nancies, can suppress the growth of human breast cancer implanted in laboratory animals (5, 6). A second hypothesis generated from this laboratory work has also received support from epidemiologic data (7). These latter findings have been interpreted by Hsieh et al. as consistent with known effects of estrogens (8). Although such post hoc explanations of new findings may be correct, we are particularly interested in epidemiologic settings in which competing theories can be pitted against one another. As we noted, the most direct evaluation of a possible association between maternal AFP in pregnancy and subsequent risk of breast cancer would involve direct measurements of AFP from screening programs for pregnant women and linkage with cancer registries (2, 3). An observed protective effect in such a study would be difficult to account for on the basis of estrogens.

More generally, we feel that in investigating the effects of reproduction on breast cancer, a full range of possible mechanisms should be considered. This includes analyses of not only hormonal factors but also substances such as AFP that are released by the fetus into maternal circulation and the immunologic aspects of pregnancy such as the fetal antigen hypothesis (9).

REFERENCES


Herbert I. Jacobson
Department of Obstetrics and Gynecology
Albany Medical College
Albany, NY 12208

W. Douglas Thompson
Department of Applied Medical Sciences
University of Southern Maine
Portland, ME 04103

Dwight T. Janerich
Utah Cancer Center
School of Medicine
University of Utah
Salt Lake City, UT 84132

RE: "DETERMINATION OF LUNG CANCER INCIDENCE IN THE ELDERLY USING MEDICARE CLAIMS DATA"

We read with interest the paper by McBean et al. (1) comparing lung cancer incidence estimates derived from the Surveillance, Epidemiology, and End Results (SEER) program and from Health Care Financing Administration (HCFA) hospitalization data for comparable states and years. The intent of that analysis was to evaluate incidence estimates based on HCFA data and their potential applicability to states that are not covered by the SEER program. As the authors discussed in the paper, a number of limitations exist in both data systems, most of which could lead to underestimation of the number of incident cases. We recommend that the authors apply simple, complementary approaches to assess the extent of undercounting inherent in both databases. These approaches, commonly referred to as capture-recapture (CR) methods, have been widely employed by wildlife and fisheries biologists to adjust population estimates to reflect incomplete case-finding (2). CR methods are used to compare the number of cases uniquely identified by a source with the number of those cases identified by more than one source (source overlap cases). This ratio provides an estimate of the number of potentially missed cases.

There is an increasing body of literature documenting the rationale and methodology of CR in the epidemiologic setting. These methods have been used to augment data from existing registries, surveys, and other record systems to
estimate the occurrence of a number of conditions, including the incidence of congenital rubella (3), myocardial infarction (4), insulin-dependent diabetes mellitus (5), and dog bites (6), as well as the frequency of birth defects (7), the prevalence of acquired immunodeficiency syndrome (8), pertussis hospitalization and mortality rates (9), and numbers of injection drug users (10). While the method is not without limitations, the CR approach can aid researchers in estimating the level of case ascertainment even for apparently exhaustive disease surveys (5,11). Quantification of the degree of underascertainment allows the researcher to correct the estimate, yielding more accurate rates than would otherwise have been produced.

CR technology has been applied to cancer registries by Robles et al. (12), who used three sources of ascertainment: hospital discharge records, pathology reports, and death certificates. They found that the level of ascertainment of lung cancer incidence was quite high; however, despite the use of multiple sources, there were still missing cases. Although CR was used to assess ascertainment in that cancer study, the authors did not correct for underascertainment. To our knowledge, ascertainment-corrected incidence estimates have not been attempted in cancer research, in contrast to other fields of disease research.

McBean et al. have a unique opportunity to begin to employ national database records to assess cancer incidence on a broad, national level. The accuracy of the incidence estimates, however, must be called into question if regional and temporal undercounting is ignored. For assessment of ascertainment and correction for incomplete case-finding, we suggest that the two-sample approach is too limited and would recommend that the monitoring of cancer incidence be accomplished using three or more readily obtainable sources. Many types of routinely collected data can be useful as sources for estimating rates. These could include death certificates and the individual source components of the SEER registries. Log-linear methods are available with which to compare three or more rosters of lung cancer cases, evaluate each source’s underascertainment, and derive ascertainment-corrected incidence rates (13).

Evaluation of and control for undercounting is simple and need not be expensive. We believe that when incidence or prevalence rates are presented, there should always be a formal evaluation of the degree of ascertainment, and the rates should be corrected for undercounting (5). The development of broad, very accurate yet inexpensive national data on cancer incidence is indeed feasible. McBean et al. have taken the first step; we now need to move on to the second: improving the accuracy of rates through formal evaluation and ascertainment adjustment (14).

