Health and Economic Benefits of Well-Designed Evaluations: Some Lessons From Evaluating Neuroblastoma Screening

Lee Soderstrom, William G. Woods, Mark Bernstein, Leslie L. Robison, Mendel Tuchman, Bernard Lemieux

**Background:** Well-designed evaluations of health services are frequently made today. However, the extent of the evaluations’ benefits and costs is not well documented, creating uncertainty whether their use is optimal from society’s perspective. We examined these costs and benefits using data from one well-designed evaluation, the Quebec Neuroblastoma Screening Project (QNSP). It screened most Quebec newborns between 1989 and 1994 for neuroblastoma. As previously reported, the screening did not reduce neuroblastoma mortality and caused adverse health effects. **Methods:** We compared the cost of doing the QNSP with its benefits. Had the QNSP not been undertaken, neuroblastoma screening would have been implemented throughout North America. We assume that screening would have started in 1989 and ended in 2002. The QNSP’s benefits include the health costs and adverse health effects averted by not using ineffective screening during those 14 years. In our calculations we used neuroblastoma incidence data for the QNSP and for Ontario where there was no screening, detailed data describing the health services used by the patients, and Quebec cost data for those services. **Results:** The QNSP cost $8.77 million (2002 US dollars). By not implementing similar screening programs between 1989 and 2002, the United States and Canada avoided $574.1 million in health costs, the unnecessary treatment of 9223 children, and false-positive findings for 5003 children screened. **Conclusions:** The health care costs and adverse health effects averted by the QNSP justify its costs. These results show that well-designed evaluations can yield—at least sometimes—benefits substantially greater than their high costs. This raises an important policy issue: are these evaluations now being under- or over-used? [J Natl Cancer Inst 2005;97:1118–24]

During the past 20 years, it has been considered important to carefully evaluate new services for detecting, diagnosing, and treating cancer and other health problems. The evaluations must be prospective and use either a randomized controlled trial or other research design to isolate the health effects of the new service for the appropriate population. Such well-designed evaluations are typically performed after the results of other types of studies have shown that the services might be clinically effective.

Such evaluations of new health services yield several benefits. Identifying services that are clinically effective can promote their use. Identifying services that are ineffective can avert their use, thereby avoiding both their adverse health effects, if any, and wasteful health spending. In both cases, evaluations may also yield new knowledge about disease causes and processes. However, evaluations also involve burdens, being both costly and time-consuming. Because these evaluations typically require years to complete, doing them can delay access to important new services for people not participating in the evaluations.

It is important to gather information about these benefits and burdens to determine whether, from a social perspective, the current use of well-designed evaluations is optimal—that is, whether too few or too many are being done. If too few, the adverse health effects and wasteful health spending caused by the use of ineffective health services could be substantial. If too many, access to some important services could be unduly delayed and spending on evaluations excessive. Given the high rate of technological innovation in health care today, it is important to know whether the use of these evaluations is appropriate. Unfortunately, as we argue below, little information is currently available about the benefits and burdens of well-designed evaluations.

To provide some information about them, we present here a case study based on our recent evaluation of neuroblastoma screening, the Quebec Neuroblastoma Screening Project (QNSP). As explained below, undertaking this project averted the introduction of widespread neuroblastoma screening in the United States and Canada; doing that was important because the QNSP found that this screening apparently did not reduce neuroblastoma mortality and did cause adverse health effects. To illustrate the costs of doing well-designed evaluations, we report the costs of the QNSP. To illustrate the benefits from doing them, we present estimates of the health care costs and adverse health effects averted by not implementing screening throughout North America. We then compare these costs and benefits. Of course, well-designed evaluations of other services may yield different results as the approach has been applied to other services.
benefits and burdens, so one case study cannot show whether the current use of well-designed evaluations is optimal. However, we argue that our results do show the importance of determining whether their use is appropriate.

