physicians’ office and health clinics, information given to parents at birth, notices included with the Provincial “subsistence checks” which generally were mailed to all parents of infants, and even reminder inserts in diaper boxes.

Initial analyses of filters from both time periods were done in Sherbrooke utilizing thin-layer chromatography. The assays were geared towards the highest sensitivity and accepted a lower specificity: all positive filters, representing between 5 and 10% of infants screened, were then sent to Minneapolis where definitive, highly specific GC-MS assays were performed on the same sample. If the results were positive, parents were contacted and a second sample was requested, which was again studied by GC-MS. All children with a second positive sample were referred to one of the four Quebec pediatric cancer centers for uniform neuroblastoma evaluation (Table 2.2).

In the German Project on Neuroblastoma Screening, initial pilot studies examined the feasibility of performing a screening study in infants at 6 months of age (Schilling et al. 1991). Subsequently, pilot studies in Japan were instituted looking at screening at a later age; and preliminary data were emerging from Quebec suggesting a greatly increased incidence of the disease by screening at or earlier than 6 months, with no evidence of lowering the incidence of advanced-stage disease (Woods et al. 1996). Investigators worldwide believed that any reduction of mortality from a screening approach would be potentially heralded by a lower incidence of children “destined” to do poorly. Stuttgart and Hamburg researchers hypothesized that if neuroblastoma screening at 6 months of age was not going to lower mortality, perhaps screening at 1 year would be more successful, as well as potentially lower the incidence of disease by not detecting cases which would have spontaneously regressed before that age. After securing funding from the German government, screening was offered to all children at 1 year of age born in six German states, between 1 July 1994 and 31 October 1999.

German investigators hoped to achieve a compliance of over 70% to insure accurate sample-size estimates geared at lowering mortality. Unfortunately,
Table 2.2. Comparison of trials. NA not applicable

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Quebec trial</th>
<th>German trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening birth cohort period</td>
<td>1 May 1989 to 30 April 1994</td>
<td>1 July 1994 to 31 October 1999</td>
</tr>
<tr>
<td>Location</td>
<td>All of Quebec</td>
<td>Six German states</td>
</tr>
<tr>
<td>Number in cohort offered screening</td>
<td>476,654</td>
<td>2,581,188</td>
</tr>
<tr>
<td>Age at screening</td>
<td>3 weeks and 6 months</td>
<td>1 year</td>
</tr>
<tr>
<td>Screening compliance</td>
<td>89% at 3 weeks 73% at 6 months 92% overall</td>
<td>61%</td>
</tr>
<tr>
<td>Concurrent control cohorts¹</td>
<td>I. Rest of Canada</td>
<td>Remaining ten German states</td>
</tr>
<tr>
<td></td>
<td>II. 4 Specific control groups:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ontario</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minnesota</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greater Delaware Valley</td>
<td></td>
</tr>
<tr>
<td>Number in control cohorts</td>
<td>Rest of Canada, 1,509,000</td>
<td>2,117,000</td>
</tr>
<tr>
<td></td>
<td>Specific control groups, 2,718,000</td>
<td></td>
</tr>
<tr>
<td>Ascertainment procedures for screened and control cohorts</td>
<td>Two independent procedures, both complete with high correlation</td>
<td>One collaborative procedure, complete</td>
</tr>
<tr>
<td>Screening assays</td>
<td>Thin layer chromatography → Gas chromatography/mass spectroscopy</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>Number of (+) assays required before referral for neuroblastoma evaluation</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Type of analysis</td>
<td>Entire cohort (8% not screened)</td>
<td>Screened sub-cohort only (excludes 39% not screened)</td>
</tr>
<tr>
<td>Screen (+) requiring neuroblastoma evaluation</td>
<td>82 (1/5300 screened)</td>
<td>1754 (1 of 840 screened)</td>
</tr>
<tr>
<td>False (+)</td>
<td>39</td>
<td>1605</td>
</tr>
<tr>
<td>True (+)</td>
<td>43</td>
<td>149</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>52%</td>
<td>8%</td>
</tr>
<tr>
<td>Missed by screening (never screened)</td>
<td>66 (3) (Excludes 20 patients diagnosed clinically prior to 3 weeks of age)</td>
<td>55 (NA) (Excludes unknown number of cases diagnosed prior to screening)</td>
</tr>
<tr>
<td>Total cases</td>
<td>132</td>
<td>204</td>
</tr>
<tr>
<td>Standardized incidence ratio (SIR) for neuroblastoma, comparing the study to control groups</td>
<td>2.0</td>
<td>1.9</td>
</tr>
<tr>
<td>SIR of advanced stage-3 and stage-4 disease ≥1 year</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Uniform neuroblastoma staging, treatment, and follow-up</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Deaths:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>Screen detected</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Diagnosed prior to 3 weeks of age</td>
<td>3</td>
<td>NA</td>
</tr>
<tr>
<td>Missed by screening</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>Not screened</td>
<td>1</td>
<td>NA</td>
</tr>
</tbody>
</table>
despite major efforts, early compliance in their trial was low, less than 50%, demonstrating how difficult obtaining good compliance is. Although the ultimate compliance rates approached 65%, overall compliance was 61%. The study was approved by a state ethics committee of the German Medical Association. They agreed that parents gave informed consent by mailing in urine-saturated filter papers for testing. The parents of each child in the screening area were offered screening once, at the time of the general checkup when the child was about 1 year of age (Schilling et al. 2002). Urine collected was analyzed for catecholamines by high-performance liquid chromatography. Similar to the Quebec trial, if a child had a positive assay, a second sample was requested. If that sample was positive, parents were contacted and asked to bring their child to a center for neuroblastoma evaluation. The assay was purposely geared to be as sensitive as possible, knowing that such an approach might lead to lower specificity, thus generating a much larger number of false-positive cases than in the North American trial; hence, there were some very interesting and important differences between this and the Quebec trial (Table 2.2).

