NCRA Launches the New Center for Cancer Registry Education!

www.CancerRegistryEducation.org

The Center for Cancer Registry Education (CCRE) provides easy access to high-quality educational programming to support both seasoned professionals and those new to the cancer registry field. The website offers a variety of products and services, allowing you to tailor your training and manage CE credits.

Use your NCRA Website member log-in and password!

Questions? Call 703-299-6640 Ext. 317 or e-mail ccrc@ncra-usa.org

The CCRE Website is partially funded by the Centers for Disease Control and Prevention, Cooperative Agreement U58/5/03342.
Contents

Guest Editors’ Introduction
Increasing the Utility of Population-Based Birth Defects Surveillance Program Data ................................................................. 3
Russell S. Kirby, PhD, MS, FACE; Wendy N. Nembhard, PhD

Original Articles
The Utility of the National Death Index as a Supplemental Data Source in Ascertaining 5-Year Mortality Among Texas Heterotaxy Cases ........................................ 4
Lisa K. Marengo, MS; Adrienne T. Hoyt, MPH; Mark A. Canfield, PhD

Multiple Congenital Anomalies: Issues for Birth Defects Surveillance .............................................................. 7
Jane A. Evans, PhD, FCCMG

Using an Existing Birth Defects Surveillance Program to Enhance Surveillance Data on Stillbirths ......................................................... 13
Wes Duke, MD, MPH; Suzanne M. Gilboa, PhD, MHS

An Informatics-Enabled Approach for Detection of New Tumor Registry Cases ........................................................................ 19
Riyad Naser, MS, MBA, PhD; Judith Roberts; Todd Salter; Jeremy L. Warner, MD, MPH; Mia Levy, MD, PhD

Assessing the Quality of Race/Ethnicity, Tumor, and Breast Cancer Treatment Information in a non-SEER State Registry ........................................... 24
Abigail Silva, PhD, MPH; Garth H. Rauscher, PhD, MPH; Carol Estwing Ferrans, PhD, RN, FAAN; Kent Hoskins, MD; Ruta Rao, MD

Economic Evaluation of Cancer Registration in Europe ........................................................................ 31
R. Zanetti; L. Sacchetto; M. Calvia; A. Bordoni; T. Hakulinen; A. Znaor; H. Møller; S. Siesling; H. Comber; A. Katalinic; S. Rosso; Eurocourse WP3 Working Group

A Primer on Molecular Biology for Certified Tumor Registrars ........................................................................ 38
Shai White-Gilbertson, PhD, MSCR, CTR; Helen Gosnell, CTR; Linda Cope, CTR

Features and Other Journal Departments
Raising the Bar: 5 Steps to Recover From a Mistake ........................................................................ 42
Michele Webb, CTR

Spring 2014 Continuing Education Quiz ........................................................................ 44
Deborah C. Roberson, MSM, CTR; Denise Harrison, BS, CTR

Call for Papers ........................................................................ 46

Information for Authors ........................................................................ 47
Editors
Vicki Nelson, MPH, RHIT, CTR, Editor-in-Chief
Russell S. Kirby, PhD, MS, FACE, Guest Editor
Wendy N. Nembhard, PhD, Guest Editor
Venetta Williams, MPH, CTR, Associate Editor
Reda J. Wilson, MPH, RHIT, CTR, Editor Emeritus
Ginger Yarger, Copy Editor
Janice Ford, Production Editor

Contributing Editors
Faith G. Davis, PhD—Epidemiology
April Fritz, BA, RHIT, CTR—Cancer Registry
Denise Harrison, BS, CTR—CE Credits
Deborah C. Roberson, MSM, CTR—CE Credits
Michele A. Webb, CTR

NCRA 2013-2014 Board of Directors
President: Shirley Jordan Seay, PhD, OCN, CTR
President Elect/Secretary: Therese M. Richardson, RHIA, CTR
Immediate Past President: Sarah Burton, CTR
Treasurer Senior: Dianna L. Wilson CTR, RHIA
Treasurer Junior: Janet Reynolds, CTR
Educational Director: Paulette Zinkann, CTR
Professional Development Director: Deidre M. Watson, CTR
Public Relations Director & Liaison to JRM: Theresa M. Vallerand, CTR, BGS
Recruitment & Retention Director: Linda Corrigan, MHE, RHIT, CTR
ATP Director—East: Pamela Moats, RHIT, CTR
ATP Director—Midwest: Mindy Young, CTR
ATP Director—West: Kendra Hayes, RHIA, CTR

2013-2014 JRM Editorial Advisory Board
Tim Aldrich, PhD
Tammy Berryhill, MA, RHIA
Betty Gentry, BS, CTR
Robert German, PhD
Donna Gress, BS, CTR
Alison Hennig, BS, CTR
Monique Hernandez, PhD
Amy Kahn, MS, CTR
Leah Kiesow, MBA, CTR
Herman Menck, MBA, CDP
Kathleen Thoburn, CTR
Kevin Ward, PhD, MPH, CTR
Pamela Warren, PhD, CTR
Reda Wilson, MPH, RHIT, CTR (Editor Emeritus)
Venetta Williams, MPH, CTR (Associate Editor)

Production and Printing
The Goetz Printing Company

Indexing
The Journal of Registry Management is indexed in the National Library of Medicine’s MEDLINE database. Citations from the articles indexed, the indexing terms (key words), and the English abstract printed in JRM are included and searchable using PubMed.

For your convenience, the Journal of Registry Management is indexed in the 4th issue of each year and on the Web (under “Resources” at http://www.ncra-usa.org/jrm). The 4th issue indexes all articles for that particular year. The Web index is a cumulative index of all JRM articles ever published.
Guest Editors’ Introduction

Increasing the Utility of Population-Based Birth Defects Surveillance Program Data

Russell S. Kirby, PhD, MS, FACE; Wendy N. Nembhard, PhD

This special issue of the *Journal of Registry Management* (JRM) continues collaboration with the National Birth Defects Prevention Network (NBDPN) to promote research focused on improving and enhancing birth defects surveillance methods and use of the resulting data. The NBDPN is committed to the primary prevention of birth defects and to the improvement of outcomes for children and families living with birth defects through the use of birth defects surveillance data for research, program planning, and program evaluation. Members of NBDPN include staff from population-based birth defects surveillance programs across the United States, as well as clinicians, public health professionals, and researchers involved with birth defects epidemiology, primary and secondary prevention activities, program planning, and evaluation.

The articles included in this Spring 2014 issue of JRM were selected from those submitted in response to a call for manuscripts posted on the NBDPN website (http://www.nbdpn.org) and distributed to all state birth defects surveillance programs, NBDPN members, and the birth defects surveillance list serv. The papers included in this issue underwent both editorial and formal peer review. It is our hope that the methods and findings from these papers will contribute to the continual improvement of the science and practice of birth defects surveillance in the United States and around the world.

The contribution of the National Death Index in studies of child mortality is explored in the first paper, focusing on population-based surveillance of heterotaxy. The second paper addresses the challenges posed by multiple congenital anomalies in birth defects surveillance, in terms of classification and categorization, data management, and analysis. The third paper explores the use of a birth defects registry to improve surveillance of stillbirth, an important yet neglected pregnancy outcome.

Many dedicated individuals contributed their time and effort to assist with the publication of this issue. These include the authors of all the submitted manuscripts and the following peer reviewers: Tanya Bedard, Mark Canfield, Glenn Copeland, Adolfo Correa, Krista Crider, Suzanne Gilboa, Peter Langlois, Lisa Marengo, Leslie O’Leary, Lisa Miller, Russ Rickard, Carol Stanton, Jean Paul Tanner, and Ying Wang. We also thank the members of the NBDPN Publications and Communications Committee, and Vicki Nelson for her assistance with the submission and publication of these manuscripts. We also recognize Cara Mai for assisting in coordinating this special issue and her tremendous support of birth defects surveillance at the state and national levels. We also thank the Division of Birth Defects and Developmental Disabilities at the National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, for its support of the NBDPN.

*Department of Community and Family Health, College of Public Health, University of South Florida, Tampa, Florida. †Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, Florida.*
Abstract: Introduction: The Texas Birth Defects Registry (TBDR) is an active surveillance system which covers all pregnancy outcomes and routinely links birth defects cases to in-state vital records. This study describes the value of using the National Death Index (NDI) data to supplement Texas state death certificates from vital records for a birth defects survival analysis. Methods: The cohort for this study were live-born cases with heterotaxy, a complex birth defect, delivered to Texas residents between 1999 and 2006, with a 5-year follow-up period for survival determination. Cases were linked to their Texas birth and death certificates, if present. Any live-born case that did not link to a Texas death certificate was sent to the NDI to search for any deaths that occurred. Results: We identified 366 heterotaxy cases that were live-born in delivery years 1999–2006, 134 of which were linked to a Texas death certificate. The 232 remaining cases were sent to the NDI to search for a death certificate not found previously. This resulted in only 2 additional out-of-state deaths. Discussion: Future quantification of NDI yields for birth defects survival studies would assist with further assessing the efficacy of utilizing the NDI for capturing early childhood mortality in states that routinely link to in-state death certificates.

Key words: birth defects, child mortality, heterotaxy syndrome, national death index, survival

Introduction

Birth defects are one of the leading causes of childhood mortality. The state of Texas is the largest active surveillance birth defects registry in the United States, with a target population of 2,993,604 births reported from 1999 to 2006. The Texas Birth Defects Registry (TBDR) covers all pregnancy outcomes and routinely links birth defects cases to in-state vital records. Heterotaxy syndrome is a rare and little studied constellation of defects which involves complex congenital heart disease and abnormal lateralization of thoracic and abdominal organs, which often require surgical intervention with subsequent high mortality. Findings from a recent Texas study revealed that approximately 30% of all Texas mothers move between conception and delivery. With a great deal of mobility within the US population, it is unknown how many individuals in this cohort moved out of Texas and whose deaths would be underascertained without using National Death Index (NDI) data. Furthermore, little evidence in the literature was found related to the underascertainment of out-of-state deaths in childhood mortality studies. To ensure that all Texas heterotaxy deaths were accurately ascertained in preparation for an upcoming survival analyses, we sent all cases without a link to a Texas death certificate to the NDI for additional follow-up. This study describes the utility of using NDI data to supplement Texas state death certificates for a birth defects survival analysis through the first 5 years of life.

Methods

The primary source of data used in this study was the TBDR and linked Texas vital records. The Registry uses an active casefinding system for structural birth defects and chromosomal anomalies. Trained program staff routinely visit medical facilities to review logs, hospital discharge lists, and other records to identify potential cases of birth defects. If a fetus of any gestational age or an infant up through the first year of life has any of the birth defects covered by the registry, and the mother resides in Texas at delivery, the medical record is abstracted. Using the state vital records system, registry records are routinely linked to birth, fetal death, and death certificates to provide or supplement sociodemographic data and vital status. It is pertinent to note that states do share death information with the state of birth for fraud control purposes. Unfortunately, this notification of death by another state can be delayed as long as a year after the date of death. For the purposes of this analysis, any Texas resident case with a death certificate was counted as having died regardless of the state where the death occurred.

The cohort of interest included live-born infants with a delivery date from 1999 through 2006, representing the first 8 years of statewide coverage. Potential cases of heterotaxy were clinically reviewed using a case definition based on the International Nomenclature Committee for Pediatric and Congenital Heart Disease. Live-born cases not linked to a Texas death certificate were compiled and sent to the NDI for a death search. The criteria for data submitted to the NDI...
is limited to precisely 100 characters with a minimal search criteria of at least 1 of the following 3 combinations plus 1 optional field:

(1) First and last name plus a Social Security number  
(2) First and last name plus month and year of birth  
(3) Social Security number, sex, and full birth date

We did not have access to the birth defect case’s Social Security number, which limited us to the second NDI submission criterion. In addition to the required fields of first and last name plus month and year of birth, we also provided the child’s middle initial, day of birth, race, sex, father’s surname, state of birth, state of residence, and 1 optional field. The optional field we supplied was mother’s maiden name.

Results

Out of nearly 3 million births during the 8-year period of interest, 366 live-born heterotaxy cases were identified out of 2,993,604 live births. Using routine linkage of registry records to Texas state vital records, we found 134 cases who died in Texas, 129 who died before 5 years of age, and 5 who died after age 5. The remaining 232 records were sent to the NDI for additional linkage. From the 232 cases sent, 74 of 232 (31.9%) cases yielded a number of NDI potential matches. As many as 17 potential matches per individual were provided by the NDI’s probabilistic matching for review. The vast majority of the potential matches were individuals with similar names, but who were clearly adults at the time of death. The NDI probabilistic year matches did not bracket the age group of interest and ranged at times to individuals who were greater than 99 years of age, thus completely out of scope for this cohort.

After review, only 2 records were found to be a true match in our cohort; both were deaths that occurred in another state. The use of the NDI yielded only 2 additional out-of-state deaths, both before 5 years of age, to the 129 of 131 (98.5%) deaths identified in Texas.

Almost half of the deaths (57 of 131, 43.5%) occurred in the neonatal period. Furthermore, the majority of the death certificates (129 of 131, 98.5%) indicated Texas as the primary state of residence at the time of death.

Discussion

Infants and children do not have as many fixed identifiers as adults. The lack of the child’s Social Security number meant that the only fields we could make available to the NDI were the child’s name, birth date, race, sex, father’s surname, state of birth, state of residence, and mother’s maiden name. Unfortunately, 2 fields that are useful for adults (Social Security number and marital status) were null for our NDI search. This lack of additional information may have limited the number of deaths successfully identified to only 2 cases.

One other recent study reported that only 0.7% additional deaths through 10 years of age were identified by supplementing a subset of their cohort with NDI search data in addition to state death data. It was unclear how many deaths occurred after 5 years of age, so we could not compare findings in this particular analysis to our own.

Our limited number of deaths found by using the NDI may have been due to the large number of infants who died in the neonatal period (~50%). This relatively short time frame may have reduced the likelihood that a family would have relocated to another state.

Further, as Texas is one of the many states that actively use birth-to-death matching for the purposes of fraud control, any out-of-state deaths should have been reported directly to Texas if the birth occurred within the state of Texas. This sharing of death information may have reduced the need for using the NDI for finding additional deaths. In this cohort, only 2 of 131 death certificates indicated death in another state, while 19 of 131 death certificates had a null state of death occurrence with a residency at death as being in Texas.

We found only 2 of 131 (1.5%) new deaths before 5 years of age by utilizing the NDI, with a total cost of $795.24, compared to a successful identification of 129 of 131 (98.5%) of deaths using routine in-state vital records linkages. This may be in part due to a lower extent of relocation among families of children with birth defects in Texas, compared to other states, or may be due to the practice of birth to death matching for fraud control.

Strengths and Limitations

A significant strength of this analysis is that Texas permits the linkage of birth defects cases to their corresponding vital records. Another study strength is that Texas participates in birth-to-death linkages for fraud control, which permits a exchange of out-of-state death information for birth defect cases delivered in Texas.

A major limitation is the small size of the cohort, which may not extrapolate to other birth defect survival studies.

Conclusions

We identified 129 of 131 (98.5%) of all infant deaths in Texas before 5 years of age through the Texas vital records linkages. With only 2 additional deaths identified through the NDI, it proved to be an inefficient resource for collecting additional childhood mortality data, particularly when child’s Social Security number was not available. Future quantification of NDI yields for birth defects survival studies would help to further assess the efficacy of utilizing the NDI for early childhood mortality in states that routinely link to in-state death certificates. The use of NDI data might be more useful for states with high out-of-state mobility or relocation.

Acknowledgements

We gratefully acknowledge Dr. Keila Lopez for the clinical review that established this cohort, as well as the Texas Birth Defects Epidemiology and Surveillance field staff for reviewing, identifying and abstracting potential cases of birth defects. We also thank the medical facilities for giving us access to their hospital discharge lists and log books, and the Vital Statistics Unit for providing us with birth, fetal death and death certificate vital records to supplement sociodemographic data. We thank the Title V Office at the Texas Department of State Health Services, which provides partial support of the Texas Birth Defects Registry.
References
4. Texas Department of State Health Services Vital Statistics Unit. Local Registrar Handbook. Austin, TX: Texas Department of State Health Services; 2010.
Multiple Congenital Anomalies: Issues for Birth Defects Surveillance

Jane A. Evans, PhD, FCCMG

Abstract: Approximately 1 in 200 individuals and 20% to 30% of those in whom any major structural malformation is found will have 2 or more serious and potentially unrelated birth defects. In addition to the challenges that multiple malformations create for affected persons, their families, and the health care system, appropriate surveillance of such complex patterns can be a concern for birth defects registries. This paper provides examples of how monitoring of multiple anomalies can be beneficial from a clinical and public health perspectives; presents a staged approach to documentation of such defects, including suggestions for their coding; describes the types of patterns in which they occur; and discusses some of the unique issues that arise with respect to statistical analysis of multiple versus isolated birth defects.

Key words: birth defects, laterality, multiple congenital anomalies, sequences, syndromes, polytopic field defects

Introduction

About 20% to 30% of individuals in whom a major structural congenital anomaly or birth defect (one with surgical or cosmetic consequences) is identified will have 1 or more additional serious and apparently unrelated malformations and be considered a case of multiple congenital anomalies (MCA). This is due both to the fact that many underlying causes of birth defects (such as a chromosomal abnormality, a single gene defect, or a teratogenic exposure) have widespread or pleiotropic effects on development and to the general tendency for nonspecific clustering of malformations.