METHODS

Quebec Neuroblastoma Screening Project

Neuroblastoma is the most common extracranial tumor in children younger than 5 years. When it is diagnosed before 1 year of age, the prognosis is favorable, but when it is diagnosed among older children, mortality is high (1). By the mid-1980s, there was much interest in population-based screening of newborns to detect tumors early. At that time, results from observational studies in Japan suggested that testing infants’ urine for catecholamine metabolites led to early detection and lower mortality (2,3), and Japan adopted screening as a national policy in 1984 (4). Therefore, some physicians in the United States and Canada were advocating mass screening here (5), but others argued that a well-designed evaluation of neuroblastoma screening was needed first (6,7,8). The latter view prevailed. Decisions about implementing screening were deferred, and the QNSP was undertaken.

The QNSP has been described elsewhere (8,9,10). In brief, parents of all babies born in Quebec between May 1, 1989, and April 30, 1994, were offered neuroblastoma screening for their newborns, which consisted of assaying urine samples collected by the parents 3 weeks and 6 months after birth for homovanillic and vanillylmandelic acid; 92% of newborns were screened at least once (Table 1). Using screening, 82 possible tumors were detected (17.2 per 100,000 births). In addition, using traditional clinical methods, physicians detected another 90 tumors (18.9 per 100,000 births) (10).

Children with tumors that were detected through screening or traditional clinical methods were referred to one of four Quebec hospitals for clinical diagnosis. Children with neuroblastoma were treated with standard therapies that were widely used elsewhere in North America. To isolate the effects of screening, neuroblastoma incidence and mortality rates for the Quebec birth cohort were compared with those for five other 1989–94 birth cohorts from the United States and elsewhere in Canada, where no screening was done but where treatment modalities were the same. One such control group comprised births in Ontario.

As reported elsewhere, neuroblastoma mortality in the QNSP cohort was higher than that observed in Ontario (Table 1) and in one other control group, and it was lower than that in three other groups; none of the differences was statistically significant (11). Moreover, screening had two adverse health effects. First, 39 of the 82 possible tumors detected via screening proved to be false-positive cases at clinical diagnosis (10). Second, many children were treated unnecessarily. Neuroblastoma incidence was 92% higher in Quebec than in Ontario (Table 1), a difference that may reflect the behavior of Quebec physicians as well as of the screening itself (10). That is, the extensive publicity in Quebec concerning neuroblastoma during the QNSP may have caused Quebec physicians to become more conscious of the disease and, consequently, to detect more tumors clinically. However, treating the additional Quebec children who had these tumors did not reduce mortality in the birth cohort. Thus, without screening, these tumors probably would not have been detected; they would have regressed, not been treated, and not caused any deaths. Our estimated incidence rate for such “silent tumors”—the difference between tumors diagnosed in Quebec and in Ontario per 100,000 births—was 13.35 (Table 1).

A German evaluation of neuroblastoma screening at 1 year of age reported findings similar to those of the QNSP (12). Given the results of these two studies, interest in neuroblastoma screening has ceased in North America, and Japan recently abandoned its screening program (4).

The QNSP was a well-designed evaluation. It was prospective and population based. A randomized controlled trial was not used; instead, the QNSP researchers decided to evaluate screening using five diverse birth cohorts not subjected to screening as control groups. This decision reflected the researchers’ belief that clinically detected “silent tumors” might prove important and that it probably would not be possible to measure the extent of them if a randomized controlled trial were used (10,11). Moreover, they believed that the five control groups from the United States and elsewhere in Canada would provide high quality data.

Had the QNSP not been done, newborn neuroblastoma screening would have been implemented in the late 1980s throughout North America. Had that occurred, we assume it would have ended in 2002, when planners would have decided that screening was ineffective, perhaps by noting that neuroblastoma mortality was the same with screening in the 1990s as it was without screening in the early 1980s. The assumption of a 2002 end date is optimistic. Given the time required for birth data and neuroblastoma mortality data to become available, for researchers to analyze them, and for planners to accept the researchers’ results, it is very doubtful that screening would have ended before 2003; it probably would have been well after 2003 before screening was stopped.