Compliance in the Quebec trial closely approximated that used to calculate sample size estimates, and further analyses of the Quebec cohort were done using the entire birth cohorts, i.e., children were included whether screened (overall 92%) or not. Because of the lower than expected compliance rate in the German trial, many analyses in their definitive paper were based on results in those individuals only screened (Schilling et al. 2002), as noted in Table 2.2. Almost 2.6 million children were born in the six states during the 5 years of the trial, with 1.5 million actually undergoing screening. On the other hand, both trials successfully utilized concurrent control groups with millions of infants born in those areas. Quebec investigators used two completely independent ascertainment procedures for identifying cases, and more importantly, neuroblastoma deaths. One procedure utilized resources set up by the pediatric oncologists in the various study and control areas noted above, with major input from the North American cooperative groups, the Pediatric Oncology Group and Children’s Cancer Group. Collectively, these groups treated 95% of all young children diagnosed with cancer in North America (Ross et al. 1996). The second ascertainment approach was performed independently by investigators at the Laboratory Center for Disease Control and Statistics Canada, part of Health and Welfare Canada, utilizing the whole of Canada without Quebec as the control. A remarkable congruence was found between the two procedures. In Germany, 10 of its 16 states in whom infants were not offered screening were used as the controls and included populations in the former East Germany. Fortunately, there were excellent childhood cancer registries in both East and West Germany before unification. Investigators were highly confident that these registries would be able to help them ascertain and follow patients (Schilling et al. 2002). Cas-

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Quebec trial</th>
<th>German trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative mortality in study (per 100,000 children)</td>
<td>4.8 (0–8 years)</td>
<td>1.3 (1–5 years)</td>
</tr>
<tr>
<td>Cumulative mortality control groups (per 100,000 children)</td>
<td>3.3–5.3 (0–9 years)</td>
<td>1.2 (1–5 years)</td>
</tr>
<tr>
<td>Standardized mortality Ratio for neuroblastoma comparing study to control groups</td>
<td>1.4</td>
<td>1.1</td>
</tr>
</tbody>
</table>

*a Two separate ascertainment procedures in Quebec  
*b Four specific areas in North America  
*c Versus rest of Canada
es were identified by the German Childhood Cancer Registry, “which receives information from all cases of childhood cancer in Germany, including all neuroblastomas. Follow-up of all cases was conducted in cooperation with the neuroblastoma treatment trial of the German Society Pediatric Hematology–Oncology” (Schilling et al. 2002). That the German ascertainment procedure was near complete was demonstrated by the fact that only two children diagnosed with neuroblastoma in Germany over a 5-year period were lost to follow-up within 5 years.