Although it does require additional analytical, epidemiological, and clinical resources, monitoring of MCA can be very advantageous from both a public health and a clinical perspective. For example, it allows identification of changes in the proportion of isolated versus multiple cases after initiation of preventative measures (eg, food fortification with folic acid), improved awareness of phenotypic variation in syndromes on a population basis, accurate assessment of the proportion of cases of a specific malformation due to a particular cause and, as teratogenic exposures rarely cause isolated major defects, better evaluation of potential environmental toxins or new drugs.

There are 2 primary and complementary approaches that can be taken to investigate MCA. One involves epidemiological evaluation to identify nonrandom statistical associations of anomalies in populations, such as investigation of the interrelationship of schisis type defects or the anomalies associated both positively and negatively with oral clefts. The other relates to morphogenesis, using the phenotypic patterns observed with, for example, limb deficiency defects or gastrochisis, to help elucidate cause and pathogenetic mechanisms. The latter approach in particular raises 2 key challenges for birth defects surveillance: documentation of the extent to which a particular anomaly is isolated or part of a pattern of multiple malformations and, for multiply malformed cases, awareness of whether the defects can be attributed to a well-established pattern (and, in some cases, a specific cause) or relate to an unknown or not previously recognized grouping of anomalies. A staged approach is optimal, documenting first the frequency of individual defects; then their distribution within specific patterns (global coding), requiring documentation of the pattern identified (eg, Holt-Oram syndrome, VACTERL association, Pierre-Robin sequence, unknown pattern of MCA); and finally, whenever feasible, clinical evaluation of the records of cases with true MCA, especially those whose cause is unknown.

Documentation of Individual Defects

Even when anomalies in a multiply malformed individual can be attributable to a specific cause or recognized pattern of MCA, it is appropriate for surveillance purposes to code all anomalies identified, especially when each defect or pattern has an independent International Classification of Diseases (ICD) code. For example, an individual with Down syndrome due to trisomy 21 (ICD-9 758.0; ICD-10 Q90.0) might also have atrioventricular (AV) canal (745.69; Q21.2) and esophageal atresia (750.3; Q39.0). Failure to code the heart and foregut malformations would lead to underestimation of the frequency of these defects in the population at large and preclude more in-depth and clinically relevant reporting of anomalies such as the proportion of AV canal defects attributable to trisomy 21 or the number of infants potentially requiring surgery for esophageal atresia.

In addition, as the range of findings in even well-recognized patterns of MCA can be very variable; only a complete record of all major defects can allow detailed
Documentation of Multiple Defects: Patterns of Anomalies

In addition to coding the individual defects in multiply malformed cases, it is clearly valuable to code a recognized pattern when this has been reliably documented. This is most straightforward when a syndrome has been diagnosed. A syndrome is 1 of 4 types of patterns of MCA generally accepted by dysmorphologists, the others being sequences, polytopic field defects, and associations (see later).

Syndrome

A syndrome is defined as a pattern of MCA that has a known cause. This would include chromosomal disorders, single gene defects and teratogenic effects such as maternal diabetes, congenital rubella or fetal alcohol embryopathy. As noted above, all defects in a syndromic case should be coded, though there is debate about minor anomalies known to be common in a particular condition. As most syndromes show variable phenotypes, awareness of all anomalies allows the relative frequency of syndromic versus nonsyndromic cases to be analyzed, may expand the phenotypic spectrum associated with a particular condition and also permits recognition of rare and potentially unrelated anomalies. For example, cleft lip is common in trisomy 13, but much rarer in trisomy 18; observation of sufficient trisomy 18 cases does, however, indicate that cleft lip can be part of the phenotypic spectrum. Postaxial polydactyly is common in trisomy 13 and in Smith-Lemli-Opitz syndrome, an autosomal recessive disorder; it is, however, not seen with trisomy 21 at frequencies higher than in the general population. Its occurrence in a case with Down syndrome would therefore be considered to be an additional anomaly unrelated to the chromosomal imbalance.

ICD codes do allow the coding of some syndromes by name. In ICD-9, there are codes for some conditions such as Ehlers-Danlos syndrome (756.83), Marfan syndrome (759.82) and velo-cardio-facial syndrome (deletion 22q11.2) (758.32). ICD-10 Q codes contain more syndrome related codes. For example, Q86 codes relate to certain exogenous causes such as fetal alcohol effects, while Q90-99 relate to chromosomal anomalies. Many Q87 codes refer to conditions with anomalies affecting multiple body systems. These are further subdivided by common phenotypic findings such as short stature and limb anomalies. Q87.89 is a catchall code for conditions not captured elsewhere. Despite the terms used in both ICD-9 and ICD-10, not all of the conditions labeled as syndromes have specific known causes; eg, prune belly (756.71), Klippel-Feil (756.16), sirenomelia, and VATER (both Q87.2).

Certainty is important with respect to a syndromic diagnosis or other patterns of MCA. In many registries, case records include a field that documents the certainty of diagnosis, differentiating those with a definitive diagnosis from those in which the diagnosis is probable or possible. The International Clearinghouse for Birth Defects Surveillance and Research (ICBDSR) recommends that questionable cases or those where a diagnosis is suspected but not confirmed by diagnostic testing should not be given a syndrome code unless it has been made by “an expert

phenotypic analysis. In VACTERL/VATER association, for example, it is well known that vertebral, anal, cardiac, tracheo-esophageal, renal, and radial ray (limb) defects are frequent, but rarely will an affected individual have anomalies in all these systems. Variation is such that 2 cases, 1 with a duplicated thumb, tracheo-esophageal fistula, and a missing rib and the other with left renal agenesis, anal atresia, ventriculo-septal defect, and a lumbar hemivertebra could both legitimately be considered VATER/VACTERL patients though they have no anomalies in common. A third case with triphalangeal thumb, anal atresia, and left multicystic kidney might also be mistakenly labeled VACTERL association until ear anomalies and deafness, unusual findings in VACTERL (itself a diagnosis of exclusion), lead to molecular testing for a SALL1 mutation and a correct diagnosis of Townes-Brocks syndrome.

From a public health perspective, documenting the frequency of a specific MCA pattern by its label may have occasional benefit; for example, documenting changes in the birth prevalence of Down syndrome with variation in age-specific fertility or the introduction of an enhanced prenatal screening modality. However, in most MCA cases, it is awareness of both the individual anomalies and the patterns in which they occur that will be of benefit in elucidating cause, monitoring the impact of environmental factors, planning for delivery of services such as surgical interventions or special needs educational programs, and in deriving valid information for counseling individuals and families as to prognosis and potential complications. Mortality or severe morbidity in an infant with trisomy 21, for example, is not due to his or her having an extra chromosome per se, but to malformations such as heart defects that are a developmental consequence of aneuploidy.

Thus, according to the National Birth Defects Prevention Network (NBDPN), surveillance normally requires coding of “all individual defects associated with a chromosomal “anomaly, syndrome or association.”10 The International Clearinghouse for Birth Defects Surveillance and Research (ICBDSR) indicates (admittedly not too clearly) that “individual defects within a recognizable syndrome should be coded all, irrespective to: severity: major/severe enough to request a medical intervention or minor/mild number of defects; commonly part and characteristic of the recognized syndrome or not”.11 EUROCAT states: “It is, however, important to report all minor anomalies for cases with major malformations or syndromes.”12

The reason for recording of minor anomalies—morphological variations seen in 4% or fewer individuals in a population and of no surgical or cosmetic importance1—in such cases is that they may help elucidate cause, identify a particular syndrome, or clarify the gestational timing of a teratogenic insult. Surveillance of minor anomalies is, in itself, an important challenge for registries that is beyond the scope of this paper. In general, however, those cases with only 1 major defect are considered isolated, regardless of the number of minor anomalies, while those with only minor malformations are usually excluded from registration.
OMIM (Online Mendelian Inheritance in Man) codes, which also provide a consistent and in-depth coding system, can also be added, especially if the opportunity for recording certainty of diagnosis exists. In this way, cases with a confirmed diagnosis and those in which one is suspected but not proven can be easily retrieved using a combination of the ICD or OMIM code in combination with the certainty variable value.

It should be noted that advances in technology will make surveillance of syndromes more complex. Identification of mutations should be documented whenever possible. However, many single gene disorders are genetically heterogeneous and could involve mutations in more than one gene; eg, tuberous sclerosis type I due to hamartin mutations and tuberous sclerosis type 2 due to tuberin defects. The increasing use of microarrays and analysis of copy number variants will identify potential mechanisms that blur the lines between chromosomal and molecular pathogenesis. The coding of such cases is not straightforward with current ICD-9 or ICD-10 codes, without loss of important data. For both autosomal dominant and sex-linked single gene disorders and certain chromosomal conditions, awareness of whether the condition is inherited or represents a de novo event should be recorded, whenever this information is available. However, this may be an unrealistic goal for many registries. Clearly, syndromic cases are ones where additional information may well become available long after the initial diagnosis and reporting of the presenting malformations. Ideally, systems should allow for capture of such information at a later date.

**Sequence**

Sequences are another pattern of MCA that can present challenges for surveillance systems. A sequence is defined as a pattern of multiple anomalies where a malformation in one organ can lead to defects in other organs or tissues that were intrinsically normal. For example, if a fetus has posterior urethral valves, this will result in a secondary urethral stenosis, which will lead in turn to upper urinary tract obstruction and, if unrecognized and untreated early in gestation, serious dysplastic changes in otherwise normal renal tissue. This is in contrast to abnormal development in syndromes, where all the organs and tissues involved have the same underlying pathogenetic defect. Urethral stenosis and renal dysplasia also occur in Meckel syndrome, but in such cases the tissues involved were abnormal to begin with.

Obviously, monitoring of such MCA cases requires awareness of the sequence of events. Club foot in spina bifida cases or related to oligohydramnios due to renal adysplasia/aplasia is a secondary sequential malformation; club foot with tibial agenesis could also be part of a sequence (failure of the tibia to develop leads to a characteristic talipes equinovarus), but both may also be a primary malformation as mutations in PITX1 can lead to an autosomal dominant disorder where club foot is common, but some gene carriers also have tibial hemimelia and/or mirror image foot polydactyly (OMIM 119800). Again, it is important to code all anomalies seen in sequences as that will aid in evaluation of severity and potential for increased mortality, as well as the need for surgical intervention or rehabilitation services. When feasible, a text description can be incorporated in the record to identify a recognized sequence such as oligohydramnios sequence.

**Polytopic Field Defects and Malformation Association**

These last 2 patterns of MCA are somewhat interrelated and require an appreciation of the concept of a developmental field as a region or part of the embryo that “responds as a coordinated unit to embryonic induction and results in complex or multiple anatomic structures.” It is important to recognize that this is a hierarchical process: “During blastogenesis (the first 28 days of development), pattern formation in a primary field leads to the establishment of upstream expression domains of growth and transcription factors, which … lay down the pattern of progenitor fields. Further spatially coordinated, temporally synchronized, and epimorphically hierarchical morphogenetic events, mostly during organogenesis, lead to the secondary fields in which attainment of final form is achieved.”

The concept of coordinated fields is an old one in embryology and was largely based on observations concerning malformations. Key attributes of fields are heterogeneity (if different causes give rise to the same malformation, the affected part is presumably a morphogenetic unit under normal circumstances), phylogeneity (similar responses in different vertebrate species), and homology (concordance of homologous structures in different parts). It is because the forebrain develops as a field that a malformation such as holoprosencephaly (742.2; Q04.2) displays both etiologic heterogeneity (eg, trisomy 13, trisomy 18, and several single gene disorders) and phylogeneity (similar defects are seen in humans, mice, chickens, etc) and why, because the radial ray and tibial ray present homologous fields, preaxial defects in upper and lower limbs may occur together (eg, in preaxial polydactyly type II [OMIM 174500] or VACTERL association). Other examples of field defects include branchial arch anomalies, bronchopulmonary-foregut malformations, conotruncal heart defects, renal agenesis, vertebral segmentation defects, and caudal dysgenesis.

Because of shared molecular determinants, spatial contiguity and/or close timing of morphogenetic events during blastogenesis, many malformations that have their origins at this time are polytopic, ie, involving 2 or more progenitor fields, often in specific combinations. Thus polytopic field defects are distinct patterns of malformations in areas of the embryo that develop spatially and temporally as coordinated units such as the limb and renal fields (leading to acrorenal field defects), heart and limb (cardiomeclic), and stomach and limb (gastromelic). Polytopic field defects may be recognized as such, but have no known cause such as would be apparent in a karyotypically normal infant with bilateral renal agenesis and hypoplastic thumb. However, polytopic fields may be affected as a consequence of pleiotropy (the propensity for some gene mutations to affect more than one phenotypic characteristic), thus the distinction between syndromes and polytopic field defects appears to be less obvious than initially thought. From a surveillance
perspective, only careful recording of all defects will allow recognition of polytopic field defects and aid in evaluation of cause where this is unknown.

The report of an international working group on concepts and terms in errors of morphogenesis defined a malformation association as essentially a diagnosis of exclusion: a nonrandom occurrence in 2 or more individuals of multiple anomalies without a recognized underlying cause (ie, not known to be polytopic field defect, sequence, or syndrome). They perceived it as referring "solely to statistically, not pathogenetically or casually related anomalies."14 A dissenting opinion was that of Lubinsky16 who suggested that "since the embryo develops in an integrated manner ... disturbances result in nonrandom patterns of anomalies ... associations are derivatives of causally nonspecific disruptive events acting on developmental fields."

Later suggestions that that the term association be reserved for the original definition of a statistical combination of anomalies (mostly of organogenesis) and what were being called associations such as VACTERL "be designated primary polytopic developmental field defects, or simply polytopic field defects"15 were largely disregarded. This was in part due to a growing awareness that, as most malformations show positive associations on a general population basis, the statistical concept alone was less helpful than a greater awareness of the hierarchical nature of developmental fields.

Current thinking is that what we often think of as associations are potentially due to early disturbances in blastogenesis when the whole embryo is acting as a single field and key events such as formation of the notochord, separation of germ layers, and establishment of body axes are going on. As they are a field defect, they are casually nonspecific, but that doesn't mean that there isn't a cause. The defects in associations largely, but not exclusively, involve mesoderm. Common ones include VACTERL, MURCS (Müllerian ducts or Wolffian ducts in males-renal-cervicothoracic somite defects) in which radial ray and ear anomalies and occipital encephalocoele may also occur, OAVS (oculo-auricular-vertebral spectrum), posterior mesodermal defects such as caudal dysplasia and sirenomelia, and laterality defects. All anomalies should be coded in such cases along with the specific code when one currently exists or a more generic one such as 759.89 or Q87.89 with a text description denoting the recognized pattern.

When attention has been paid to careful recording and analysis of all anomalies in multiply malformed cases, surveillance systems have played a major role in identifying new associations such as SCHISIS (anencephaly, encephalocele, spina bifida, cleft lip, cleft palate, diaphragmatic defects, omphalocele)17 and TARCD (tracheal agenesis, cardiac and radial ray defects, duodenal atresia)18,19 and clarifying established ones, such as documenting the occurrence of tibial defects in VACTERL20 and the need to separate those cases of VACTERL with hydrocephalus from others with VACTERL-type anomalies.21 In some cases, they have allowed algorithms to be developed to help track environmental exposures. For example, using data on over 12,000 births from the Atlanta surveillance system, it was possible to show that the combination of ear defects + conotruncal and/or aortic anomalies + microcephaly, hydrocephaly or reduction defects of the brain was associated with isotretinoin embryopathy with a sensitivity of 45.9%, a specificity of 99.9%, and a positive predictive value of 85%.22

Laterality

A further point with respect to coding multiple anomalies relates to the sidedness of defects in paired organs. An anomaly may be found on one side only (left or right), be found on both sides in a symmetric pattern, or be present on both sides, but variable in presentation or severity. With radial ray defects, for example, one case might have a duplicated thumb on the right side, another could have bilateral radial ray aplasia with absent thumbs, while a third has thumb hypoplasia on one side and a triphalangeal thumb on the other. Sidedness is a very useful criterion in interpretation of patterns of defects. For example, bilateral defects, especially when symmetric, are more often syndromic; unilateral ones relate more to field defects. In addition, the pattern of associated anomalies may be different for a right-sided defect versus a left-sided one as is seen with diaphragmatic hernia and certain heart defects23 or with ipsilateral versus contralateral anomalies in acrorenal field defects.24

Unfortunately, only a few ICD-9 or ICD-10 codes for oral facial clefts and renal and hip defects specify whether a defect is unilateral or bilateral, and none document sidedness. The Royal College of Paediatrics and Child Health adaptation of ICD-10 allows for more detail, especially with respect to orofacial clefts, and diaphragmatic and genitourinary defects. ICDBSR suggests adding a descriptor after a Q code for paired organs to indicate laterality: UL, UR, UU (unknown), BI, LU (laterality unknown).11

Clinical Review and Case Classification

All surveillance systems should have recourse to 1 or more dysmorphology experts to aid with complex coding issues and to provide assistance in determining whether the pattern of anomalies in a multiply malformed case can be considered a syndrome, sequence, polytopic field defect, or association, and coded appropriately. If it is believed that malformations in a particular individual may represent a pathogenetically related disorder, but its nature is unclear, the code 759.7 “Multiple congenital anomalies, so described” or Q89.7 “Multiple congenital malformations, not elsewhere classified” could be used to allow retrieval of such cases for further review at a later date. It should be recognized that the use of these nonspecific codes has the potential to create a large pool of otherwise uncharacterized MCA cases; however, it does allow recognition and retrieval of cases that have MCA, but would be hard to identify without inspection of individual case records to evaluate all the codes applied.