Table 1. Quebec Neuroblastoma Screening Project (QNSP); selected results

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>QNSP</th>
<th>Ontario</th>
<th>ONSP rate relative to Ontario rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>QNSP compliance rates, %*</td>
<td>3 wks</td>
<td>89</td>
<td>N/A</td>
</tr>
<tr>
<td>6 mo</td>
<td>73</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Incidence per 100,000 births, n</td>
<td>Total cases diagnosed†</td>
<td>27.91</td>
<td>14.56</td>
</tr>
<tr>
<td>Detected via screening‡</td>
<td>9.02</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Detected clinically§</td>
<td>18.89</td>
<td>14.56</td>
<td>1.30</td>
</tr>
<tr>
<td>Silent tumors¶</td>
<td>13.35</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>False positives</td>
<td>8.18</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Neuroblastoma mortality among children &lt;8 years of age per 100,000 births, ‡</td>
<td>4.78</td>
<td>4.33</td>
<td>1.10</td>
</tr>
</tbody>
</table>

*Overall compliance was 92%; 3% of the newborns were screened at 6 months, but not at 3 weeks (10). N/A = not applicable.
†Cases reported as of February 24, 2000.
‡Cases detected via screening and then confirmed clinically using a prescribed algorithm at one of four designated Quebec hospitals.
§Cases detected by a physician using conventional clinical methods (e.g., during a well-baby checkup) and then confirmed clinically using a prescribed algorithm at one of the four designated hospitals.
¶“Silent tumors” indicates the patients who were confirmed on clinical diagnosis to have a tumor and who were then treated, but whose tumor had two features: 1) if not treated, the tumor would have regressed permanently, so treating it apparently did not reduce mortality, and 2) for reasons explained in the text, the tumor would not have been detected had there been no screening program.
*Relative to Ontario, the standardized mortality ratio for Quebec was 1.11 (95% CI = 0.64 to 1.92) (11).
Had screening started in 1989, the proportion of all newborns screened across North America would probably have been less than the proportion screened in the QNSP, 92% (10). Given the uncertainty about the effectiveness of screening in 1989 and the difficulties involved in organizing a program, some states and provinces might not have implemented screening in that year. Moreover, where screening was available, compliance might have been less than 92%. Thus, we used participation rates of 92%, 50%, and 10% when we calculated the adverse health effects (i.e., silent tumors and false positives) and the health care costs avoided by not implementing screening between 1989 and 2002.

**Total Costs Avoided**

The estimation of QNSP costs and health care costs avoided is summarized here. (see Supplementary Data, available at http://jncicancerspectrum.oxfordjournals.org/jnci/content/vol97/issue15.) Costs indicate the present value of resources used in the health system, whether financed publicly or privately. Costs are expressed in 2002 U.S. dollars, and a 3% discount rate was used to calculate present values. The QNSP obtained written informed consent from the parents of newborns with positive screening tests and the approval of institutional review boards sanctioned by the U.S. National Institutes of Health.

Costs avoided by not implementing screening in 1989 were calculated in three steps. We first calculated health care costs avoided per birth (see below). Next, we calculated the present value of costs avoided for each of the 14 annual birth groups between 1989 and 2002 in the United States and Canada. We multiplied costs avoided per birth by the number of births for each group. Then, we summed these costs for the 14 groups to obtain the total costs avoided. “Net saving” resulting from the QNSP is the difference between total costs avoided and QNSP evaluation costs.

Costs avoided per birth is the difference between the lifetime costs per birth of managing neuroblastoma using screening and the costs of managing it without screening. Lifetime costs would be higher with screening because the incidence of neuroblastoma in the QNSP was higher than in Ontario and in the other control groups (10). Not only were some cases detected via screening in the QNSP, but also—reflecting the silent tumor effect—more cases were detected in it using conventional clinical methods than were detected in Ontario using those same methods. Having more cases detected when screening was used meant higher costs for detection with screening, higher costs to confirm the diagnosis clinically, higher monitoring costs for what turned out to be false positives from screening itself, and higher treatment and follow-up costs for children diagnosed with neuroblastoma.

To calculate these costs with screening, we used detection and incidence rates for the QNSP. These costs without screening were calculated using incidence data for the 1989–1994 birth cohort in Ontario, where there was no screening. The Ontario and Quebec health systems are very similar. This Ontario cohort was also one of the control groups used in the QNSP, and the results obtained using it were the same as those for the other control groups (10, 11).