There were considerable differences in the sample sizes of the two studies. The much smaller Quebec trial was able to perform its study with fewer expected cases and deaths, in large part because it was investigating mortality from birth rather than from 1 year of age, as in the German study: about one-fifth of all neuroblastoma deaths occur in the first year of life. However, their study may have been underpowered (Esteve et al. 1995) had results been less conclusive. The much larger German trial still relied on sample-size estimates that required a 50% reduction in mortality to see “significant benefit for the study population” (Schilling et al. 2002). Finally, both studies nicely utilized uniform neuroblastoma evaluation, treatment, and follow-up in all study cases and in many of the control areas.

2.6.2 Studies’ Results

Despite some substantial and interesting differences between the two trials, overall results were strikingly similar, and hence very revealing vis-à-vis neuroblastoma behavior. Firstly, the Quebec trial nicely confirmed reports from smaller studies in Japan that highly sensitive assays measuring catecholamine metabolites would markedly raise the incidence of neuroblastoma in infants screened at 6 months of age or younger (Woods et al. 1996). The incidence of neuroblastoma almost doubled over controls in the Quebec birth cohort. The potential for neuroblastomas to regress had been documented for over 30 years (d’Angio et al. 1971), but the magnitude of such regressing cases was never appreciated. The Quebec data suggest that in countries in which there is a very strong medical surveillance of infants, neuroblastoma incidence may rise, as previously noted in studies from Denmark (Carlsen 1986). The data also suggest that as newer perinatal technologies become widespread, such as intrauterine ultrasonography, neuroblastoma incidence will also rise. Furthermore, initial studies examining the incidence of neuroblastoma in children born in Quebec during the 5 years immediately after screening was discontinued, 1 May 1994 to 30 April 1999, document that there has been a reduction in cases compared with the screened population, although not to baseline (WGW: personal observation).

Although many investigators may have predicted the marked rise in neuroblastoma incidence in Quebec, the German results vis-à-vis incidence were almost “shocking,” they, too, found a doubling of the neuroblastoma incidence (Table 2.2), all over the age of 1 year. These data strongly suggest that tumors destined to regress may be present and excrete catecholamines for a much longer time than neuroblastoma researchers previously would have hypothesized.

Both studies documented significant neuroblastoma deaths in the study population. In the Quebec trial, there were 22 deaths, with none in the screened detected cases (Woods et al. 2002); however, three infants diagnosed prior to screening at 3 weeks of age, all with extremely high catecholamine levels that would have been detected by screening, and all with stage 4-S disease, died. Two of these infants had classic stage 4-S disease with rapidly expanding liver masses leading to respiratory compromise, despite heroic surgical attempts at relief. The third infant, despite a clinical stage of 4-S, had unfavorable biologic features, including amplified MYCN gene. The patient responded initially to chemotherapy but ultimately relapsed and died; otherwise, only one child in the Quebec population who died was not screened. In the German trial, investigating only those individuals who were screened, there were 17 deaths, 14 in children missed by screening, and 3 who died after preclinical detection. Of these three, “two children died from complications from surgery (one with stage 2-B disease, and the other with stage 3 disease), and one died from complications of chemotherapy (for stage 2-B disease)” (Schilling et
Despite no deaths in the screened detected children in the Quebec trial, in 1 child with stage 2-B neuroblastoma that was detected by screening at 6 months and who was treated with doxorubicin and cyclophosphamide, a secondary leukemia with an abnormality in chromosome 11q23 subsequently developed. That child underwent bone marrow transplantation and is alive but has severe graft-vs-host disease. An additional child whose disease was detected by screening is in a persistent vegetative state as a result of complications of surgery for severe gastrointestinal obstruction and necrosis. The gastrointestinal problems were attributed to adhesions that resulted from the surgery to remove the neuroblastoma 7 years previously (Woods et al. 2002).