Although likely not feasible at the individual program level due to small numbers of cases, it is possible in large data sets to use an algorithm to derive a subset of cases that have 2 or more unrelated major structural anomalies
that do not appear to be part of a specific syndrome or sequence. These can then be examined in greater detail for phenotypic findings and potential etiology. Application of this algorithm to the 2004 EUROCAT data indicated that it was “a feasible, efficient, and transparent way to improve classification of congenital anomalies for surveillance and research.”3 In total, 1,862 (10.5%) of 17,733 cases were initially considered to be potentially unexplained MCA as opposed to isolated defects/sequences, chromosomal disorders, or other syndromes. In-depth review by 3 geneticists led to removal of a further 557 cases, leading to a final figure of 1,305 (7%). The analysis documented that the proportion of unexplained MCA cases varied considerably by malformation type; eg, 11% for congenital heart defects, 24% for hydrocephalus, and 35% for bilateral renal agenesis. As would be expected, there were occasional discrepancies between the geneticists as to whether the reported MCA in a particular case represented a true multiple, a sequence, or an isolated defect. In some situations, such disagreements helped refine surveillance practices as they were followed by written rules for future coding decisions.

Another issue relates to the extent to which fetuses and infants with malformations are investigated clinically. The identification of even 1 malformation may lead to a more detailed physical examination and the use of diagnostic imaging or genetic testing. While such investigations may ultimately increase the frequency with which MCA cases are detected and recorded as such, ascertainment bias alone is unlikely to have a major impact on the strength of associations between specific anomalies, unless subsequent clinical evaluations are limited in scope or excessively targeted. An example might be restricting the further evaluation of an infant with a radial ray deficiency to external examination alone, rather than considering such investigations as a skeletal survey, abdominal ultrasound, hearing assessment, and karyotype or microarray. From a surveillance perspective, it would be ideal to be able to record both prenatal and postnatal investigations and their results, especially whether an autopsy or whole body MRI imaging was carried out in deceased individuals.

Statistical Issues
Patterns of MCA are recognized because certain malformations occur together more often than would be anticipated by chance. Thus analysis of patterns of birth defects, or birth defects clustering as it is sometimes referred to, is often part of the interpretation of surveillance data. This is a complex issue from a statistical standpoint and a detailed review is beyond the scope of this commentary. However, a few basic points will be discussed. The simplest measure of whether 2 defects, A and B, are seen together more often than expected is the syntropy index, where the observed (O) number of cases with both defects in a given population is compared to the expected number (E) as the product of the frequencies of A and B in the same population. A ratio of O/E greater than 1.00 would indicate nonrandom clustering and is observed for almost all pairs of anomalies. In fact, this method frequently gives very high indices due to the fact that birth defect clustering is both generalized and nonspecific.4

There is also the issue that to have MCA, a case must have at least 2 anomalies, yet the population data used to derive E is largely composed of cases with isolated A or B. Thus an alternative method of analysis was proposed in which the expected figures were based only on the proportion of cases in the population that were multiply malformed.2 Ratios derived in this way tended to be much lower and it was shown that this method was biased by the tendency for all the anomalies other than A and B to cluster together. A further refinement was suggested by Khoury et al25 whereby the expected ratio was derived using only the MCA cases that had at least 1 of A or B. When they applied this technique to analyze cases with VACTERL type anomalies from the Atlanta Birth Defects Registry, it was obvious that observed-to-expected ratios were higher than seen using all multi-malformed cases, but still much less than the unadjusted rates; respective ratios for the combination of renal and vertebral defects, for example, were 11.4, 2.1 and 187.3.25

Using this methodology, these authors also showed that some of the midline malformations often considered part of SCHISIS association or seen in other recognized patterns do not show significant pairwise associations even when the ratios were compared using cases with at least 1 of the defects. While omphalocele + anal atresia and neural tube defect + oral cleft had a highly significant ratios of 13.3 and 4.3 respectively, neural tube defect + tracheoesophageal defect (O/E 1.4) and oral cleft + diaphragmatic hernia (O/E 1.8) no longer showed significant association, indicating that the clustering of midline malformations is not generalized and nonspecific. Further analysis of a hypothetical set of cases with central nervous system and/or ear defects in a population potentially exposed to isotretinoin indicated that the strength of the clustering between these 2 defect classes depended to a great extent on the frequency of the exposure in the population, a factor that will need to be taken into account when surveillance data are used to assess patterns of anomalies associated with environmental factors.29

Many other methods of statistical analysis exist, including numerical taxonomy26,27 and log-linear28 and multiple regression29 models. Multivariate techniques have also demonstrated the fact that there are many confounders that can influence the results, such as whether or not autopsy data were available, the specific surveillance program, and demographic factors such as maternal age.29

Summary
While MCA do create significant challenges for surveillance, attention to identifying and analyzing cases with complex patterns of birth defects can provide much useful information for end users of surveillance data. It is hoped that the information and suggestions provided in this opinion paper will encourage such activities to be incorporated into registry systems. They are intended to highlight the need and opportunity for many individuals with complementary expertise in disciplines such as epidemiology, medical
genetics, statistical analysis, knowledge transfer, and public health to work together to turn surveillance data into positive action for the benefit of affected individuals, their families and caregivers, and society at large.

Acknowledgments

The author would like to thank local colleagues, members of the Canadian Congenital Anomalies Surveillance Network and of the Section of Maternal and Child Heath, Public Health Agency of Canada for useful discussions.

References

Abstract: Background: Fetal death certificates (FDCs) are the main source of stillbirth surveillance data in the United States, yet previous studies suggest FDCs have incomplete ascertainment. In 2005, the Centers for Disease Control and Prevention (CDC) funded 2 pilot programs to determine the feasibility of expanding existing birth defects surveillance systems employing active casefinding methods to conduct surveillance of stillbirths. The objectives of this analysis were to: 1) estimate the completeness of ascertainment of stillbirths identified through one of the pilot programs, the Metropolitan Atlanta Congenital Defects Program (MACDP), and 2) compare the prevalence of stillbirths obtained through active casefinding (MACDP) with data available from FDCs. Methods: Stillbirths in metropolitan Atlanta were independently ascertained by both FDC and MACDP in 2006 and 2008. Capture-recapture methods were used to estimate the total number of stillbirths in the surveillance area. The sensitivities for capturing stillbirths were estimated for FDCs, MACDP, and both sources combined. Prevalence estimates for each data source and for the combined data sources were calculated using a denominator of live births plus FDC-identified stillbirths. Results: An estimated 1,118 stillbirths occurred in metropolitan Atlanta. MACDP captured 863 and FDCs captured 862. There were 198 stillbirths captured by MACDP and not reported by FDC, and 197 stillbirths identified by FDCs that were not initially captured by MACDP. The estimated sensitivities were 77.1%, 77.2%, and 94.8% for FDCs, MACDP, and both sources combined, respectively. The stillbirth prevalences for 2006 and 2008 using FDC data alone were 8.2 and 7.4 per 1,000 live births plus stillbirths, respectively, and 9.9 and 9.3 per 1,000 live births plus stillbirths, respectively, using both data sources combined. Conclusions: Leveraging the resources of existing birth defects surveillance programs in combination with FDCs could improve population-based ascertainment of stillbirths.

Key words: birth defects, prevalence, sensitivity, stillbirth, surveillance

Introduction

Stillbirth is an important public health concern. Despite improvements in prenatal and perinatal care in recent decades, stillbirth occurs in approximately 1 out of every 200 pregnancies and has a tremendous emotional and psychological impact upon families.1,2

Although reporting requirements vary, stillbirth is a reportable event in all 50 states as well as US territories. Data on stillbirths are regularly collected, analyzed, and reported by the National Center for Health Statistics (NCHS) through collaborative agreements with states as part of the National Vital Statistics System.3 Based on these data provided to NCHS, in 2005 the prevalence of stillbirths in the United States was 6.22 per 1,000 live births plus stillbirths.4 The use of vital records for surveillance purposes, however, has been problematic.5,6 The American College of Obstetricians and Gynecologists (ACOG) has published recommended guidelines for conducting postmortem stillbirth evaluations.7 However, several studies have shown that data on fetal death certificates (FDCs) not only yield inaccurate and incomplete information with respect to certain variables such as maternal health conditions, presence of a birth defect, and causes of death, but they also potentially underestimate the true prevalence of this event.10-15 Without reliable population-based data, the conduct of epidemiologic studies of risk factors and causes of stillbirth are challenging.

In 2005, the Centers for Disease Control and Prevention (CDC) funded 2 pilot projects—1 in Iowa and 1 in metropolitan Atlanta—to assess the feasibility of expanding existing population-based birth defects surveillance programs to include surveillance of stillbirths with or without birth defects. The hypothesis was that using the infrastructure of established birth defects surveillance programs employing active casefinding methods to collect, analyze, and report data on stillbirths could enhance existing surveillance information on stillbirths. These enhancements would need to demonstrate improvements not only in the quantity and quality of information collected, but also completeness of case ascertainment. In 2008, Duke et al evaluated a revised data collection tool for use in the surveillance of stillbirths as part of the Metropolitan Atlanta Congenital Defects Program (MACDP).16 After linking MACDP-identified stillbirths with FDCs, the analysis demonstrated that overall there was less missing information for critical variables, such as birth weight and fetal sex, compared with corresponding information on FDCs. Also, the amount and quality of clinical and pathological information abstracted

__________

National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia.
Address correspondence to Wes Duke, MD, MPH; National Center on Birth Defects and Developmental Disabilities, Division of Birth Defects and Developmental Disabilities, Mail Stop E-86, Atlanta, GA 30333. Email: CB7@cdc.gov.
CDC Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.
from the medical record through MACDP surveillance was improved and could potentially allow for a better understanding of the contributing factors associated with the fetal death. The current paper reports the results of a follow-up study; the objectives were to evaluate the completeness of case ascertainment and compare prevalence of stillbirths identified through MACDP, FDCs, and both data sources combined for the years 2006 and 2008.

**Methods**

**Pregnancy Outcome Determination**

Pregnancy outcome classification was based on the definitions for live birth, fetal death, and induced termination of pregnancy provided by the 1992 Revision of the Model State Vital Statistics Act and Regulation (Model Law, page 2). There is no universally accepted definition of stillbirth that includes the criteria for gestational age or birth weight. For surveillance purposes, stillbirth was defined by MACDP as a fetal death occurring at 20 or more weeks of gestation or 350 or more grams if the gestational age is not known. The gestational age used was the age of the fetus as indicated by the physician in the medical record. Lastly, the Model Law defines an induced termination of pregnancy as “… the purposeful interruption of an intrauterine pregnancy with the intention other than to produce a live born infant and which does not result in a live birth and ... excludes management of prolonged retention of products of conception following fetal death.” While fetal heart tones may be present and documented in the medical record prior to the induction of labor, the “intention” is not always clear; therefore, assessing misreporting of these outcomes as fetal deaths is problematic. For MACDP stillbirth surveillance purposes, these cases are ascertained and reported as stillbirths resulting from medical intervention along with the indication for induction of labor (Figure 1).

**Case Ascertainment**

Stillbirths were independently ascertained though FDCs provided by the state of Georgia and MACDP. Georgia requires all fetal deaths to be reported if brought to the attention of a health care provider; more information on fetal death registration requirements in Georgia can be found in Chapter 31 of the Official Code of Georgia (O.C.G.A § 31-10-18). Prior to 2006, MACDP received FDCs on an ongoing basis as 1 data source for ascertainment of birth defects; however, due to an administrative lapse, FDC for 2006 and later were not obtained until late 2009, allowing for the independence of sources in casefinding for the current assessment. A complete file of FDCs for 2007 was never obtained, necessitating the exclusion of that year from this analysis.

MACDP is a population-based active surveillance system ascertaining structural and chromosomal anomalies among pregnancies resulting in a live birth, stillbirth, or termination. Trained medical records abstractors visit area birthing hospitals, pediatric hospitals, and other clinical providers including prenatal diagnostic centers and genetics clinics located in the 5 central counties of metropolitan Atlanta to identify and abstract information on potential cases. In 1994, MACDP abstractors began to visit the outpatient offices of area perinatologists and maternal-fetal medicine specialists to abstract information about pregnancies diagnosed prenatally with congenital abnormalities. Clinical reviewers review each potential case and determine eligibility for inclusion in the surveillance system and code the birth defects. MACDP methods for birth defects surveillance have been previously described. In 2006, after revisions to the data collection tool and surveillance methods, MACDP began active surveillance for all stillbirths, with or without birth defects. The sources for active ascertainment of stillbirths by MACDP largely overlapped with the sources for birth defects ascertainment but included a few additional sources such as emergency department records and autopsy and placental histopathology reports. In addition, mothers diagnosed with an intrauterine fetal death at 20 or more weeks of gestation in the specialty clinics previously mentioned were also ascertained with follow-up attempted at the delivering hospital. Stillbirths which occur without any resulting contact with a health care provider (eg, no emergency department visit, hospitalization, or visit to selected Atlanta-area prenatal care providers) are not able to be ascertained by MACDP. Furthermore, MACDP does not have access to abortion clinic records and any stillbirths or terminations occurring at such facilities would be missed.

Live birth certificates and FDCs for 2006 and 2008 were obtained from 2 departments within the Georgia Department of Public Health. In 2006, data came from the Office of Health Indicators for Planning and in 2008 from the Office of Vital Statistics. The records of stillbirths identified through MACDP were linked to FDCs for the same birth cohorts (2006 and 2008) by means of a deterministic matching process with multiple iterations using the following variables: mother’s name and race, father’s name, gender of fetus, date of event, hospital, county of residence, and mother’s address at the time of delivery. Manual matches were also attempted for stillbirths that did not link. For those stillbirths in the FDC that did not link to a stillbirth in MACDP, abstractors were asked to locate
the medical records for those stillbirths and abstract the relevant information if the mother was a resident of the 5-county surveillance area.

Data Analysis

To evaluate the total number of stillbirths occurring among the surveillance population and the relative contribution of each data source for casefinding (active surveillance through MACDP; passive surveillance through FDC), capture-recapture methods were used. Briefly, this method can be used to estimate total prevalence and to evaluate the relative contribution of independent case sources. The number of stillbirths missed by both sources was estimated by the product of the number missed by each source, divided by the number identified by both sources. These stillbirths missed by both sources were then added to the total number identified by either source to estimate the total prevalence. The prevalence of stillbirths was then calculated by each data source alone as well as for both data sources combined.

Table 1. Fetal Deaths Reported by Gestational Age, Georgia and Metropolitan Atlanta, 2006 and 2008

<table>
<thead>
<tr>
<th>Gestational age</th>
<th>Georgia 2006</th>
<th>Georgia 2008</th>
<th>Metropolitan Atlanta 2006</th>
<th>Metropolitan Atlanta 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>&lt; 20 weeks*</td>
<td>6,894</td>
<td>71.9</td>
<td>5,061</td>
<td>62.1</td>
</tr>
<tr>
<td>≥ 20 weeks</td>
<td>1,141</td>
<td>12.6</td>
<td>1,111</td>
<td>13.4</td>
</tr>
<tr>
<td>20–27 weeks</td>
<td>722</td>
<td>4.6</td>
<td>652</td>
<td>7.9</td>
</tr>
<tr>
<td>28+ weeks</td>
<td>419</td>
<td>6.8</td>
<td>459</td>
<td>5.5</td>
</tr>
<tr>
<td>Missing</td>
<td>985</td>
<td>2.5</td>
<td>2,092</td>
<td>25.3</td>
</tr>
<tr>
<td>Total</td>
<td>9,020</td>
<td>100</td>
<td>8,264</td>
<td>100</td>
</tr>
</tbody>
</table>

*Fetal deaths before 20 weeks of gestation, also known as miscarriages or early fetal losses, were not considered in this analysis.

2006 data were from Georgia Department of Public Health, Office of Health Indicators for Planning.

2008 data were from Georgia Department of Public Health, State Office of Vital Records.

Table 2. Distribution of Stillbirths by Source of Identification and Estimated Total Stillbirths in Metropolitan Atlanta, 2006 and 2008

<table>
<thead>
<tr>
<th>Identified by MACDP</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified by FDC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>665</td>
<td>197*</td>
<td>862</td>
</tr>
<tr>
<td>No</td>
<td>198</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>863</td>
<td>255</td>
<td>1,118</td>
</tr>
</tbody>
</table>

MACDP, Metropolitan Atlanta Congenital Defects Program; FDC, fetal death certificates.

*Includes 30 stillbirths for which no medical record could be found and 26 that were delivered in a county outside MACDP catchment area.

**Estimated stillbirths missed by both sources.

Results

In 2006 and 2008, there were 2,252 stillbirths reported in the state with just under half of these occurring among mothers residing in the 5-county metropolitan Atlanta area (Table 1). Because Georgia law requires that all fetal deaths be reported regardless of gestational age if brought to the attention of a health care provider, the majority of fetal deaths in Georgia are losses before 20 weeks of gestation (Table 1). The year 2008 had substantially more fetal deaths with a missing gestational age than the year 2006 because of differences in data sources from the state. The data from 2006 from the Office of Health Indicators for Planning had missing clinical estimates of gestational age recoded as the gestational age based on the last menstrual period, if available. The data from 2008 were the raw vital statistics data that did not undergo this assignment process.