Costs avoided for screening itself per birth were calculated using detailed QNSP data; research-related costs (e.g., data collection) were excluded. However, after discussions with QNSP screening staff, it was assumed that screening costs per birth would be 23% higher elsewhere in North America than in Quebec. Screening in the QNSP was piggybacked on a well-established genetic screening program for all newborns in Quebec, so screening costs were probably lower in Quebec than what they would be elsewhere, where there is no comparable genetic screening program.

The other costs avoided were calculated using detailed data indicating the specific health services used by three samples of patients (see below). To determine the cost of the hospital services used by the patients, we gathered detailed financial data for the two hospitals where 75% of the QNSP patients were diagnosed and treated. To determine the cost of the physician services used, we used fees paid by Quebec’s public medical care plan. Prices for prescription drugs used at home are based on prices paid by Quebec’s public drug plan (13).

Health services used for clinical detection were determined using information for 50 children whose tumors were detected clinically; the children’s primary care physicians listed for us the services they used before referring the children to a hospital for clinical diagnosis. To calculate clinical detection costs avoided per birth, we multiplied the average detection costs for these 50 cases by the difference between the incidence of clinically detected cases in the QNSP and the incidence of neuroblastoma in Ontario.

Monitoring costs avoided per birth were estimated using the QNSP incidence of false positives from screening. To determine the specific services required for monitoring, we obtained information about the clinic visits and the tests used to monitor 14 randomly selected QNSP false positives.

Services used for diagnosis, treatment, and follow-up included inpatient and outpatient hospital and physician services, prescribed drugs used in hospital and at home, and home care services. We enumerated diagnosis, treatment, and follow-up services used at diagnosis and during the subsequent 48 months by reviewing hospital charts for 42 randomly selected patients with neuroblastoma. Almost all services were provided at the hospitals, although the charts also indicated some services provided elsewhere. Given the data on services used and on costs for each service, we calculated the average cost of diagnosis, treatment, and follow-up services per patient for each of the nine Evans stage–age groups (14). With Evans staging, patients are distinguished by their age at diagnosis (<12 months, ≥12 months) and by the invasiveness of their cancer.

Finally, to estimate the diagnosis, treatment, and follow-up costs avoided per birth, we assumed for each stage–age group that diagnosis, treatment, and follow-up costs per patient were the same, with or without screening. Given our estimates of these costs for each clinical stage, we calculated the total costs for the diagnosis, treatment, and follow-up costs per birth: one with screening, using the QNSP incidence rates for the nine groups, and one without screening, using the Ontario incidence rates for the groups. Diagnosis, treatment, and follow-up costs avoided per birth are the difference between the costs with screening and the costs without it.

Sensitivity analysis was used to assess the importance of possible data problems for our baseline results. First, we recalculated our results making allowance for possible errors in our cost data for particular services. For example, we determined the effect of the hospital and physician costs per case being actually two standard deviations lower than the estimated cost per case used in the baseline calculations. We also calculated the effect of screening costs per birth outside of Quebec being the same as in the QNSP,
not 23% higher as assumed in the baseline calculations. Next, we
determined the effect on our results of the appropriate discount
rate being 0 or 5%, not 3%, and of screening lasting 10 years, not
14 years, if the QNSP had not been done. Finally, we calculated
the effect of having lower participation in the North American
screening programs than in the QNSP. The QNSP participation
rate was 92%; we also considered scenarios with 50% and 10%
participation.

**QNSP Evaluation Costs**

Evaluation costs include research costs, screening costs, and
the costs of the additional clinical services used by the QNSP
birth cohort because of screening. Research costs are the differ-
ence between the total financing received by the QNSP and the
cost of screening and other clinical services used by the birth
cohort. Screening costs were based on QNSP births and our
estimate of screening costs per birth. The costs of added clinical
services were based on QNSP births and on our estimates of the
added costs avoided per birth for clinical detection, monitoring,
diagnosis, treatment, and follow-up.