As an intermediate end point, both studies examined the incidence of advanced-stage neuroblastoma (INSS 3–4) in children over 1 year of age. Both showed, if anything, an increase in that incidence in the screened groups (Table 2.2). These results suggested that the screening procedure or the public health interventions instituted as part of the screening projects actually raised the incidence of advanced-stage disease in older infants, perhaps through clinical detection of cases that may have spontaneously regressed, the “halo effect.” In the Quebec trial, there was a significant increased incidence in neuroblastoma over 1 year of age (Woods et al. 1996). In the German trial, even the incidence of advanced-stage disease over the age of 2 years was not lowered by the screening procedure (Schilling et al. 2002).

Finally, and most importantly, cumulative mortality in the study populations in both the Quebec and German trials was not reduced compared with appropriate controls. Mortality was higher in the Quebec screened cohort than in the German, but included a 9-year analysis (Woods et al. 2002). Only preliminary mortality results were presented for the German trial, examining cumulative figures between 1 and 5 years of age (Schilling et al. 2002); however, in examining standardized mortality ratios (SMR) of neuroblastoma comparing study versus control groups, neither showed any reduction, with an SMR of 1.4 in Quebec and 1.1 in Germany (Table 2.2). In fact, in the Quebec trial, examining the rate of death due to neuroblastoma compared with the rest of Canada, as compiled by Statistics Canada, the overall SMR for Quebec was 1.39, with 95% confidence intervals of 0.85–2.30. Figure 2.2 displays cumulative deaths in the Quebec population versus the four control populations whose deaths were ascertained by study investigators.
2.7 Biologic, Psychologic, Economic, and Clinical Aspects of Neuroblastoma Screening

2.7.1 Biologic Aspects

Had neuroblastoma screening actually been associated with a reduction in mortality, one should have seen children with unfavorable biology detected preclinically with subsequent good outcomes. In general, this was not the case. Even in early Japanese trials, virtually all children with neuroblastomas detected by screening demonstrated favorable biologic features, including histology, triploid DNA content, and lack of MYCN amplification (Kaneko et al. 1990). In only one international trial, that conducted in Austria, were any substantial number of patients found through screening with unfavorable biology (Kerbl et al. 1997): results from this study are a bit controversial because of various methodologic issues. Preliminary results from the Quebec trial documented that, similar to the Japanese uncontrolled studies, virtually all children detected clinically had favorable biologic features (Brodeur et al. 1998, 2001); however, the vast majority of children who died after being missed by screening had unfavorable biologic features; for example, amplified MYCN oncogene identified in 11 of 19 patients studied (Woods et al. 2002). The German project is expected to publish biologic results in the future.

In summary, the current data overwhelmingly suggest that patients with favorable biology neuroblastoma are able to be successfully detected preclinically; however, those with poor biologic characteristics are missed by screening at 3 weeks, 6 months, and 1 year of age. This suggests that such tumors are either in general not present at these ages, or small enough not to be excreting catecholamines in excess of normal urinary amounts, with subsequent other cellular events leading to a great expansion of the cancer, often with metastatic spread, and clinical detection at an advanced stage.

2.7.2 Psychologic Aspects

Unfortunately, very few studies have examined the potential psychological implications of screening infants for neuroblastoma (Bell et al. 1994). Investigators in the Quebec trial tried unsuccessfully to obtain funding for what they believed to be an important secondary aim of their trial. Austrian investigators fortunately were able to conduct interviews on parents of children who underwent neuroblastoma screening with negative results (Dobrovoljski et al. 2003). They found that a large portion of parents of infants who were referred to cancer centers because of elevated catecholamines and were found not to have neuroblastomas remained very concerned about their children, even years later. Hence, the screening procedure was felt to be very psychologically stressing with long-term consequences. The Quebec screening trial was geared toward a very high specificity, and in the end, less than 1 in 10,000 normal children were evaluated at medical centers for neuroblastoma and found not to have the cancer. The number and percent of such children who falsely tested positive was a log higher in the German study, and remains high in Japan today.