MACDP captured 863 stillbirths and FDCs captured 862. Of these, 665 stillbirths were independently captured by both sources (Table 2). MACDP captured an additional 198 stillbirths for which no FDC could be found. Similarly, a total of 197 stillbirths were identified based solely on the FDC for casefinding and abstraction. Using capture-recapture methods, 58 stillbirths were estimated as missed by both sources ((198 × 197)/665 = 58), resulting in 1,118 total stillbirths (95% confidence interval [CI], 1,052–1,183) in metropolitan Atlanta during 2006 and 2008. The estimated sensitivities for capturing a stillbirth were 77.1% (95% CI, 74.7%–79.7%) for FDC alone, 77.1% (95% CI, 74.7%–79.7%) for MACDP alone and 94.8% (95% CI, 93.5%–96.1%) for both sources combined. Of the 197 stillbirths identified solely through FDC, medical records were sought but not found for 30 of them, and 26 additional stillbirths occurred among mothers who resided within the catchment area, but delivered in a facility outside of it. The medical records for these 26 stillbirths were not sought by MACDP.
Furthermore, MACDP captured 61 stillbirths for which induction of labor was performed due to the fetus being affected by a birth defect. Of these, 31 linked to FDC and 30 could not be linked. MACDP ascertained another 49 stillbirths for which induction of labor was performed secondary to a pregnancy complication such as preeclampsia, premature rupture of membranes, or chorioamnionitis. Thirty-seven of these were issued a FDC and 12 did not link (Table 3). The 30 cases for which the medical record could not be found and the 26 cases that were delivered outside the catchment area are not included in the assessment of ascertainment by pregnancy outcome reported in Table 3.

Lastly, there were an additional 114 stillbirths identified through FDC with a gestational age of 20 or more weeks that were subsequently excluded after reviewing the medical record. The reasons for excluding these cases are listed in Table 4. Forty-five cases were excluded after review of the medical record clearly indicated that the death occurred before 20 weeks of gestation. Thirty-four cases were singleton stillbirths for which 2 identical FDCs were generated. Twenty cases had medical record documentation that the fetus was born alive and expired shortly after birth. Another 13 cases were excluded because the mother did not reside in the surveillance catchment area and 2 cases had the wrong year of birth on the FDC.

There were 55,707 and 54,581 live births and stillbirths (the stillbirths in the denominator were based on the number ascertained by FDC) delivered in the metropolitan Atlanta surveillance area in 2006 and 2008, respectively. Using only those stillbirths identified from FDCs, the prevalence of stillbirth was 8.2 per 1,000 live births plus stillbirths in 2006 (95% CI, 7.4–8.9) and 7.4 per 1,000 live births plus stillbirths in 2008 (95% CI, 6.7–8.2). Using only ascertainment by MACDP, the estimates were 8.0 and 7.6 per 1,000 live births plus stillbirths (95% CIs, 7.3–8.7 and 6.9–8.4), respectively. Using both sources for ascertainment yielded estimates of 9.9 and 9.3 per 1,000 live births plus

### Table 3. Distribution of Stillbirths by Pregnancy Outcome, Initial Source of Identification, and Linkage Status, Metropolitan Atlanta, 2006 and 2008

<table>
<thead>
<tr>
<th>Pregnancy Outcome</th>
<th>Linked Identified by</th>
<th>Did Not Link Identified by</th>
<th>Total Identified by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MACDP</td>
<td>FDC</td>
<td>MACDP</td>
</tr>
<tr>
<td>Stillbirth with birth defect</td>
<td>82</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Stillbirth without birth defect</td>
<td>520</td>
<td>130</td>
<td>139</td>
</tr>
<tr>
<td>Stillbirth due to induction of labor for birth defect</td>
<td>31</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Stillbirth due to induction of labor for other pregnancy complication*</td>
<td>32</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>665</td>
<td>139</td>
<td>198</td>
</tr>
</tbody>
</table>

**FDC, fetal death certificate; MACDP, Metropolitan Atlanta Congenital Defects Program.**

*Complications such as chorioamnionitis, premature rupture of membranes, and pre-eclampsia

**Does not include 30 stillbirths for which no medical record could be found and 26 that delivered in a county outside MACDP catchment area.

### Table 4. Reasons for Excluding Stillbirths Identified Through Fetal Death Certificates as Occurring at 20 or More Weeks of Gestation, Metropolitan Atlanta, 2006 and 2008

<table>
<thead>
<tr>
<th>Reason for Exclusion</th>
<th>N (%).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical record indicated less than 20 weeks*</td>
<td>45 (39.5)</td>
</tr>
<tr>
<td>Duplicate FDC</td>
<td>34 (29.8)</td>
</tr>
<tr>
<td>Live birth</td>
<td>20 (17.5)</td>
</tr>
<tr>
<td>Non-resident**</td>
<td>13 (11.4)</td>
</tr>
<tr>
<td>Wrong year of FDC</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>Total</td>
<td>114 (100)</td>
</tr>
</tbody>
</table>

FDC, fetal death certificate.

*Medical record review clearly indicated the death occurred prior to 20 weeks of gestation indicating that the death was misclassified as a death occurring at less than 20 weeks on the FDC.

**Medical records documented these mothers as not residing in the metropolitan-Atlanta surveillance catchment area.

### Figure 2. Prevalence of Stillbirths by Data Source, Metropolitan Atlanta, 2006 and 2008

[Graph showing the prevalence of stillbirths by data source for 2006 and 2008.]

FDC, fetal death certificate; MACDP, Metropolitan Atlanta Congenital Defects Program.
stillbirths (95% CIs, 8.1–10.7 and 8.5–10.1), respectively (Figure 2). Prevalence estimates were compared including and excluding stillbirths occurring after induction of labor as a medical intervention (Figure 3). Excluding stillbirths occurring after induction of labor naturally reduced the prevalence; we observed a greater reduction for MACDP identified stillbirths than FDC-identified stillbirths.

**Table:**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Prevalence per 1,000 Live Births and Stillbirths</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDC Only</td>
<td>7.0</td>
</tr>
<tr>
<td>MACDP Only</td>
<td>7.2</td>
</tr>
<tr>
<td>FDC + MACDP</td>
<td>7.2</td>
</tr>
</tbody>
</table>

*Excludes stillbirths occurring after induction of labor.
**MACDP numerator includes stillbirths identified through FDC and linked (n = 139)
 FDC, fetal death certificate; MACDP, Metropolitan Atlanta Congenital Defects Program.

**Discussion**

The use of FDC alone for population-based surveillance of stillbirths is limited, and uncertainty about the utility of FDC data for risk-factor analysis has been previously noted. Complete and reliable surveillance data is needed if hypothesis-driven epidemiologic studies are to be conducted. The current study demonstrated that expanding the capabilities of an existing birth defect surveillance system to include active ascertainment of stillbirths, and combining that information with what is gathered from FDCs, is feasible and results in the ascertainment of cases that would have otherwise been missed by either system alone. Expanding MACDP to include surveillance of stillbirths required a few modifications to the birth defects surveillance protocol, such as accessing emergency department records (to capture cases arriving as an emergency and potentially getting discharged without hospital admission), and autopsy and placental histopathology information. There was no additional staff required to implement stillbirth surveillance as sources for casefinding overlapped with sources already used for birth defects surveillance and were already being reviewed by clinical abstractors.

Each data source—FDC and MACDP—has limitations. The current analysis, as well as previous studies, indicate that FDCs underreport stillbirths, as well as often misreport the pregnancy outcome and contain large amounts of missing information for critical variables such as gestational age, birth weight, and cause of death. With respect to cause of death, this may in part be explained by the fact that the majority of FDCs are completed before all postmortem evaluation information is available. For MACDP, our analyses suggest that a large number of stillbirths would have been missed if not for the availability of FDCs as a source for casefinding. Routine procedures for MACDP normally involve obtaining FDC on a monthly basis, from which stillbirths can be identified on an ongoing basis. This was not the case for our study years, allowing for the application of capture-recapture methods to estimate the number of stillbirths occurring in the surveillance population. Active ascertainment by MACDP was only able to collect what was available from the medical record. However, when both MACDP and FDCs were used together, they ascertained more stillbirths than either system captured independently. The factors influencing case ascertainment within each data source are not clear. The underascertainment of stillbirths is not likely a random event; it may be associated with factors such as maternal race/ethnicity, gestational age, delivery facility, or the cause of death, or perhaps factors that are not even recorded. More analyses need to be undertaken to better understand the role that these factors may play in the ascertainment of stillbirths. This could provide potentially valuable information to inform training needs and strategies to improve the reporting process.

Active casefinding of stillbirths has several strengths. Trained abstractors visit area hospitals, locate medical records for potential cases, and record the relevant information. The abstracted information for each potential stillbirth is systematically reviewed by 1 or more MACDP clinicians to ensure that inclusion criteria are met and to designate the appropriate outcome classification. Previous studies by Duke et al, using data from MACDP, have demonstrated that active ascertainment and medical chart review improves upon the quantity and quality of the data collected. In addition, the in-depth medical record review resulted in more accurate classification of pregnancy outcomes, which provided insight into the potential misclassification of pregnancy outcomes by FDCs. This information is important to better understand stillbirth prevalence estimates that are based on FDCs alone. Active ascertainment can allow for the inclusion or exclusion of stillbirths resulting from the medical induction of labor or stillbirths that were actually live births. As shown in Table 3, about 50% of the inductions performed for a fetus affected by a birth defect linked to a FDC, whereas about 75% of those inductions performed for other pregnancy complications were issued a FDC. It is not possible to know if the cases that did not link were issued an induced termination of pregnancy (ITOP) certificate, data which are unidentified and unlinkable with information from other sources. Anecdotally, the majority of inductions performed in the setting of a fetus affected by a birth defect are done subsequent to the administration of intrauterine potassium chloride, and identifying these events through the review of medical records is relatively straightforward. Therefore, from a surveillance perspective, the intent of the procedure is apparent, and ITOP certificate should have been issued. However, many of these birth defects can
be considered lethal anomalies and should be considered when understanding fetal mortality rates. On the other hand, inductions performed in the context of other clinical scenarios, such as severe chorioamnionitis, are most often conducted in the best medical interest of the mother and the intent may very well have been to produce a live birth. It is likely that these clinical situations explain the differences in whether a FDC was issued or not. Active casefinding allows for these events to be captured and documented based on the thorough review of medical records, potentially improving our understanding of the impact of these events on estimates of the prevalence of stillbirths. These distinctions cannot be made when using FDC data alone.

We capitalized on a lapse of availability of FDCs to MACDP for case ascertainment. Having 2 independent data sources for stillbirth ascertainment allowed us to conduct a capture-recapture analysis to estimate the total number of stillbirths occurring within metropolitan Atlanta, and the number potentially missed by the 2 data sources working independently. This normally cannot be done when FDCs are obtained and used on an ongoing basis as a source for casefinding.

This analysis is subject to several limitations, however. First, we were limited to only 2 years of data; a similar analysis is planned for stillbirths occurring in 2009 and later. Second, we did not assess or compare data quality between sources, an important next step to further demonstrate the utility of this approach to stillbirth surveillance. Third, it was not possible or practical to capture every fetal death. Stillbirths that occurred to mothers residing in metropolitan Atlanta but delivering outside the catchment area were missed by MACDP and were therefore not subjected to medical chart review. However, they could be identified if issued a FDC and were included in the analysis as shown in Table 2. Similarly, medical records for terminations and stillbirths delivered at abortion clinics are not accessible by MACDP. Lastly, it is not clear why such a large number of medical records could not be located (n = 30). This may reflect inadequate staffing and resources to conduct an exhaustive search for the medical record as many health care facilities store medical records offsite after a certain length of time.

Fetal death reporting by states to the National Vital Statistics System is and will remain the core infrastructure for stillbirth surveillance in the United States; however, expanding existing birth defects surveillance programs to include active ascertainment of stillbirth is potentially a valuable approach to help address our current knowledge gaps about the frequency and risk factors for stillbirths. More importantly, improvements to surveillance data on stillbirths will require multidisciplinary efforts to increase and standardize the use of ACOG recommended clinical guidelines for postmortem stillbirth evaluation.

References


Original Article

An Informatics-Enabled Approach for Detection of New Tumor Registry Cases

Riyad Naser, MS, MBA, PhD\textsuperscript{a,b}; Judith Roberts\textsuperscript{a}; Todd Salter\textsuperscript{b}; Jeremy L. Warner, MD, MS\textsuperscript{b,c}; Mia Levy, MD, PhD\textsuperscript{d,e}

Abstract: Tumor registries are held to a very high standard for identifying and reporting new analytic cancer cases. However, current approaches to new case detection are often inefficient and costly. Efficient and effective detection of new cancer cases has the potential to maintain a high accuracy of reporting while reducing costs, increasing timeliness of reporting, and ultimately advancing cancer research. We describe the development, implementation, and evaluation of an informatics tool that integrates multiple data sources to support the workflow of new case identification at the Vanderbilt University Medical Center (VUMC) tumor registry office. The new system reduced the total number of potential cases to analyze from roughly 13,000 to 2,500 records per month. This resulted in an efficiency gain of roughly 80 man-hours per month with a respective annual savings of approximately $50,000. Further iterative refinement of this approach along with support for case abstraction could result in further efficiencies.

Key words: accession, new cases identification, tumor registry, workflow

Introduction

Cancer registries have been in place in the United States since the 1930s, with a goal of creating national statistics on cancer incidence and mortality.\textsuperscript{1} Since then, tumor registries have played a significant role in the advancement of cancer research. They have been used as multipurpose tools for various epidemiology and cancer research disciplines. For example, they have been used to identify causes of cancer in geographical regions, cancer trends, intervention efficacy and treatment effectiveness, and survivorship analysis.\textsuperscript{2,7} Tumor registries offer the potential to help assess the cancer burden in a geographical region with resultant impact on the proper allocation of resources and initiatives to advance cancer research.\textsuperscript{8}

Tumor registry staff perform several key tasks including: 1) detection of new analytic cases, 2) abstraction of new analytic cases, 3) long-term follow-up of vital status of analytic cases, 4) transmission of results to the state registry, and 5) generation of population-based reports for institutional operations and research. Most institutions that care for cancer patients must maintain a highly trained staff to perform these tasks. Such individuals are referred to as certified tumor registrars (CTRs). Unfortunately, commercially available software tools support these tasks to varying degrees. In particular, the processes for detecting and abstracting new analytic cases have been variously reported as ineffective, inefficient, erroneous, and time-consuming.\textsuperscript{9} Informatics-based solutions for new case detection and case abstraction could lead to more complete and timely reporting with fewer staff members resulting in decreased costs.\textsuperscript{8,9,11} We describe the creation, implementation, and evaluation of an informatics tool to detect potential new cancer cases at the Vanderbilt University Medical Center (VUMC) tumor registry office.

Background

Vanderbilt University Medical Center Tumor Registry Office

The VUMC tumor registry office is certified by both the American College of Surgeons (ACoS) and the Commission on Cancer. The office identifies and abstracts approximately 5,000 new analytic cases per year, and is staffed by 1 manager, 8 nurses, and 1 data-management specialist. All tumor registry nurses are certified tumor registrars.\textsuperscript{3} In accordance with ACoS requirements, the VUMC tumor registry maintains over 90% completeness of information on treatment and follow-up.\textsuperscript{11}

Preimplementation Case Detection Process

The VUMC tumor registry uses 2 primary data sources to detect new analytic cases: 1) billing data from the finance system, and 2) pathology data from the laboratory information system. Figure 1 illustrates how the finance report is processed for new case detection. A finance report is automatically generated each month from a query of billing records in the VUMC enterprise data warehouse (EDW). The query looks for International Classification of Diseases (ICD)-9 billing records in the ranges of M 8***2, 3, and M 9**2, 3. For each ICD-9 code detected, a new row is generated containing patient demographics, the date of the encounter, and the ICD-9 code. As such, if a patient is seen multiple times in 1 month, or has multiple relevant ICD-9 codes, then the patient will have multiple rows in the same file. Each monthly finance report contains an average of 13,000 records.

The tumor registry director manually downloads the comma-delineated file and imports the file into Microsoft Excel. The Excel file is then imported into the tumor
registry software application. The import process takes approximately 3 to 4 hours and often fails due to the large data set. Additional manual data processing in Excel is normally needed after creating the registry suspense list. This manual process involves high-level elimination of duplicates, removing nonreportable cases, removing odd formatting, and breaking down the data set into equal and manageable parts for each nurse. Each nurse then receives a multicolor-coded Excel sheet labeled according to each nurse’s name that contains an average of around 2,000 records. The in-depth investigation of each medical record involves various steps, many systems, and takes many days to complete. Investigating the entire Excel spreadsheet file takes approximately 6 to 7 weeks. From the 13,000 records, the registry nurses abstract roughly 400 to 600 new cases every month.

The surgical pathology report is generated each month from a query of pathology cases in the VUMC laboratory information system. The query looks for pathology cases annotated with Systematized Nomenclature of Medicine—Clinical Terms (SNOMED) codes in 2 ranges of interest: M-8***2, 3 and M-9**2, 3. The report contains records of cancer patients with a tissue diagnosis at the VUMC as well as pathology consults from patients diagnosed at outside institutions. The comma-delimited file contains partial demographic information, the date of specimen acquisition, the date of specimen acquisition, and the respective SNOMED code.