**Health Effects Avoided**

We calculated the number of false-positive cases avoided in
two steps: First, using the QNSP incidence rate for false posi-
tives, we estimated the number of them for each of the 14 annual
birth groups in Canada and the United States. Then, we summed
these 14 estimates to obtain total false positives avoided. We used
an analogous method to calculate silent tumors avoided.

**RESULTS**

The QNSP evaluation cost $8.77 million (in 2002 US dol-

s), 85% of which was financed by grants from the United
States National Institutes of Health (Table 2); 44% of this cost
was for screening and clinical services, and 56% was for re-
search. By not implementing neuroblastoma screening between
1989 and 2002, health costs of $970,000 per 100,000 births were
avoided; 73% of these costs would have been for screening itself
(Table 3). The costs avoided for particular types of diagnosis,
treatment, and follow-up services used are described in Table 4.
For example, screening would have increased inpatient days by
411 days per 100,000 births, reflecting the increased incidence
of neuroblastoma cases with screening. The average cost per

### Table 2. Quebec Neuroblastoma Screening Project (QNSP): costs of the QNSP evaluation

<table>
<thead>
<tr>
<th>Cost</th>
<th>Amount, 2002 U.S. $</th>
</tr>
</thead>
<tbody>
<tr>
<td>QNSP evaluation funding</td>
<td></td>
</tr>
<tr>
<td>Research grants (public and private)</td>
<td>4262000</td>
</tr>
<tr>
<td>Quebec government (publicly funded services)</td>
<td>942600</td>
</tr>
<tr>
<td>Parents (postage for mailing urine samples)</td>
<td>534500</td>
</tr>
<tr>
<td>Total</td>
<td>8772000</td>
</tr>
<tr>
<td>QNSP evaluation costs</td>
<td></td>
</tr>
<tr>
<td>Added health care costs for QNSP birth cohort</td>
<td></td>
</tr>
<tr>
<td>Screening</td>
<td>3187700</td>
</tr>
<tr>
<td>Added clinical detection costs</td>
<td>3800</td>
</tr>
<tr>
<td>Monitoring of false positives</td>
<td>35100</td>
</tr>
<tr>
<td>Added diagnosis, treatment, and follow-up costs</td>
<td>598700</td>
</tr>
<tr>
<td>Research costs</td>
<td>4944900</td>
</tr>
<tr>
<td>Total</td>
<td>8772000</td>
</tr>
</tbody>
</table>

### Table 3. Added health care costs per 100,000 births with an operating neuroblastoma screening program

<table>
<thead>
<tr>
<th>Type of added cost</th>
<th>Cost per 100,000 births, 2002 U.S. $</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening costs</td>
<td>709400</td>
</tr>
<tr>
<td>Added clinical detection costs</td>
<td>9900</td>
</tr>
<tr>
<td>Costs of monitoring patients with false-positive results</td>
<td>7800</td>
</tr>
<tr>
<td>Added diagnosis, treatment, and follow-up costs</td>
<td>252100</td>
</tr>
<tr>
<td>Total</td>
<td>970200</td>
</tr>
</tbody>
</table>

In total, the United States and Canada avoided $574.1 million
in health costs by not using neuroblastoma screening between
1989 and 2002 (Table 5). The net saving—the difference between
costs avoided and the costs of the QNSP—was $565.4 million,
which is 64.5 times the evaluation costs of $8.77 million.
Furthermore, 5003 false-positive cases and 9223 silent tumors
were avoided.

These baseline results proved robust in our sensitivity analy-
isis. (Supplementary Data, available at http://jncicancerspectrum.
oxfordjournals.org/jnci/content/vol97/issue15.) For example, using
the minimum plausible estimate of health costs and the maxi-
mum plausible cost for the QNSP, net saving was still 49.6 times
the evaluation costs. Thus, even if hospital wages and staffing
were lower elsewhere in North America, it seems implausible
that they could have been so much lower that costs avoided would
have been less than the costs of the QNSP evaluation. Similarly,
if the cost of doing the screening elsewhere in North America
were the same as in Quebec—not 23% higher as assumed in the
baseline calculations—net saving would still be 55.3 times the
QNSP evaluation costs.