REFERENCES

Claudia S. Moy
Wilmer Ophthalmological Institute
School of Medicine
The Johns Hopkins University
Baltimore, MD 21205

Daniel J. McCarty
Ingrid M. Libman
Ronald E. LaPorte
RE: “EPIDEMIC POLIOMYELITIS IN THE GAMBIA FOLLOWING THE CONTROL OF POLIOMYELITIS AS AN ENDEMIC DISEASE. II. CLINICAL EFFICACY OF TRIVALENT POLIO VACCINE”

During the recent extended case-control studies of polio vaccine efficacy in The Gambia, only the total viral particle content of various vaccine samples was assayed (1). All live poliovirus vaccine titrations, before and during field usage, should be aimed at quantifying the individual serotype to eliminate the statistical flaws inherent in the assay procedure.

Unfortunately, the assay procedures used on live poliovirus vaccines for determination of their total viral quantum are, at best, directed toward type 1. That is well illustrated during computation of the total viral quantum of theoretical vaccine specimens with varying degrees of increase or decrease in different serotypes (table 1). The best case would be the hypothetical sample 1, with a simultaneous 99 percent reduction in type 2 and type 3 contents from 10^6 to 10^3 TCID_{50} [median tissue culture infective dose] (or plaque-forming units) and from 10^{5.5} to 10^{3.5} TCID_{50}, respectively. That loss would not be evident during the trivalent titration procedure.

Alternatively, the 90 percent reduction in type 1 content from 10^6 to 10^5 TCID_{50} without any simultaneous reduction for types 2 and 3 (sample 2) would be manifest in the test report as a mere 50 percent loss in viral content. The simultaneous 90 percent loss in type 1 and 99 percent loss in types 2 and 3 (specimen 3) would be reported as a 90 percent loss. The 99 percent loss in type 1 quantum, with (sample 4) or without (sample 5) a 99 percent loss in types 2 and 3, would be apparent as a 50 percent or 98 percent loss, respectively.

The total virus titrations fail in precise quantification of any increase in type 2 or 3 quantum. Even a hypothetical increase of 900 percent in the type 2 and 3 content without a type 1 rise (sample 6) would be apparent as a 500 percent rise in total viral quantum.

The ambiguous information on type 2 and 3 viral contents apparent in the hypothetical specimens with a lower or higher trivalent titer (table 1) would have been applicable right through the 1970s during laboratory studies on samples retrieved from the field. During the 1970s, the total viral contents of 191 samples were titrated in primary rhesus monkey kidney cell cultures (2). Apart from 113 samples with titers of 10^{5} TCID_{50} three samples were below 10^{4} TCID_{50}, 11 were between 10^{4} and 10^{5} TCID_{50}, 42 were between 10^{5} and 10^{6} TCID_{50}, and 42 titrated between 10^{5.7} and 10^{5.9} TCID_{50}. The recent report (3) on 122 specimens for which total virus content was quantified is disappointing; 12 samples were below 10^{5.5} TCID_{50}, 16 were between 10^{5.6} and 10^{5.83} TCID_{50}, and 92 were between 10^{5.84} and 10^{6.13} TCID_{50}. Moreover, the trivalent titer exceeded 10^{6.13} TCID_{50} in two specimens (3). The inherent variability of infectivity assays in cell cultures would be eliminated through parallel assays of unknown samples and the reference preparation. The procedure would be repeated at least twice, the mean values calculated

<table>
<thead>
<tr>
<th>Hypothetical sample</th>
<th>Live viral quantum (TCID_{50})* with the following poliovirus type:†</th>
<th>Total reported type 1 trivalent titer for types 1 + 2 + 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10^6</td>
<td>10^3 (99)†</td>
</tr>
<tr>
<td>2</td>
<td>10^6 (90)</td>
<td>10^5</td>
</tr>
<tr>
<td>3</td>
<td>10^6 (90)</td>
<td>10^5 (99)†</td>
</tr>
<tr>
<td>4</td>
<td>10^5 (99)†</td>
<td>10^5</td>
</tr>
<tr>
<td>5</td>
<td>10^6 (99)†</td>
<td>10^5 (99)†</td>
</tr>
<tr>
<td>6</td>
<td>10^6</td>
<td>10^6 (+900)†</td>
</tr>
</tbody>
</table>

* TCID_{50} median tissue culture infective dose.
† During trivalent titration of a standard formulation of vaccine with 10^6 TCID_{50} for type 1, 10^5 TCID_{50} for type 2, and 10^{5.5} TCID_{50} for type 3, the trivalent titer was 10^{5.15} TCID_{50}.
‡ Numbers in parentheses, percentage loss or gain in viral quantum.