Inefficiencies in the new case detection process described above stem predominantly from the nature of the input data sources and the inability to compare them to cases already abstracted in the tumor registry software. These inefficiencies have resulted in the need to hire additional registry staff in order to meet timely reporting requirements. Given that the number of analytic cases continues to grow at VUMC each year, we needed to take steps to improve the efficiency of the case detection process so as not to increase our costs while maintaining high-quality and timely reporting. With these goals in mind, we developed a set of informatics tools to help improve the efficiency of the case detection process for the VUMC tumor registry.

**Methods**

Prior to any system modifications or design, we conducted an analysis of the current workflow, information systems, and data artifacts previously described. The purpose of the analysis was to identify interventions that could improve the efficiency of the new case detection process. The analysis consisted of personnel interviews, direct observation, and collection of data artifacts. Five individuals in the tumor registry department were selected for interviews based on their diverse roles in data collection, analysis, and abstraction. A total of 15 hours of personal interviews, 30 hours of observation time, 12 reports, and 35 screen shots were collected and analyzed over a period of 6 months.

We first analyzed the finance reports to determine the percent of potential new cases that were duplicate rows for the same patient or already entered into the tumor registry database as analytic cases. This became the basis for a strategy to dramatically reduce the number of cases that would need to be manually evaluated each month. From our analysis we created a classification algorithm for potential new cases, and a system that automatically runs the algorithm over multiple data sources and provides a user interface to support the workflow of the registry nurses for new case detection. The system was implemented in December 2012.

We evaluated the impact of the new system in several ways. First, we evaluated the completeness of the system for potential new case detection to ensure that no cases were missed. Second, we evaluated the impact of the system on staff efficiency by measuring the change in the number of potential new cases that needed to be manually assessed. To do this, we compared the number of cases that needed to be manually assessed before and after the tool was implemented.

**Results**

**Algorithm for Detection of Potential New Reportable Cases**

Figure 2 illustrates the algorithm for classification of potential new analytic cases. It assumes the existence of a registry database that contains patients who were previously found to be analytic cases at time point t1, as well as a second database that contains information about cases that had been previously screened for inclusion but were not found to be reportable cases at time point t1. The algorithm takes as input a new case with an eligible code (billing or pathology) at time point t2 (where t2 > t1) and assesses if the record is: A) a case that has never been previously assessed, B) a case that has been previously assessed but not found to be a reportable case, or C) a case that has been previously assessed and found to be a reportable case. Cases categorized as A or B would be further assessed as potential new reportable cases by the registry nurses whereas cases categorized as C would be assessed for potential secondary cancers.

Cancer patients often have multiple clinical encounters over several months to several years, generating numerous billing codes that would fall into category C. In the old, inefficient process, these cases would need to be evaluated as potential new cases each time the patient had a new eligible code. Since the registry software system already contained a process for triggering yearly case review to assess vital status and presence of any secondary cancers,
Time point t1 occurs before time point t2. A patient medical record number (MRN) with a reportable code at time point t2 is being assessed as a potential new analytic case. The algorithm classifies the case as: A) an MRN that has not previously been evaluated, B) an MRN that has previously been evaluated at time point t1 and determined NOT to be a reportable case, or C) an MRN that was previously evaluated at time point t1 and found to be a reportable case. Category A and B cases are assessed for potential new reportable primary cancer diagnosis at time point t2. Category C cases are assessed for potential secondary reportable cancer case.

Figure 2. Classification of Potential New Analytic Cases

Figure 3. New Casefinding Process

Initial import eliminates records if they exist in the registry system, removes records that are nonreportable, and rejects records that were reviewed prior.

it was believed that cases falling into this category could be safely eliminated using the new case detection algorithm.

Likewise, the set of codes used to screen for potential new cases may identify patients who do not qualify as reportable cases at time point t1 (category B). These patients may likewise have multiple clinical encounters over time. However, it would be inefficient to assess these patients for conversion from a nonanalytic case to an analytic case at each encounter. Since the standard for reporting newly diagnosed cancer cases is within 6 months of diagnosis8 and the registry assesses cases approximately 6 months after the date of the clinical encounter, it was determined that category B cases only needed to be manually assessed once per year. All category A cases would be included for manual review.

System Description

The new approach consists of the following processes: structuring new queries to retrieve the appropriate number of results, building assessment tool Web pages to present the results for the users who will make the determination, recording the results in the database, and utilizing the recorded data into the continuous filtering process of subsequent datasets. Oracle was used to house the structured queries and to process imported files. Some of these queries are conducted at the time data is imported into the system, and others are a runtime process that is triggered by the users as they start the evaluation process. However, freely available open source options such as MYSQL could be used to replace Oracle. Other freely available open source tools utilized to build the tool were PHP programming language and Apache Tomcat Web server. All programming was performed by one of our in-house programmers in collaboration with one of the authors.

Within our Oracle database, we built a new infrastructure consisting of 5 tables: 1) master patient table to house records being imported from the 2 data sources, 2) patient history table for records being evaluated and their outcome, 3) reportable codes table provided by the tumor registry group, and 4) a replica of the tumor registry database to lower the performance impact on the tumor registry system.

The PHP Web interface consists of a simple on-demand tool interface with actionable buttons and 3 tabs: potential cases, second review, and an administrative dashboard for reports. The system utilizes 2 layers of security: network access managed by the university and application security managed on the application level. The application allows only users who have a network access and who are also designated and approved by the tumor registry manager.

Tumor registry staff were granted access to the tool to run their report based on a date range input parameter against the encounter date. The generated list is based on an optimized query that only selects and retrieves records whose encounter date falls within the date range and only single records are retrieved. However, the outcome of this query undergoes a number of refining phases prior to presenting the tumor registry nurse with the final list to evaluate (Figure 3).

First, any record that exists in the tumor registry table is eliminated as it is not a new case according to the category A definition provided earlier. Second, ICD-9 codes are evaluated for each record against a reportable codes table identified by tumor registry staff as valid reportable codes. Records whose ICD-9 codes are not present within the ICD-9 reportable codes table are then eliminated. Third, if the record has been evaluated before and has an occurrence in the patient history table, the record is then evaluated further for 2 criteria. The record is dropped if the difference between the first evaluation date and current visit date is less than a year. If the record was evaluated over a year ago, the ICD-9 is inspected if it is the same code that was entered in the history table. If so, the record is dropped. Otherwise, the record is left on the list for evaluation by the tumor registry nurses. The above refining processes are automated (Figure 3). Tumor Registry nurses are only presented with a refined list that includes last name, first name, middle initial, Social Security number, date of birth, medical record number, and actions (Figure 4). For each case presented, the tumor registry nurse does a preliminary evaluation of the record in the electronic health record and determines if to the case should be abstracted. If the case is determined to be an analytic case, it is marked as “TR Case” and included in downstream lists of cases to be further abstracted. If it is
The new system implemented in December 2012 while the tumor registry number of cases (Table 1 and Figure 5) found that the new process identified roughly the same cases recorded in the same month of the previous year and analytic cases. We also compared the number of analytic cases recorded in May 2011 and 2012. As shown in Table 1 and Figure 5, the total number of processed cases significantly reduced between the same time period in 2011 and 2012. For the month of May alone, the tumor registry reviewed 9,614 fewer records than were reviewed the same month of the year before, corresponding to a 78% reduction in cases needing evaluation. This corresponded to a major decrease in the number of cases each registry nurse had to evaluate, from 1,500 down to 350 per month. As such, the tumor registry nurses were able to finish initial review of the month of May 2012 within 3 days, a process that used to take 7 to 10 days. This time savings directly corresponds to cost savings, with an estimated $50,000 yearly savings in eliminating staff overtime alone due to the growing workload and concerns of timeliness and completeness of registry data.

**System Evaluation**

Evaluation of the completeness of the algorithm for detecting reportable cases. Since the function of the tumor registry is to report all analytic cases, we wanted to ensure that the new system did not eliminate any analytic cases for review. To validate the algorithm, we applied it to old finance input files from 3 months prior to the system implementation and confirmed that the algorithm correctly identified all of the cases that were manually classified as analytic cases. We also compared the number of analytic cases recorded in the same month of the previous year and found that the new process identified roughly the same number of cases (Table 1 and Figure 5).

Evaluation of improved efficiency. The system was implemented in December 2012 while the tumor registry staff was still evaluating potential cases from the billing reports of previous months. Table 1 and Figure 5 display a comparison of the total number of cases manually reviewed and those classified as analytic cases in the months of May, June, and July for the years 2011 and 2012. Over the 3 months that were assessed after deployment, there was an average of 31% fewer cases that needed to be manually assessed by registry staff.

The tool was first released to production in early December 2012. At that time, all tumor registry staff started to use the tool as their primary method of case investigation. Next, registry staff began evaluating data from May, June, and July of 2012. As shown in Table 1 and Figure 5, the total number of processed cases significantly reduced between the same time period in 2011 and 2012. For the month of May alone, the tumor registry reviewed 9,614 fewer records than were reviewed the same month of the year before, corresponding to a 78% reduction in cases needing evaluation. This corresponded to a major decrease in the number of cases each registry nurse had to evaluate, from 1,500 down to 350 per month. As such, the tumor registry nurses were able to finish initial review of the month of May 2012 within 3 days, a process that used to take 7 to 10 days. This time savings directly corresponds to cost savings, with an estimated $50,000 yearly savings in eliminating staff overtime alone due to the growing workload and concerns of timeliness and completeness of registry data.

**Discussion**

Different approaches have been reported for case identification that include natural language processing algorithms and commercially available artificial intelligence based products such as E-Path. Unlike the automated case-finding engine developed at the University of Michigan, our approach utilizes an automated self-enhanced algorithm for case identification and duplicates reduction that is based on time sensitivity, ICD-9 codes, and approved reportable codes. However, it is likely that such a tool has improved since its development. Other approaches for case identification were reported using Medicare claims and veterinary pathology reports.

Our novel approach to the case-finding process has several benefits and limitations. The primary benefit is that it directly supports one of the tumor registry’s primary missions of complete and timely case-finding and abstraction. We have demonstrated that the method improves the efficiency of case-finding by eliminating redundant work and streamlining access to enterprise data sources. We have also demonstrated that the method does not increase the number of missed cases. These new efficiencies allow the registry staff to complete their tasks in a timely manner at reduced cost to the institution.

In addition, the tool was specifically developed to support the workflow of a large tumor registry staff working in parallel to review cases. Deployed in December 2012, the tool continues to be used every day by the VUMC registry staff. Furthermore, the system has the potential to gain efficiency the longer it is in use. As more cases are flagged in the system as “non-analytic” cases, the pool of cases...
requiring monthly review will likely continue to diminish. This form of intelligent learning is a major contribution to efficiency gains of the system.

While we have only demonstrated the use and evaluation of this approach at a single institution, the principles and tools could be applied to other environments with appropriate technical support. Limitations include the lack of integration of other clinical data sources such as pathology reports. We are actively working to incorporate this particular data source in future iterations of the system.

Finally, the scope of this work is limited to refinement of the new case identification process. It does not currently support the data abstraction process, another area of great need that could have major impact on the completeness and timeliness of tumor registration. However, the efficiencies gained from the decreased load of cases to review allow the registry staff more time to abstract analytic cases.

Conclusion
We have developed an information system to improve the efficiency of new analytic case detection at the VUMC tumor registry. We have demonstrated that this approach decreases the number of cases to be reviewed each month without increasing the false negative rate. We believe this approach is potentially generalizable to other institutions beyond the specific implementation of the tools developed. While we have addressed an important problem of new case identification, gaps still remain in assisting tumor registry staff in case abstraction, a ripe area for additional informatics tool support.

Acknowledgement
The authors would like to thank the staff of the tumor registry office for their assistance and cooperation.

References
Original Article

Assessing the Quality of Race/Ethnicity, Tumor, and Breast Cancer Treatment Information in a Non-SEER State Registry

Abigail Silva, PhD, MPH; Garth H. Rauscher, PhD, MPH; Carol Estwing Ferrans, PhD, RN, FAAN; Kent Hoskins, MD; Ruta Rao, MD

Abstract: Introduction: The quality of population-based cancer registries are largely defined by the completeness, accuracy, and timeliness of incident cases and demographics reported. However, both Surveillance, Epidemiology, and End Results (SEER) cancer registries and non-SEER population-based state cancer registries have been regularly used to examine treatment patterns. While the quality of treatment data in SEER cancer registries has often been examined and improved, the quality of such data in non-SEER state registries has rarely been assessed. Methods: We used self-reported (SR) and medical record (MR) abstracted data from a population-based breast cancer study for comparison with information contained in the Illinois State Cancer Registry (registry). Using either MR or SR as the gold standard, we estimated concordance, kappa, and sensitivity for the presence or absence of surgery and initiation of chemotherapy, radiation and hormone therapy, as well as tumor characteristics, race/ethnicity and insurance status. Results: The accuracy of most of the data elements examined was generally high. For instance, there was almost perfect agreement between SR race/ethnicity and registry documentation (k = 0.92). MR and registry data on tumor stage, grade, ER/PR status, and node status had substantial agreement (k = 0.78–0.88). In regard to treatment information, surgery was rarely underdocumented in registry data, while radiation and chemotherapy were modestly underdocumented (8%–16%). On the other hand, per SR or MR, the registry generally failed to document hormonal treatment in a large proportion of cases (0.38 and 0.52, respectively). Health insurance information in the registry was also not well documented. There was only moderate agreement (k = 0.41) between SR and registry health insurance status, with uninsured patients being the least likely to be documented as such in the registry (sensitivity = 0.37 vs 0.96 and 0.63 for public and private insurance status, respectively). Discussion: While some registry data elements are quite reliable, others warrant concern and must be interpreted with great caution. Understanding the strengths and limitations of a population-based non-SEER state cancer registry data can be useful to researchers who use these data sources to examine population cancer patterns or carry out cancer studies.

Key words: breast cancer, cancer registry, data quality

Introduction

The quality of population-based cancer registries is largely defined by the completeness, accuracy, and timeliness of incident cases and demographics reported. However, both Surveillance, Epidemiology, and End Results (SEER) cancer registries and non-SEER population-based cancer registries have been regularly used to examine treatment patterns, such as in the case of breast cancer. While the quality of treatment data in SEER cancer registries has often been examined and is continuously being improved, the quality of such data in non-SEER state registries has rarely been assessed.

The most comprehensive evaluation of various cancer registry data elements (e.g., race/ethnicity, tumor stage, and insurance type) included both SEER and non-SEER registries. Unfortunately, the results were not stratified by SEER status and so the quality of data in non-SEER registries is not well known. The quality of the data between SEER and non-SEER registries may differ because they have different structures and funding sources. For example, SEER registries are likely better resourced as they receive funding from the National Cancer Institute and many also receive support from the Centers for Disease Control and Prevention. Furthermore, SEER sites are continuously involved in data quality assessments.

A better understanding of the quality and limitations of non-SEER cancer registry data elements is required in order to better interpret study findings. Self-reports (SR) and medical record (MR) data obtained from a recent population-based study of breast cancer patients allowed for the examination of race/ethnicity, health insurance, tumor, and treatment data found in a non-SEER state registry.
Methods

Patient Population

Interview and MR data were obtained from the Breast Cancer Care in Chicago (BCCC) study which aimed to examine the biological, social, environmental, and behavioral factors associated with the racial/ethnic (black and Hispanic vs white) breast cancer disparity in stage at diagnosis and prognosis. Briefly, non-Hispanic (nH) white, nH black, and Hispanic patients aged 30 to 79 years, who were diagnosed with first primary in situ or invasive breast cancer in Chicago between 2005 and 2008, were identified through monthly casefinding activities carried out across 56 hospitals in the Chicago region. The final interview response rate was 56% (proportion interviewed among total estimated eligible sample). The BCCC study sample included 989 women of which 397 were nH white, 411 were nH black, and 181 were Hispanic. Of the 989 women interviewed, 849 consented to an MR review, including linkage with the Illinois State Cancer Registry (registry). Only patients who provided MR consent and had single invasive primary tumors were included in this data quality study (N = 824).

Self-Reported and Medical Record Abstraction Data and Variables

Face-to-face interviews were conducted (in Spanish or English) a median of 3.5 months after diagnosis. The majority of patients (88%) had received surgery by the time of the interview. All participants answered questions on various topics including sociodemographics and cancer treatment. Patients also provided the name of the diagnosing and treatment facilities so that MRs could be abstracted.

Race/ethnicity, health insurance, and treatment information was obtained during the interview. A patient’s race/ethnicity was categorized into nH white, nH black, or Hispanic based on the following 2 multiple-choice questions: “Do you consider yourself to be of Hispanic or Latino origin? (Yes/No)” and “What race do you consider yourself to be (white, black, Asian, etc)?” Patients were also asked “What kind of insurance did you have at the time the problem was discovered that turned out to be breast cancer?” This was followed by a series of yes/no questions regarding private insurance (eg, Medicare with private supplement, HMO, PPO) and public insurance (eg, Medicare, Medicaid, military). All questions related to radiation, chemotherapy, and hormonal treatments followed the same general structure (Table 1). For example, regarding radiation, patients were asked the following yes/no questions: “Were you offered radiation therapy or was it suggested that you accept this treatment? If yes, have you agreed to have radiation therapy? If yes, have you begun radiation therapy yet?”