### Table 4. Added diagnosis, treatment, and follow-up costs by type of service

<table>
<thead>
<tr>
<th>Services</th>
<th>Added units per 100,000 births</th>
<th>Cost per added service, 2002 U.S. $</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient nursing and hospital costs, inpatient days</td>
<td>411</td>
<td>197</td>
</tr>
<tr>
<td>Outpatient services, outpatient contacts</td>
<td>464</td>
<td>74</td>
</tr>
<tr>
<td>Drugs, doses prescribed†</td>
<td>3214</td>
<td>3</td>
</tr>
<tr>
<td>Diagnostic and therapeutic procedures, if†</td>
<td>493</td>
<td>46</td>
</tr>
<tr>
<td>Radiology examinations</td>
<td>2071</td>
<td>9</td>
</tr>
<tr>
<td>Laboratory tests</td>
<td>49</td>
<td>426</td>
</tr>
<tr>
<td>Surgical procedures</td>
<td>180</td>
<td>33</td>
</tr>
<tr>
<td>Physician, n§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consultations</td>
<td>254</td>
<td>†</td>
</tr>
<tr>
<td>Diagnostic and therapeutic procedures</td>
<td>2713</td>
<td>15</td>
</tr>
</tbody>
</table>

*Units of measurement are indicated for each service.
†Includes services provided to inpatients and outpatients.
‡Refers to various procedures, including: audiology consults, bone marrow
biopsies, aspirations and biopsies, dietitian consults, electrocardiograms,
euroencephalograms, inhalotherapy, occupational therapy, physiotherapy, psychol-
ogy consults, radiation therapy, and transfusions.
§Allowance was also made for the added cost of physicians’ inpatient ward
visits, emergency room visits, and outpatient clinic visits. Given the costing
methods used, it is not possible to separate the costs for these services from those
for consultations.
Furthermore, when a discount rate as high as 5% was used, the ratio of net saving to evaluation costs was still 62.8. The net saving ratio was unaffected by using different exchange rates. If the screening programs elsewhere would have lasted only 10 years—not 14 years as assumed in the baseline calculations—net saving would still be 48.5 times the cost of the QNSP. Finally, if the proportion of newborns screened throughout North America were lower than in the QNSP, the net saving would still be substantial. With 50% participation, the net saving would have been 39.0 times larger than evaluation costs. With only 10% participation, the net savings would have been 6.7 times those costs (false-positive cases = 600 and silent tumors = 1106).

**DISCUSSION**

The investment in the QNSP yielded substantial benefits. This evaluation averted at least 14 years of ineffective screening across North America, thereby preventing important adverse health effects and much wasteful health spending. The spending avoided greatly exceeded the cost of the evaluation; in the baseline case, net saving was 64.5 times evaluation costs. These conclusions hold even if participation in the screening programs implemented was much less than that in the QNSP. Moreover, our results likely underestimate the QNSP’s benefits because, if screening had started in 1989, it probably would have been well after 2003 before the screening’s ineffectiveness was recognized and screening was stopped.

The high ratio of net saving to evaluation costs means that the QNSP was an unusually good investment. This can be shown with the following hypothetical investment opportunity: suppose one could make an $100 investment that would yield $50 annually for the next 14 years. Clearly, such an investment would be very attractive. This investment’s net yield would be $482; this is the difference between the present value of the 14 annual yields discounted at 3% yearly ($582) and the investment cost ($100). However, the ratio of the investment’s net yield to its cost is only 4.82 ($482/$100). This is much lower than 64.5, the ratio of the QNSP’s net savings to its costs. To have a ratio of 64.5, the investment’s annual yield would have to be $563, not $50. Then the $100 investment would provide a net yield of $6450 over 14 years. Very few investment opportunities provide such a large yield.