2.7.3 Economic Aspects

Very little has been written on the potential cost-effectiveness of neuroblastoma screening. Because screening has been found to be ineffective, one could argue that such studies would by necessity be negative. The Quebec investigators, however, did prospectively examine cost-effectiveness and the data have yet to be published; however, preliminary results lead to a very important and provocative conclusion: over $8 million USD were spent on the Quebec trial in funds provided through the peer-review grant mechanism of the National Cancer Institute. In addition, significant resources were provided by the Quebec Institute of Genetic Medicine for neuroblastoma screening, including costs associated with setting up the infrastructure for metabolic screening that enabled this study to be done as economically as possible. The German Trial cost more than $20 million USD; hence, at first glance these were highly expen-
sive “negative” studies. But over 4 million children are born in the U.S. every year, compared with 100,000 in Quebec. To put in place an effective infrastructure to screen a large portion of American newborns would have cost easily hundreds of millions of dollars. As importantly, such an infrastructure would have cost tens of millions of dollars to maintain on an annual basis. As noted above, major pediatric voices clamored for institution of neuroblastoma screening in the U.S. before definitive trials proving or disproving its efficacy were performed. Not only did the Quebec and German trials show that neuroblastoma screening was ineffective, but ultimately they saved the American, Canadian, German, and other health care system billions of dollars over a generation. The economic value of well-done research cannot be overestimated, even if results obtained are negative.

2.7.4 Clinical Implications

With the determination that a substantial number of preclinically detected neuroblastomas undergo spontaneous regression, it is highly likely that a substantial amount of favorable-biology neuroblastoma detected clinically would also spontaneously regress; hence, the results of the neuroblastoma screening studies may have practical implications for the care of infants with clinically detected disease. Yamamoto and colleagues have now defined criteria for observing patients with neuroblastomas detected by screening without incurring any untoward risk. The criteria include the identification of small masses on radiographic examinations but no invasion of the intraspinal canal or infiltration around the great vessels; relatively moderate catecholamines secretions; and parental consent (Yamamoto et al. 1998). Their initial results reveal that a substantial proportion of observed tumors regress, and even those infants that need subsequent treatment do well. It is therefore likely that a similar proportion of infants in whom neuroblastoma is detected clinically at less than 6 months of age can also be observed for potential regression of the tumor, rather than undergo major surgery.

2.8 Conclusions

The idea that one could detect childhood cancer preclinically by screening has been and remains an appealing prospect. In well-performed trials in the only childhood cancer in which proper studies could be performed at the end of the twentieth century, neuroblastoma screening for elevated urinary catecholamines led to a marked increase in the incidence of the disease with no reduction in its mortality; hence, in 2004 using the markers studied, neuroblastoma screening has been and should be abandoned throughout the world: in Japan, screening was finally halted in March of 2004 (Tsubono et al. 2004). In the future, there may be better opportunities as more selective markers for poor-biology neuroblastoma are discovered that can be utilized as screening tools. In the meantime, one needs to remember that even collecting urine from a wet diaper may have horrendous long-term consequences, as evidenced by the outcome of some infants screened in the Quebec, German, and Japanese trials. Physicians should always practice the “golden rule” of medicine: primum non nocere.

References


Kerbl R, Urban C, Ladenstein R et al. (1997) Neuroblastoma screening in infants postponed after the sixth month of age: a trial to reduce “overdiagnosis” and to detect cases with unfavourable” biological features. Med Pediatr Oncol 29:1–10

LaBrosse EH (1968) Biochemical diagnosis of neuroblastoma; use of the urine spot test. Proc Am Assoc Cancer Res 9:39


Matthay KK, Sather HN, Seeger RC et al. (1989) Excellent outcome of stage II neuroblastoma is independent of residual disease and radiation therapy. J Clin Oncol 7:236–244


Tuchman M, Robison LL, Maynard RC et al. (1985) Assessment of the diurnal variations in urinary homovanillic acid and vanillylmandelic acid excretion for the diagnosis and follow up of patients with neuroblastoma. Clin Biochem 18:176–179