Certified tumor registrars employed by the registry reviewed pathology records, the hospital tumor registry, or both, depending on the protocol at the individual hospital. Using a structured data collection form, tumor and treatment information were abstracted from the MRs. Data regarding the initiation of radiation or chemotherapy were coded as “yes” whenever there was documentation of treatment receipt (eg, treatment start or stop date, treatment status). Hormonal treatment initiation was coded as “yes” if there was evidence of a treatment prescription (eg, name of drug, date of prescription). In the few cases where records showed different tumor grades or stage, the highest grade or stage was documented.

Registry Data and Variables

The state registry is the only population-based source for cancer incidence information in Illinois. Cancer cases are collected through mandated reporting by hospitals, ambulatory surgical treatment centers, non-hospital affiliated radiation therapy treatment centers, independent pathology labs, physicians and through the voluntary exchange of cancer patient data with 11 other states. Data is collected in accordance with the Illinois State Cancer Registry Standards for Data Reporting and Data Dictionary. However, data variables are set by national standard setters (eg, SEER, North American Association of Central Cancer Registries).

Table 1. Treatment-Related Questions from In-Person Interviews

<table>
<thead>
<tr>
<th>Radiation Treatment</th>
<th>(If yes) Have you agreed to have radiation therapy?</th>
<th>(If yes) Have you begun therapy yet?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemotherapy Treatment</td>
<td>(If yes) Have you agreed to have chemotherapy?</td>
<td>(If yes) Have you begun chemotherapy yet?</td>
</tr>
<tr>
<td>Hormone Treatment</td>
<td>(If yes) Have you agreed to have hormone therapy?</td>
<td>(If yes) Have you begun hormone therapy yet?</td>
</tr>
</tbody>
</table>
which govern the rules of data collection and the values for each data variable. The registry receives more than 90,000 cancer reports resulting in about 65,000 incident cases per year. Including the registry manager, there are currently 13 registry staff members representing a combined total of 187 years with the registry.

We provided the registry with the following information on the 849 consenting patients: first name, last name, middle name, birth date, race, address, ZIP code, home phone, cell phone, date of diagnosis, biopsy dates, International Classification of Diseases codes, and tumor site. During the early part of 2012, a deterministic match was carried out in 2 parts. First, a “perfect” match was one that matched on last name, first name, date of birth, and gender. Second, successive match passes were run on those cases that weren’t contained in the “perfect” match group. For each pass, registry staff reviewed whatever data elements weren’t matching. If they determined that the case was truly a match, then it was marked as such.

Of the 849 BCCC participants, a match was found for 846. Among those matches, 824 patients had a single primary tumor while 22 had multiple primary tumors. Race, Hispanic ethnicity, type of payer, tumor, and treatment information were obtained for the patients for whom a match was found. With respect to Hispanic ethnicity, the registry provided the Hispanic identification algorithm (NHIA) variable which was derived according to the North American Association of Central Cancer Registries. The NHIA uses information on race, maiden name, surname, sex, and place of birth to enhance the ethnic identification of the Hispanic population in the United States.29

### Statistical Analysis

First, differences in the distribution of sociodemographic characteristics by MR consent were assessed using Pearson’s chi-square test. Second, using either the MR or SR data as the gold standard, the sensitivity (ie, proportion of the gold standard correctly classified by the registry), percent agreement, and kappa statistic for race/ethnicity, insurance, and tumor information in the registry were estimated. For the treatment information, the false-negative rate (1-sensitivity) was calculated in an effort to quantify the level of potential underascertainment of treatment in the

---

**Table 2. Medical Consent Rate by Selected Characteristics**

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nH black</td>
<td>411</td>
<td>(86)</td>
</tr>
<tr>
<td>nH white</td>
<td>397</td>
<td>(85)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>181</td>
<td>(87)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50 years old</td>
<td>317</td>
<td>(86)</td>
</tr>
<tr>
<td>50–59 years old</td>
<td>308</td>
<td>(86)</td>
</tr>
<tr>
<td>≥ 60 years old</td>
<td>365</td>
<td>(86)</td>
</tr>
<tr>
<td>Primary language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>850</td>
<td>(86)</td>
</tr>
<tr>
<td>Non-English</td>
<td>123</td>
<td>(86)</td>
</tr>
<tr>
<td>Annual household income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ $30,000</td>
<td>365</td>
<td>(88)</td>
</tr>
<tr>
<td>&lt; $30,000</td>
<td>595</td>
<td>(86)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ High school</td>
<td>369</td>
<td>(90)</td>
</tr>
<tr>
<td>&gt; High school</td>
<td>620</td>
<td>(83)</td>
</tr>
<tr>
<td>Treatment facility type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCI/academic</td>
<td>656</td>
<td>(85)</td>
</tr>
<tr>
<td>Other</td>
<td>274</td>
<td>(86)</td>
</tr>
<tr>
<td>Time from diagnosis to interview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 90 days</td>
<td>323</td>
<td>(86)</td>
</tr>
<tr>
<td>≥ 90 days</td>
<td>666</td>
<td>(86)</td>
</tr>
<tr>
<td>Total</td>
<td>989</td>
<td>(87)</td>
</tr>
</tbody>
</table>

*aP < .10

**Table 3. Quality of Race/Ethnicity and Insurance Information in the Illinois State Cancer Registry**

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Sensitivity</th>
<th>Agreement</th>
<th>K</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nH black</td>
<td>347</td>
<td>0.98</td>
<td></td>
<td></td>
<td>(0.95–0.99)</td>
</tr>
<tr>
<td>nH white</td>
<td>325</td>
<td>0.95</td>
<td></td>
<td></td>
<td>(0.92–0.97)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>152</td>
<td>0.88</td>
<td></td>
<td></td>
<td>(0.81–0.92)</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>111</td>
<td>0.37</td>
<td></td>
<td></td>
<td>(0.28–0.48)</td>
</tr>
<tr>
<td>Public</td>
<td>138</td>
<td>0.96</td>
<td></td>
<td></td>
<td>(0.90–0.99)</td>
</tr>
<tr>
<td>Private</td>
<td>574</td>
<td>0.63</td>
<td></td>
<td></td>
<td>(0.59–0.68)</td>
</tr>
</tbody>
</table>

CI, confidence interval.

*bCount of cases per self reports.

*cSelf reports were used as the gold standard.

*dMissing responses were excluded from estimates.
### Table 4. Quality of Breast Cancer Tumor Characteristics in the Illinois State Cancer Registry

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Sensitivity&lt;sup&gt;abc&lt;/sup&gt;</th>
<th>95% CI</th>
<th>Agreement&lt;sup&gt;a&lt;/sup&gt;</th>
<th>95% CI</th>
<th>K&lt;sup&gt;c&lt;/sup&gt;</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>186</td>
<td>0.95</td>
<td>(0.90–0.97)</td>
<td>0.91</td>
<td>(0.89–0.93)</td>
<td>0.88</td>
<td>(0.85–0.91)</td>
</tr>
<tr>
<td>I</td>
<td>271</td>
<td>0.96</td>
<td>(0.92–0.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>237</td>
<td>0.88</td>
<td>(0.84–0.92)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>90</td>
<td>0.81</td>
<td>(0.72–0.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>14</td>
<td>0.71</td>
<td>(0.43–0.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>141</td>
<td>0.88</td>
<td>(0.82–0.93)</td>
<td>0.86</td>
<td>(0.83–0.88)</td>
<td>0.78</td>
<td>(0.74–0.82)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>290</td>
<td>0.91</td>
<td>(0.87–0.94)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>300</td>
<td>0.79</td>
<td>(0.74–0.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ER/PR status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>552</td>
<td>0.98</td>
<td>(0.97–0.99)</td>
<td>0.98</td>
<td>(0.97–0.99)</td>
<td>0.94</td>
<td>(0.91–0.97)</td>
</tr>
<tr>
<td>Negative</td>
<td>148</td>
<td>0.96</td>
<td>(0.92–0.99)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lymph node status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>194</td>
<td>0.94</td>
<td>(0.90–0.97)</td>
<td>0.96</td>
<td>(0.94–0.98)</td>
<td>0.92</td>
<td>(0.88–0.95)</td>
</tr>
<tr>
<td>Negative</td>
<td>367</td>
<td>0.96</td>
<td>(0.92–0.99)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI, confidence interval.  
<sup>a</sup>Count of cases per medical record abstractions.  
<sup>b</sup>Medical record abstractions were used as the gold standard.  
<sup>c</sup>Missing responses were excluded from estimates.

registry. For each measure, 95% confidence intervals (CIs) were computed. All analyses were performed with SAS version 9.2 (SAS Institute Inc).

### Results

Overall, 87% of the 989 study participants consented to having their tumor, diagnostic, and treatment information abstracted from their MRs, including the registry. There were no differences in medical consent rates by race/ethnicity, age, primary language, annual household income, type of treatment facility, and time from diagnosis to interview (Table 2). However, participants with a high school education were less likely to provide MR consent than participants with less education (90% vs 83%, respectively).

### Race/Ethnicity and Insurance Type

Table 3 illustrates that the race/ethnicity information in the cancer registry was very accurate for patients who reported to be nH black or nH white (sensitivity 0.97 and 0.98, respectively). However, Hispanic ethnicity was less likely to be correctly coded (0.88).

There was only moderate agreement between SRs and the registry in regard to health insurance information (k = 0.41). The registry had correct documentation for patients with public insurance (sensitivity = 0.96). However, there was a substantial lack of consistency when it came to patients who reported no insurance or private insurance. Only a minority of uninsured patients and slightly more than half of privately insured patients were correctly classified in the registry (sensitivity = 0.37 and 0.63, respectively).

In addition, among patients who reported to be uninsured at the time of diagnosis, the registry documented that 53% had public insurance (data not shown). Similarly, 36% of patients who reported having private insurance were coded as having public insurance (data not shown).

### Tumor Characteristics

There was substantial agreement between the registry and MRs in regard to tumor stage, grade, ER/PR-receptor status, and lymph node status (Table 4). Sensitivity was also generally high. Interestingly, sensitivity was lower for later stages and higher grade.

### Treatment

Table 5 shows that when surgery is reported or documented in the MR, it is virtually always documented in the registry. There was also a high level of agreement on surgery type (mastectomy or breast-conserving) between the registry and the MRs and SRs (k = 0.86–0.88). Concerning radiation and chemotherapy, the data show that the registry underdocumented treatment in a small proportion (0.09–0.16) of patients that reported treatment or had documentation of such in the MR. Per SR or MR, the registry generally failed to document hormonal treatment in a large proportion of cases (0.38 and 0.52, respectively).

### Discussion

This study examined the quality of several often used data elements of a non-SEER state cancer registry. The data quality of race/ethnicity, tumor, radiation, and
chemotherapy information was generally high. Conversely, there was evidence that hormonal treatment and insurance information in the registry was less adequate.

Our study results on race/ethnicity and tumor characteristics are largely consistent with that of other studies. For instance, our findings are similar to that reported by 3 other studies that compared race/ethnicity information between SEER and SRs.\textsuperscript{27,30,31} Those studies also concluded that while SEER adequately classified nH white and black patients, they tended to underclassify Hispanic patients. An evaluation of 7 state cancer registries also showed comparable results to ours in terms of tumor grade and ER/PR status but reported substantially lower accuracy and sensitivity for TNM stage.\textsuperscript{21}

Several studies have examined the quality of breast cancer treatment in population-based registries. Using MRs as the gold standard, other studies have also found that surgery was rarely underdocumented (0.02–0.05) while hormonal treatment was more underdocumented (0.40–0.64).\textsuperscript{16,18,21} However, these studies reported a higher level of underdocumentation for radiation (0.22–0.28) and chemotherapy (0.16–0.44) than we did (0.16 and 0.08, respectively). Several validation studies have provided support for the use of SRs to broadly measure the receipt of breast cancer treatment.\textsuperscript{32,36} Using SRs as the gold standard, we observed a similar pattern in treatment underdocumentation as when using MRs as the gold standard. A previous study which used SRs to evaluate radiation receipt in the Los Angeles and Detroit SEER registries found underdocumentation to be 0.32 and 0.11, respectively.\textsuperscript{17} Our findings on radiation mirrored Detroit’s.

Altogether, our results suggest that some population-based cancer registries may be a reasonable data source for general information on receipt of surgery, radiation and chemotherapy. However, the data should be used with caution as the quality may vary by jurisdiction.\textsuperscript{19} In terms of hormonal treatment initiation, most, if not all, cancer registries are an inadequate source for such information. Documentation of “prescribing” or “administering” hormonal treatment is particularly questionable given that this treatment is typically taken orally and outside the health care setting. At most, this variable should be interpreted as a recommendation for treatment. Even then, because hormonal treatment follows surgery and other adjuvant treatment, it may not be recommended until many months after diagnosis. This may contribute further to underreporting or underdocumentation.

The study findings on treatment information are not surprising given that cancer registries are not designed to collect validated treatment information and so their use as a gold standard should be questioned.\textsuperscript{1,2,17} In order to improve treatment information, SEER is linked to Medicare files.\textsuperscript{13,15} Other cancer registries have also attempted to improve their treatment information by linking the registry to other administrative data including electronic medical records, insurance claims, and Medicaid.\textsuperscript{37-39} In order to improve hormonal treatment information, it may prove most useful to make linkages with pharmacy data.\textsuperscript{40,41}

In terms of insurance, German et al reported moderate agreement (k = 0.41–0.48) between (SEER and non-SEER) state registries and MRs.\textsuperscript{21} They also showed that sensitivity was highest for public (0.73–0.74) and no health insurance (0.62–0.73) and lowest for private insurance (0.51–0.63). Using SRs as the comparison group and gold standard, we found a similar level of agreement with the registry and same general pattern for sensitivity with one exception. Only 37% of patients who reported to be uninsured at the time of breast cancer diagnosis were documented as such in the registry. In fact, most of these reportedly uninsured patients (53%) were subsequently classified as publicly insured in the registry. It is likely that patients who reported being uninsured at the time of their initial breast cancer discovery were in fact enrolled in Medicaid by the time their data were sent to the registry. Finally, the majority (68%) of patients who had missing insurance information in the registry actually reported having private insurance (data not shown). We found that the health insurance information in this non-SEER cancer registry was not well documented. Although many studies use health insurance information found in cancer registries,\textsuperscript{3,11,42-45} they should do so with caution or seek ways to improve its quality through linkage with other administrative data.\textsuperscript{14,37,46,47}

This study extends previous evaluation work on the quality of population-based cancer registry data in 2 important ways. First, unlike earlier studies that have solely examined SEER data\textsuperscript{13,15,19} or SEER and non-SEER data jointly,\textsuperscript{21,22} our study examines data quality in a non-SEER cancer registry exclusively. Non-SEER cancer registries represent approximately 70% of the US population and so the data quality of these registries must be better understood because they are often used to examine cancer treatment patterns.\textsuperscript{3,4,10,11,48,49} Second, with 1 exception,\textsuperscript{21} most evaluations have examined the quality of a small number of data elements such as race/ethnicity\textsuperscript{27,30,31} or treatment.\textsuperscript{16-19,23} In

**Table 5. Potential Level of Treatment Underdocumentation in the Illinois State Cancer Registry**

<table>
<thead>
<tr>
<th>Treatment Received</th>
<th>n</th>
<th>(1-Sensitivity)\textsuperscript{a}</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per medical record abstractions\textsuperscript{b}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>741</td>
<td>0.02</td>
<td>(0.01–0.04)</td>
</tr>
<tr>
<td>Radiation</td>
<td>353</td>
<td>0.16</td>
<td>(0.12–0.20)</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>236</td>
<td>0.08</td>
<td>(0.06–0.13)</td>
</tr>
<tr>
<td>Hormonal</td>
<td>33</td>
<td>0.52</td>
<td>(0.35–0.68)</td>
</tr>
<tr>
<td>Per self-report\textsuperscript{c}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>726</td>
<td>0.01</td>
<td>(0.01–0.02)</td>
</tr>
<tr>
<td>Radiation</td>
<td>246</td>
<td>0.09</td>
<td>(0.06–0.14)</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>304</td>
<td>0.09</td>
<td>(0.06–0.13)</td>
</tr>
<tr>
<td>Hormonal</td>
<td>177</td>
<td>0.38</td>
<td>(0.31–0.45)</td>
</tr>
</tbody>
</table>

CI, confidence interval.

\textsuperscript{a}Medical record abstractions were used as the gold standard.

\textsuperscript{b}Self-reports were used as the gold standard.

\textsuperscript{c}Missing responses were excluded from estimates.
this analysis, we assessed the quality of race/ethnicity, insurance status, tumor characteristics, and treatment.

There are limitations to this study that are worth noting. First, several potential sources of error may explain the discordance between the registry and the gold standard (SR or MR). For instance, disagreement with SR may occur as a result of unclear wording, misunderstanding of the interview questions, or inaccurate patient recall. However, all interview questions were tested and refined by cognitive interviewing techniques, which would decrease the likelihood of unclear wording or misunderstanding of the questions. Also, the interviews were conducted face-to-face by well-trained interviewers, which would make misunderstanding less likely in this study. Inconsistencies between the registry and MR may be due to erroneous information in the MR or incorrect data abstraction. To minimize the latter, MR abstractors used standardized forms to collect tumor and treatment information. However, the study results could be biased if some abstractors captured the data more accurately than others. Second, we evaluated a non-SEER registry using data from the third largest city in the United States. According to 2006–2010 state cancer registry data, 18% of breast cancer cases resided in Chicago which is home to the state’s only 2 NCI-designated cancer centers and approximately half of the academic comprehensive cancer programs. As such, our findings may not be generalizable to other non-SEER registries or less urban regions. Indeed, studies have documented variation in data quality across state registries. Moreover, patients from rural areas may have a different cancer profile than those from urban areas. Finally, given our interview response rate was 56% (proportion interviewed among total estimated eligible sample), selection bias cannot be ruled out.