This case study shows that well-designed evaluations can yield substantial benefits relative to their burdens. Other evaluations, of course, may yield different benefits than the QNSP and involve different burdens. Nevertheless, they could be very good investments even if they do not yield benefits nearly as great, relative to their burdens, as did the QNSP. As the foregoing example shows, if the net saving of the QNSP had been only 4.8 times its costs, not 64.5, the QNSP evaluation would still have been a very attractive investment.

Little is known, however, about the benefits and burdens of well-designed evaluations of other health services. A search of the literature published between 1980 and 2004 using the Medline, Health Star, and Cochrane databases yielded only two studies. Both report the cost-effectiveness of investments in evaluations in which the services evaluated proved clinically effective. Drummond et al. (15) reported in 1992 that the investment in an evaluation of photocoagulation therapy for diabetic retinopathy was cost-effective. That evaluation showed that the therapy was effective, thereby stimulating its subsequent use; Drummond et al. calculated that the value of the benefits resulting from the increased use of the therapy was substantially greater than the evaluation’s costs. Similarly, Detsky (16) argued in 1989 that the investments in seven other evaluations were cost-effective. However, he calculated the expected health benefits and expected evaluation costs at the time the evaluations were being planned, not the actual benefits and costs resulting from the evaluations.

We found no studies of the cost-effectiveness of investments in evaluations in which the services evaluated proved ineffective. The benefits of such evaluations probably vary, in part because the extent of the adverse health effects identified in such studies varies. In some of these studies the researchers report finding evidence of minimal adverse effects (17); in others, the researchers report finding serious adverse effects (18). Furthermore, the extent of the net saving yielded by these evaluations is unknown because, when services prove clinically ineffective, researchers do not calculate the costs avoided by not using them.

The benefits and burdens of a well-designed evaluation depend on which implementation strategy is adopted for the service being evaluated—as well as on whether the service proves to be effective. The benefits of the QNSP that we report here stem from the fact that screening implementation decisions were delayed until after the QNSP was completed—because a “postevaluation” implementation strategy was followed. Alternatively, a “preevaluation” strategy might have been followed; in this case screening would have been implemented before the QNSP evaluation was completed and used until the QNSP results were published. This strategy is widely used; for example, it has been used with prostate cancer screening based on the prostate specific antigen test (19,20).

---

**Table 5.** Quebec neuroblastoma screening project (QNSP): net saving and adverse health effects avoided by not implementing screening before evaluating it

<table>
<thead>
<tr>
<th>Type of saving/health effect</th>
<th>Canada and U.S.</th>
<th>Only U.S.</th>
<th>Only Canada</th>
<th>Only Quebec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health costs avoided</td>
<td>574 136</td>
<td>527 032</td>
<td>47 105</td>
<td>9695</td>
</tr>
<tr>
<td>QNSP evaluation costs</td>
<td>8770</td>
<td>7484</td>
<td>1286</td>
<td>1209</td>
</tr>
<tr>
<td>Net saving</td>
<td>565 366</td>
<td>519 548</td>
<td>45 819</td>
<td>9486</td>
</tr>
<tr>
<td>Net saving/evaluation costs</td>
<td>64.5</td>
<td>69.4</td>
<td>35.6</td>
<td>7.0</td>
</tr>
<tr>
<td>Adverse health effects avoided, n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases of false-positives avoided</td>
<td>5003</td>
<td>4583</td>
<td>420</td>
<td>97</td>
</tr>
<tr>
<td>Cases of silent tumors avoided*</td>
<td>9223</td>
<td>8448</td>
<td>775</td>
<td>179</td>
</tr>
<tr>
<td>Total</td>
<td>14 226</td>
<td>13 031</td>
<td>1195</td>
<td>276</td>
</tr>
</tbody>
</table>

**“Silent tumors” indicates the patients who were confirmed on clinical diagnosis to have a tumor and who were then treated, but whose tumor had two features: 1) if not treated, the tumor would have regressed permanently, so treating it apparently did not reduce mortality, and 2) for reasons explained in the text, the tumor would not have been detected had there been no screening program.**
Had this preevaluation strategy been followed with neuroblastoma screening, the use of screening would have caused some adverse health effects and wasteful spending, so the benefits of the QNSP probably would have been smaller than those reported here. However, there are two reasons for expecting some benefits: First, some health planners might have doubted screening’s effectiveness, so they would have elected to limit the scope of their screening programs until the QNSP results were available. Our calculations using participation rates as low as 10% indicate that in this situation the evaluation could still have yielded substantial benefits relative to its costs. Second, for reasons already discussed, if no evaluation had been done, it probably would have been after 2003 before planners realized that screening was clinically ineffective and stopped it. If so, evaluating screening would have shortened the period in which screening was used. If the evaluation resulted in screening being stopped in 2003 rather than in 2010 or 2015, it would have averted some adverse health effects and wasteful spending.