Non-SEER state cancer registry data are often used to examine population cancer treatment patterns or to carry out cancer studies. This study contributes to our understanding of the accuracy of some of the most frequently used registry data elements. Such information can help researchers better plan their studies and interpret their findings if they can identify which data elements are reliable and which warrant concern.

Acknowledgements

The authors would like to thank Lori Koch and Dr. Tiefu Shen at the Illinois Department of Public Health’s State Cancer Registry (funded through a cooperative agreement with the Centers for Disease Control and Prevention) for providing the data and some technical assistance.

References


Economic Evaluation of Cancer Registration in Europe

R. Zanetti; L. Sacchetto; M. Calvia; A. Bordon; T. Hakulinen; A. Znaor; H. Møller; S. Siesling; H. Comber; A. Katalinic; S. Rosso; Eurocourse WP3 Working Group

Abstract: Background: Little has been reported on costs of cancer registration, and standard indicators have not yet been identified. This study investigated costs and outcomes of a sample of 18 European registries covering a population of 58.8 million inhabitants. Methods: Through a questionnaire, we asked registries for real cost data including personnel, information technology (IT), and infrastructure. Staff costs were grouped by professional position and by activity performed. As outcomes, besides the production of current data, we considered publications in peer-reviewed journals (last 5 years’ impact factor [IF]) and characteristics of registry websites. Results: In our sample, the average cost of cancer registration per inhabitant was €0.27 at purchasing power standard (PPS) (range €0.03–€0.97), while the mean cost per case registered was €50.71 PPS (range €6–€213). Personnel costs accounted for an average of 79% of total resources. Resources spent in routine activities (an average of 51%, range 28%–87%) were predominant with respect to those allocated to research, with a few exceptions. Website quality seemed to be independent of total registry budget. Conclusions: The variance in costs of cancer registration across Europe can be attributed mainly to the type of registry (whether national or regional), the size of the covered population, and the national economic profile, expressed as gross domestic product.

Key words: cancer registry, costs, economic evaluation

Introduction

Cancer registration started in the middle of the 1900s in the United States (Connecticut, 1935) and in Europe (Denmark, 1943). During the following years, it expanded in North America and in North Europe (where complete national coverage was reached in several countries), and spread throughout the rest of the world, including to several populations in developing countries. As a consequence, the first compendium of data on cancer impact worldwide harbored the title, *Cancer Incidence in Five Continents (C15C).*1,2

In the subsequent decades, the expansion has been remarkable, mainly in Asian countries. Today, more than 500 population-based cancer registries are active, covering a population of more than 1 billion people. Impact indicators such as incidence, mortality, survival, and prevalence have been the constant products regularly published. The contribution to etiological studies (cohort, case-control, and correlation design) prevailed in a first phase, while the contribution to clinical studies, health care planning, and evaluation was later expanded.
In the frame of the Eurocourse Project (www.eurocourse.org), we decided to perform an evaluation on costs and outputs in the European setting with the following aims: to obtain detailed and updated information on operational costs, to standardize them with respect to the economic level of different countries, to propose indicators of cost suitable for standardization, and to investigate possible indicators of quantitative outcomes.

**Materials and Methods**

We surveyed a sample of the population-based European cancer registries, focusing on the year 2010. Among the 206 members of the European Network of Cancer Registries (ENCR), we drew a convenience sample of 20 units, balancing type of coverage (national vs regional), population size, length of activity, and the economic conditions of the country of the registry.

We prepared a detailed questionnaire and interacted by email and phone with each registry in the sample for increasing the compliance. We gathered information on costs (human resources, administration, infrastructure, IT, and other activities) and outcomes (scientific publications, characteristics, and contents of registry websites). Human resources were indicated in full-time equivalent (FTE) units, divided into categories of homogeneous salaries: registrars, programmers, statisticians, medical doctors, epidemiologists, administrative personnel, and others. For each professional profile, we asked for the amount of time spent on different activities (data collection; data processing and analysis; management and administration; research; communication; and other) and their corresponding average gross salary. Besides salaries, duty travel expenses were also considered. For administration and infrastructure, items such as costs of office rental, maintenance, electricity, and stationery were considered. IT costs included leasing or purchasing hardware and software. Other activities were included such as communication, training, and fundraising.

On the side of outcomes, we considered peer-reviewed publications authored by a staff member, excluding working group contributions. Registries were ranked according to the total impact factor (IF) of the last 5 years. The elements for evaluating websites of registries were existence, accessibility, languages, data, periodicity of updating, interactivity, and availability of the electronic version of the printed reports. Each of those items was scored from 0 to 10. The total score was expressed in the same manner. Economical values, provided in euros or in other European currencies, were converted to euros at

<table>
<thead>
<tr>
<th>Table 1. Participating Registries (Pseudonymized by Flower Names)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Respondent/ included</strong></td>
</tr>
<tr>
<td>Broomland</td>
</tr>
<tr>
<td>Cyclamenland</td>
</tr>
<tr>
<td>Cypressland</td>
</tr>
<tr>
<td>Daisylan</td>
</tr>
<tr>
<td>Edelweissland</td>
</tr>
<tr>
<td>Palmland</td>
</tr>
<tr>
<td>Jasmineland</td>
</tr>
<tr>
<td>Irieland</td>
</tr>
<tr>
<td>Mintland</td>
</tr>
<tr>
<td>Oakland</td>
</tr>
<tr>
<td>Orchidland</td>
</tr>
<tr>
<td>Peonyland</td>
</tr>
<tr>
<td>Pineland</td>
</tr>
<tr>
<td>Primroseland</td>
</tr>
<tr>
<td>Poppyland</td>
</tr>
<tr>
<td>Roseland</td>
</tr>
<tr>
<td>Sunflowerland</td>
</tr>
<tr>
<td>Tulipland</td>
</tr>
<tr>
<td>Violetland</td>
</tr>
<tr>
<td>Willowland</td>
</tr>
</tbody>
</table>
purchasing power standard (PPS)\(^*\) for 27 European countries (EU27).\(^{9,10}\) In the following, values have been always expressed in € EU27 PPS, but shortly referred as the euro.

We calculated the following indicators:

- Cost per registered case (including only cases of malignant tumors)
- Cost per number of inhabitants observed

Costs for running the registry were evaluated according to productive factors: human resources (HR), infrastructure, IT, and disposable. Costs were also analyzed according to different phases of work such as data collection, data management, research, and communication (including providing and attending training courses). Due to some confidential aspects of budget data, results were presented anonymously (each registry was given the name of a plant or flower, assigned independently from local botanic).

**Results**

**Sample and Compliance**

Nineteen out of the 20 sampled population-based cancer registries completed the questionnaire. Among the respondents, 1 cancer registry was excluded because it was in a situation of organizational transition at the time focused on by the survey. The final sample consisted of 18 registries, 10 regional and 8 national, representing 14 countries with various economic and geographical characteristics, including western, eastern, northern and Mediterranean, and European Union members and nonmembers (Table 1).

The registries started their activities between 1950 and 2006. According to the size of population covered, we grouped cancer registries in 4 classes: less than 500,000; from 500,001 to 2 million; from 2,000,001 to 5 million; and more than 5 million of population observed. At least 1 regional and 1 national registry were included in each group (Table 1).

**Costs**

In our sample, the overall cost in 2010 was €15.6 million, with a population covered of 58.9 million inhabitants and 308,464 cancer cases registered in the year. Figure 1 shows the cost per inhabitant observed. Values ranged from €0.03 to €0.97; the average was €0.27, while the median was €0.26 (values weighted on population size). Figure 2 shows the cost per registered case with a range of €6 to €213, an average of €50.71, and a median of €54.67 (values weighted on number of cases registered).

The costs appeared to be strongly associated with the type of organization, with an average cost per inhabitant of €0.36 for regional and €0.21 for national registries. The costs also seemed to be related to population size—€0.49 per inhabitant for small registries with a population below 2 million; €0.24 per inhabitant for large registries (population greater than 2 million) (Table 2). The analysis of variance highlighted that the effects were statistically significant (\(P < .05\), while their interaction was not. Cost per inhabitant in small regional registries was 3 times that of the large national registries (Table 2).

Figure 3 presents the relationship between the national economic conditions (here summarized by the gross domestic product [GDP]) and the total cost per observed

---

*PPS is an artificial currency unit, used by Eurostat (the European statistical agency) to express national accounts aggregates adjusted for price level differences using Purchasing Power Parities (PPPs). In their turn, PPPs are currency converters and spatial price deflators for comparison among countries (in our case for the 2 groups, those in the Euro Area and those with national currency); they are mainly based on market prices of goods and services. As the PPS conversion little influences Euro–USD comparison, in practice the reader of the present paper can simply read PPS Euro as real Euro and convert it. Euro/USD rates for 2010 (arithmetic mean between exchange rates at 01.01.10 and 31.12.10) were as follows: 1 Euro = 1.3884 USD; 1 USD = 0.7213 Euro.
inhabitant. The curvilinear model ($R^2 = 0.78; P < .001$) showed that an increasing expenditure for cancer registries was associated with the economical wealth of the country. The same data could also be observed as 3 clusters, with a central cluster and 2 others at each of the extremes. At the lowest extreme, there were 3 points seemingly inelastic to the cost; at the highest, there were 2 points seemingly inelastic to the GDP. However, most of the observations fell in the central cluster with an important variation in costs for countries with similar GDP.
Table 3. Human Resources in Population-Based Cancer Registries in Europe

<table>
<thead>
<tr>
<th>Professional profiles of workers in population-based cancer registries in Europe</th>
<th>Registrar</th>
<th>Programmer</th>
<th>Statistician</th>
<th>Epidemiologist</th>
<th>Physician</th>
<th>Administration</th>
<th>Director</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. FTE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of total FTE units</td>
<td>47.70%</td>
<td>9.40%</td>
<td>5.60%</td>
<td>11.60%</td>
<td>2.70%</td>
<td>8.00%</td>
<td>7.00%</td>
<td>7.90%</td>
</tr>
<tr>
<td>b. Annual costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average unit cost</td>
<td>35,869</td>
<td>41,860</td>
<td>42,766</td>
<td>62,331</td>
<td>83,811</td>
<td>37,600</td>
<td>93,221</td>
<td>29,183</td>
</tr>
<tr>
<td>% of HR budget</td>
<td>38.20%</td>
<td>8.80%</td>
<td>5.30%</td>
<td>16.10%</td>
<td>5.10%</td>
<td>6.70%</td>
<td>14.70%</td>
<td>5.10%</td>
</tr>
<tr>
<td>c. Activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data collection</td>
<td>68.80%</td>
<td>16.60%</td>
<td>7.20%</td>
<td>0.00%</td>
<td>13.30%</td>
<td>0.60%</td>
<td>0.00%</td>
<td>17.90%</td>
</tr>
<tr>
<td>Data Processing and Analysis</td>
<td>30.80%</td>
<td>74.90%</td>
<td>27.40%</td>
<td>20.60%</td>
<td>43.20%</td>
<td>3.00%</td>
<td>16.90%</td>
<td>5.80%</td>
</tr>
<tr>
<td>Management and administration</td>
<td>0.00%</td>
<td>2.50%</td>
<td>3.30%</td>
<td>8.20%</td>
<td>15.50%</td>
<td>82.00%</td>
<td>32.20%</td>
<td>5.60%</td>
</tr>
<tr>
<td>Research</td>
<td>0.30%</td>
<td>6.00%</td>
<td>61.70%</td>
<td>69.70%</td>
<td>23.70%</td>
<td>0.40%</td>
<td>42.60%</td>
<td>61.50%</td>
</tr>
<tr>
<td>Communication</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.50%</td>
<td>1.50%</td>
<td>4.30%</td>
<td>14.00%</td>
<td>6.80%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Other</td>
<td>0.10%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>1.50%</td>
<td>9.20%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

a. Percentage of total FTE units by main professional profiles. b. Average unit cost (€ PPS) by main professional profiles and their percentage by human resources budget. c. Percentage of time devoted to different tasks by main professional profiles.

Figure 9. Total Costs for 2010 (Expressed in Euro PPS EU27) vs Average Impact Factor for the Last 5 Years (Size of the Registry as Dimension of the Circle)

Figure 10. Total Costs in 2010 (Expressed in Euro PPS EU27) vs Website Score

Figure 4 shows the relationship between length of activity and cost per inhabitant. Data points are scattered and no trend can be highlighted ($R^2 = 0.31; P = .016$).

Analysis of cost by productive factors indicated that almost all the registries were labor-intensive. On average, 79% of resources were spent on personnel, 10% on infrastructure and connected utilities, 6% on IT, and only 5% on materials (for example, printing books or bulletins and organization of events) (Figure 5).

Most of staff resources were devoted to data collection (24% of total budget), data management/analysis (21%), and research (21%), while management and administration, communication, and other activities received only 13% of total resources (Figure 6).

Tables 3a and 3b report the distribution of human resources according to their functions, both in terms of percentage of FTE units and salaries: 57% of staff was represented by registrars and IT technicians, 27% by scientific tasks, and 16% by administration and services.

Table 3c cross-analyzes professional profiles and activities, showing a clear correspondence between skills and main tasks performed, but also showing almost every professional profile being involved in the full process of data production and results dissemination.

With respect to the main routine product (incidence cases registered), it resulted that, on average, it took 1 FTE personnel unit to register 1,120 cases per year.

Figure 7 shows for each registry the total costs per inhabitant served and the weighted average, divided into detailed items of cost. In Figure 8, the same data, presented...
as percentages of the total, were grouped into broad items of cost.

Outcomes

With respect to peer-reviewed publications, Figure 9 shows the association between total costs and the average IF for the last 5 years, taking into account also the size of the population covered. Most of the registries fell in the bottom of the IF scale with just 4 having middle or high scores for IF. There was no evidence of a clear trend in total costs, nor in the size of the registry represented by the dimension of the circles. However, the majority of the registries presented moderate levels of IF.

Finally, with respect to websites (Figure 10), registries seemed to cluster in 3 groups. A good deal of registries with limited budget (under €800,000) obtained a good score, satisfying the majority of the criteria adopted (existence of a website, accessibility, languages, and availability of updated data). A second group scored low due to the absence of a dedicated registry website. A third group reached maximal scores with apparently no direct quantitative dependence with their level of budget. On the whole, more than two-thirds of the registries obtained a high or very high score.

Discussion

Our study represents the first effort in quantifying the costs of cancer registration in Europe in relation to various indicators of activities, performances, and expenditures. Unlike the American study that examined different registries pertaining to the same common program, in this study, we had to deal with different countries, currencies, administrative organizations, and historical, clinical, and administrative backgrounds. Direct comparison in the analyses of detailed items of cost seemed difficult to establish. Nevertheless, it was in some way reassuring that comparable macro results have been found by both the American study and the European study. In the 5 registries considered by the American analysis, the cost per registered case ranged from approximately $30 ($22.6) to just over $100 ($75.2), with an average of $54.80 ($41). Our corresponding results ranged from $6 to $213 with an average of €51. The range in European data was wider, but the average value was comparable. Both in the United States and in Europe, the first item of cost was personnel, in almost identical proportions (79% in both the United States and Europe).

Costs per unit showed a similar pattern when considered per inhabitant or per registered case. The main factor influencing the cost appeared to be the type of organization, whether regional or national, with the national scheme appearing to perform better at each level of size. The relationship between cost per inhabitant and the GDP per capita of the country was direct and strong, when considered on the whole range of GDPs. Focusing on the middle of the GDP scale (between €25,000 and €30,000 in 2010) the huge variation in cost was not easy to explain. At the 2 extremes, as one could expect, wealthy countries spent much more than poor ones; however, from our data, it was not clear why the cost per inhabitant varied of nearly 1 order of magnitude for registries at the middle of the GDP scale (Figure 3). Length of operation did not seem to affect the costs (Figure 4).

Most of resources were spent on personnel. Staff worked primarily in collecting, quality controlling, and analyzing data, followed secondly by research. This appeared coherent with the fact that cancer registration was a labor-intensive activity with relatively low technologies, and that routine data collection and analysis came first in the development of a registry. Personnel skills and tasks appeared quite well-defined, but with some overlapping. This seems natural for small organizations such as registries, but is also a testament to good multidisciplinary efforts and integration of the work.

For all examined registries, the majority of the resources went to routine production of basic indicators and their publication and disseminations. The quantity of resources dedicated to research varied more among registries than the quantity devoted to the routine of data collection and producing routine statistics on trends. Comparing absolute and percentage figures, it can be seen that registries with an intense research activity do not necessarily expand their total budget accordingly, but, in some cases, they seem to finance research through a proportional reduction of routine costs, probably obtained by better management.