Our study and that of Drummond et al. show that at least some well-designed evaluations yield substantial benefits relative to their burdens. This raises a question: Is the current use of well-designed evaluations optimal? That is, is current funding for these evaluations appropriate? From a social perspective, current funding is not optimal if doing more evaluations would yield sufficient additional benefits to justify their added burdens. Identifying additional effective services could promote their use. Identifying additional ineffective services could avert their adverse health effects and wasteful health spending.

Although the question whether socially optimal funding is available for well-designed evaluations is important, a full discussion of it is beyond the scope of this article. Complicated issues are involved. One is that funding decisions involve much uncertainty; when decisions are being made regarding total funding for these evaluations, it is uncertain what the benefits and burden of the evaluations eventually funded will prove to be. In addition, as already explained, the benefits and burdens of doing an evaluation depend on the implementation strategy followed for the service being evaluated. Thus, assessing the appropriateness of current funding also requires an assessment of the appropriateness of the current mix of implementation strategies being used.

Given the lack of information about the benefits and burdens of well-designed evaluations, it is not surprising that there has apparently been no assessment of the optimality of the current funding for them. Our search of the literature between 1980 and 2004 yielded no such studies. Various methods for improving evaluation practices have been discussed theoretically (21,22,23), but their utility remains questionable (24).

The principal limitation of this study is that it provides information about the benefits and burdens of only one evaluation. Our QNSP results, which seem quite robust, indicate the benefits of that evaluation were substantially greater than its costs. However, given the current lack of comparable studies of other health service evaluations, there is no way of knowing how typical our results are. Indeed, two features of the QNSP might appear to make it an atypical evaluation. First, it evaluated a service that proved ineffective. But the results of Drummond et al. indicate that evaluations in which services prove effective can also yield substantial benefits relative to their costs. Second, a postevaluation implementation strategy was followed for neuroblastoma screening. But, as we have shown, if a preevaluation implementation strategy had been followed, the QNSP evaluation could still have yielded benefits.

Our results also indicate why concern about the current use of well-designed evaluations is justified. If they are being underused, our QNSP results warn that the adverse health effects and wasteful health spending caused by the use of ineffective health services could be substantial. Moreover, our finding that there is little information available about the benefits and burdens of well-designed evaluations implies that researchers, funding agencies, health planners, and governments have not been in a position to make well-informed judgments about the optimality of the funds available. Consequently, if funding has been appropriate, that is probably an accidental occurrence. The importance of having good information about the appropriateness of funding is underscored by the widely accepted belief that the development of new services will continue to be a major feature of health care in the future.

REFERENCES

(19) The International Prostate Cancer Screen Trials Evaluation Group. Large-scale randomized prostate cancer screening trials: program performances in...


NOTES

Supported by Grant CA 46907 from the U.S. National Institutes of Health, by the Quebec Network of Genetic Medicine, Montreal, and by Grant 2691 from the National Cancer Institute of Canada.

We very much appreciate the comments of Maurice McGregor, M.D., on an earlier version of the manuscript and the research help provided by the many people—staff, clinical investigators, and researchers—involved in the Quebec Neuroblastoma Screening Project, including Tim Byrne, Patricia Campion, Claude Fortin, Robert Giguère, Doreen Lalonde, Louise Renaud, and Martine Therrien. We also thank our two reviewers for their helpful comments. However, we are responsible for any problems in the final version of our report.

Manuscript received January 17, 2005; revised May 9, 2005; accepted June 21, 2005.