As outcome, we considered the IF of the last 5 years of publications and the registry websites. IF was high for a minority of registries, and this seemed more associated with expenditure in research than with total expenditure. It should be taken into account that routine data were mainly published in reports and compendia, and not in articles in peer-reviewed journals. Moreover, this survey was not able to measure the scientific production and publications that were based on registry data but not authored by registry staff. This point should be considered in future development of this kind of analyses.

With respect to the scoring of websites as an indicator of outcome, results showed that most registries had well-developed and efficient websites, irrespective of budget. This proved the cleverness of registries with fewer resources, or at least demonstrated that new technology allows the creation of good websites at low expenditure.

This economic evaluation was urged for at least 2 reasons. First, the extension of coverage recommended by international agencies, patient groups, and professional societies, as well as by national and regional governments, requires a revision of costs and considerations on the most effective organization model. Second, on the side of benefits, it is becoming clearer that the quantity and quality of the scientific production of a cancer registry is not independent from its size, and that there is probably a threshold in size for balancing efficiency and scientific excellence.

The strengths of our study are represented by the balance of the sample, the high response rate, and the efforts in devising comparable economic indicators. Our sample can be considered representative of different characteristics of population-based cancer registries around Europe, including geographical location, size, length of operation, and economic level of the region/country. Furthermore, our active follow-up of sampled registries by email and phone
granted a high response rate. The information gathered was very detailed and its accuracy has been checked against information from other sources.

The economic amounts were made comparable across countries through a well established algorithm of standardization, the purchasing power parities (PPPs).9-10

The proposed indicators (cost per inhabitant covered and cost per new registered case) were simple and will be easy to replicate in further studies. The cost per inhabitant allowed estimating the total cost needed for expanding registration to uncovered areas and for realizing a complete continental coverage. Moreover, it is directly comparable to any other public service provided to the population. The cost per unit of cancer registered was directly comparable with other costs of treating a case of cancer, and so could be expressed as a fraction of the total cost of a cancer case. However, it is necessary to notice that this indicator was calculated considering only malignant cases in the denominator, despite the fact that several registries register considerable amounts of benign cases, from around 3% to 25%. Although this influences the cost per case, we preferred building our indicator on the basic common functions of measuring malignancies.

Among the possible weaknesses, this survey referred to 2010, after which the economic cycle had its effects, and the economic features of some of the examined countries could have been altered by the process of entering in the Euro area in the period focused by the survey. The survey was not able to investigate in detail the degree of automation in the registration procedures and its effect on costs. Most of the registries were not autonomous institutions and administrative entities, but were part of larger institutions or administrations; this implies that they did not have a formal budget and that we had established their total budget by summing up single items. This made it more likely to overlook single items of costs and so more likely to underestimate than overestimate the total costs (in particular, writing off costs of resources could have been omitted).

In our opinion, these preliminary results deserve to be further investigated on a larger set of registries or even on the totality of European registries, and also in the prospective of monitoring costs in parallel with the United States and other emerging experiences in different continents.12

Overall, this study highlights variations in cancer registration costs and outcomes across Europe, and provides a basis for estimating the cost of extending the coverage of registration.

Acknowledgments

We are grateful to the staff of the cancer registries who accorded us their cooperation. Special thanks go to the EUROCHIP group, and to Eva Stelianova-Foucher and Mark O’Callaghan at the International Agency for Research on Cancer (Lyon) who helped in the conduction of the survey.

References

Abstract: Certified tumor registrars (CTRs) are expected to have expertise in cancer staging, treatment, and patient follow-up and an overall knowledge of the cancer disease process. As medicine becomes more personalized, the prognosis for individual cancer patients is beginning to include more molecular markers, and CTRs are being asked to record these results along with traditional anatomic information about the disease. Molecular markers, also called biomarkers, are measured using a variety of techniques, including fluorescent in situ hybridization (FISH), immunohistochemistry (IHC), enzyme-linked immunosorbent assay (ELISA), and reverse transcription polymerase chain reaction (RT-PCR). This primer will provide an overview of these techniques so that CTRs can more efficiently search medical records for information and more accurately record these data items into abstracting templates.

Key words: enzyme-linked immunosorbent assay, flow cytometry, fluorescent in situ hybridization, immunohistochemistry, reverse transcription polymerase chain reaction

Introduction

Currently, a cancer registry collects multiple types of site-specific factors, some specifically coding for molecular markers. Certified tumor registrars (CTRs) must therefore be familiar with a variety of molecular markers as well as the traditional measures of anatomical spread. These trends are likely to increase as more predictive markers are discovered and validated. Therefore, a review of the most common laboratory techniques used to measure and describe cancer markers is a good reference for the CTR’s library. We intend for this primer to be such a resource for CTRs.

The focus of this paper will be on describing the techniques of fluorescent in situ hybridization (FISH), immunohistochemistry (IHC), flow cytometry, enzyme-linked immunosorbent assay (ELISA), and reverse transcription polymerase chain reaction (RT-PCR). As each is covered, examples of currently coded site-specific factors will be referenced. As cancer surveillance standard setters are phasing out Collaborative Staging, the American Joint Committee on Cancer (AJCC) language of “nonanatomical factors” may supersede the “site-specific factor” phrase. However, the information presented here will be relevant to any use of FISH, IHC, flow cytometry, ELISA, or RT-PCR, regardless of which molecular markers are being tracked at any time or which staging system is in use. These techniques have been standardized over decades of use and function both as tools for discovery and tools for standard measurement of patient samples.

A brief overview of the fundamental challenges of molecular biology is needed before moving into specific techniques. As the name implies, molecular biology is the study of the very small parts of individual cells. This includes the study of cellular deoxyribonucleic acid (DNA), for instance, or any one of the many proteins which make up a cell. While we can see a cell and even its parts, such as the nucleus and mitochondria, under a microscope, we cannot see either individual proteins or the discrete nucleic acids which form DNA and define genes. We know, largely through indirect evidence, that DNA contains the set of instructions for how to make each part of a cell. Virtually every other part of the cell is protein, for which the DNA holds the building instructions. There are exceptions to this, such as functional ribonucleic acid (RNA) and lipids, but we will focus on the basic pathway of DNA instructions being used to build proteins (Figure 1). It is critical to understand this fundamental process in cancer cell biology, although it cannot be directly visualized. Instead, we rely on the techniques referenced above. With the exception of RT-PCR, which will be discussed separately, the techniques share a conceptual basis. In each case, the cellular target of interest (ie, tumor marker) is labeled with an engineered probe which is specific to the target. This probe must be specific to the target, must tightly bind to the target, and must be capable of taking up a label of some kind which can be observed by current technology. Luckily, the probes, which are developed in laboratories, can be modified and optimized to do these jobs.

This approach depends on researchers discovering the right markers to measure in cancer cells—markers that will be meaningful and predictive. This work involves using many of the methods which will be discussed here as diagnostic techniques. In both research and clinical practice, these techniques allow us to visualize and measure a specific, otherwise invisible, target against the background of many other invisible cellular components. Once a target is identified, an appropriate probe and label system allow us to gather information about it. This concept will be referenced repeatedly and is important to make clear. The target is a marker created by the cancer cell, which indicates
something about its state. The **probe** is a tool employed to selectively bind only to the target. The **label** attaches only to the probe and can be measured. Some labels are fluorescent, some turn a visible color, but all exist to allow us to observe a target. Our discussion of how to use probes to quantify molecular markers begins with the DNA itself and the technique known as FISH.

### Fluorescent In Situ Hybridization

Normally, each mammalian cell has 1 set of genes inherited from the mother and 1 set from the father. However, in cancer cells, the process of DNA replication during cell division is disorganized, resulting in a variety of mistakes. One such mistake involves an area of DNA being replicated over and over, leaving the cells with too many copies of a gene residing in the affected segment of DNA. The result of having too many gene copies, and therefore too many copies of the instructions for making one protein. Each mRNA is processed by specialized machinery in the cell to produce a single protein. These transformational steps, from DNA to mRNA to protein, are indicated with full arrows. Once a protein is manufactured, it may reside inside the cell, be embedded in the cell membrane, or may be secreted outside of the cell. These trafficking possibilities are indicated with dashed arrows. Proteins which are retained by the cell can be evaluated using IHC or flow cytometry techniques. Proteins which leave the cell and enter circulation can be monitored by taking a blood sample and performing an ELISA test.

In the case of FISH, the target we are attempting to visualize is the DNA. The probe is also made of DNA which binds to the target DNA. This binding would be invisible to us except that the probe is attached to a molecule which emits fluorescent light when viewed under the correct type of microscope. Our eyes can then register the number of times this light signal accumulates inside a single cell. Generally, 2 signals per cell indicate a healthy state. The signal should accumulate both at the place in the cell where the maternally inherited copy of the gene resides and at the place where the paternally inherited copy of the gene resides: 2 spots of light. A cell which accumulates many bright spots has excessive copies of the gene of interest, indicating a potentially tumorigenic change.

An excellent example of a disease where FISH is routinely used is breast cancer. The gene for HER2 is often overexpressed in this disease. The protein that the gene produces resides in the cell membrane and responds to a growth signal in the surrounding environment. When active, the HER2 protein begins the process of initiating cell growth or division. When many HER2 proteins are produced, more growth factors are sensed and more cell growth is initiated. Therefore, HER2-positive cancers tend to have more aggressive growth and a poorer prognosis, making FISH an important tool for understanding the disease.

### Immunohistochemistry

IHC is a method for visualizing changes in the expression of proteins. In this case, the protein is the target, unlike the FISH technique in which DNA is the target. The probe is an antibody which can recognize and bind to the chosen protein target. This antibody will attach to the target so strongly that a series of treatments to visualize the antibody will not affect the binding. In order to make the visualization function, the antibody is labeled with a series of reagents which result in a colorimetric response. In other words, where the antibody has attached to its target, the sample will show up as colored, typically brown. Again, the HER2 protein is an example of one application of this technique. Tumors with a high HER2 IHC reactivity—those which turn brown over much of their area—are likely to be more aggressive than tumors with little change in color. These changes are read by pathologists and can be found in pathology reports on biopsies and resections.

One question that may arise after gaining an understanding of FISH and IHC is why one test would be preferable to the other. If both tests can give information about the prognostic factor HER2, does it matter which test is used? The answer sometimes depends on the gene being measured. Not all proteins accumulate in the cell due to duplications of their genes. Some proteins may become resistant to degradation and accumulate for that reason. Another possibility is the activation of a gene which is not normally turned on, so that any resulting protein is pathological. The best example of this phenomenon is the site-specific factor Ki67. The Ki67 protein is only expressed in cells that are multiplying, so the appearance of Ki67 indicates that a growth program has been turned on. This
simple marker of mitosis must be measured at the protein level, as the difference between normal cells and malignant cells is the activation of the gene, rather than the amplification of the gene. HER2 DNA duplication is known to play a role in protein overexpression, so either test (FISH or IHC) can be used to identify overexpression, depending on the available facilities for sample characterization. This nuance helps to demonstrate how important it is that site-specific factors are not only identified but that their mechanisms of dysregulation are understood.

Flow Cytometry
Flow cytometry measures specific protein expression using the same principles of target-probe-label that IHC uses. However, it is tailored for cells that are in suspension, such as cells in a blood sample. One of the challenges of hematological malignancies is making a specific diagnosis. Blood is composed of a varied and complex array of cell types and determining which particular cell type has become cancerous is a primary goal. Fortunately, each cell type produces proteins which are specific to the function of that particular cell type. These proteins can serve as the targets for flow cytometry. A blood sample can be probed with a labeled antibody and the relative strength of the signal indicates the concentration of that cell type in the overall population. This measurement is extremely exact because the method of flow cytometry is to feed each individual cell through a reader to collect information on the strength of the label in that one cell. Thousands of cells are typically read to get a precise picture of how common a given target is in a patient sample. This process can be repeated until a target is shown to be amplified well above normal levels. Once a target has been identified as being overabundant, it can be matched to the cell type which produces that specific target and a specific diagnosis can be made. Toward this end, many membrane-bound proteins are examined in samples from patients with new hematological malignancies. For example, CD38 is the name of a marker for plasma cells, which are an antibody-producing subset of B cells. When plasma cells become cancerous, more CD38-bearing cells are present in the blood and the diagnosis of multiple myeloma is made.

Enzyme-Linked Immunosorbent Assay
The ELISA test depends on antibodies, much like the IHC and flow cytometry techniques. However, this test has a special application which differentiates it from the other techniques described above. For all other measures of biomarkers, a sample of the malignant tissue is required; the correct probe for the chosen target is applied to the sample and then measured. This yields information about the carcinogenic changes of the tumor itself, an invaluable asset for physicians and researchers. However, getting a sample of a tumor is not a trivial task. Some tumors are widely dispersed or difficult or dangerous to surgically access. An ELISA test is an alternative to obtaining tumor tissue. When research has demonstrated that a certain type of cancer tends to overproduce a class of secreted proteins which leave the cancer cell and enter circulation, a simple blood sample can be tested for elevated levels of those proteins.

The classic example of ELISA testing is the prostate-specific antigen (PSA) test. PSA is a protein produced by cells of the prostate gland. Unlike Ki67 or membrane-bound markers on blood cells, PSA is not confined to the cell that made it. Instead, it originates in one cell type and then circulates in the blood. Therefore, the concentration of PSA in the blood is an indirect measure of the state of the prostate gland. Elevated levels suggest an increased activity of the prostate gland tissue, possibly from cancerous growth of the tissue.

A variety of cancer markers can be measured by ELISA, but all must share the characteristic of being released into the bloodstream. In ovarian cancer, a type of membrane-bound protein is often overexpressed and the outer most section is cleaved and released. This marker, called CA125, can then be measured in a blood sample.

To perform an ELISA test, the probe antibody is typically anchored to a solid surface, such as a small plastic well. The patient’s sample is then placed in the well so that the target proteins can bind to the probes. The unbound sample is then removed, leaving only the layer of probe and the layer of target protein from the patient. Finally, a suspension that contains a second set of probe antibodies—these labeled for measurement—is placed in the well and allowed to bind to the trapped target proteins. After this second round of binding, the suspension is removed, along with any unbound labeled probe. The remaining signal is read by a machine sensitive to the label, which is often colorimetric. This type of ELISA is called a “sandwich” because the patient’s target proteins are trapped between the first layer of probe and the second, labeled, layer of probe. The more patient protein is bound to the first probe layer, the more labeled probe is retained when it is incubated inside the testing well, and the more signal the machine reading the sample will record.

Reverse Transcription-Polymerase Chain Reaction
The RT-PCR technique enables the identification and measurement of a different type of change than the other methods discussed. For this method, cancer cells are required. Often, a tissue sample from a resection can be used or, in the case of hematological cancers, a blood sample is sufficient. From this sample, a laboratory technician can isolate strings of messenger RNA (mRNA), an essential intermediary between the DNA that codes for a protein and the protein itself. When a protein is needed, copies of the instructions for making it are generated from the DNA template in the nucleus and these copies, the mRNA, are transported to the protein-making machinery of the cell, located outside of the nucleus. This is a normal cellular process referred to as transcription and the mRNA should exactly replicate the DNA sequence, allowing us to analyze the mRNA strings to see what is happening with the DNA.

The RT part of RT-PCR stands for reverse transcription and simply indicates that the mRNA isolated from the sample is used to create, or transcribe, a DNA copy of the mRNA sequences. DNA is harder than mRNA and can better withstand the rest of the process. Why not
simply isolate the cellular DNA to start with? Because this procedure deals with tiny amounts of genetic material and there are many more copies of each gene’s information in the mRNA pool compared to the single, master DNA copy. In addition, these copies of the master instructions include only the abridged version of the original, omitting much of the DNA sequence which pertains to regulation of expression while retaining the blueprint of the protein. This shorter, more abundant source of transcripts is the raw material of the measurement. The rest of the RT-PCR process is replication of the templates, a necessary task to get enough material to analyze. PCR stands for polymerase chain reaction, the process of amplifying selected sequence of nucleic acid out of the total cellular supply. After RT-PCR, we are left with a large pool of our DNA sequence of interest. This material is then processed to yield the actual string of DNA components which make up the sequence. The well-known bases of DNA, the adenosine (A), thymine (T), cytosine (C), and guanine (G), can be analyzed by a computer program and compared to the normal sequence. Changes in the patient’s DNA indicate that the instructions for making its protein have been altered.

Changes to a DNA sequence are called mutations, and these can be major drivers of changes in cell behavior. Therefore, the RT-PCR technique provides a special kind of information about the sample that is tested. While other available techniques are measures of the amount of target, RT-PCR can identify changes to the nature of a target. This is an important distinction because some pathological proteins are not overexpressed; rather, they are faulty. The change in a mutated protein can be very small. Even a change to a single base pair (represented by the letters A, T, C, or G) in the instructional DNA sequence can make a functional difference in the gene product. An IHC test might not show any meaningful difference between healthy tissue and cancer tissue in this case, since the proteins in each could be too similar for a probe to distinguish between them, and they might also be produced in the same amount.

Two common examples of biomarkers that require RT-PCR testing are Janus Kinase 2 (JAK2) and the K-ras oncogene. In each case, the proteins are part of a signaling pathway that normally responds to cues from outside the cell. In the case of K-ras, the signaling pathway involved prompts the cell to grow and divide. This is an important ability for epithelial cells in particular because old cells are replaced by new cells as needed. The lining of the digestive system is a good example of an anatomic area that undergoes such an organized program of replenishment. If, however, a cell acquires a specific mutation in the K-ras gene, the resulting K-ras protein made from the faulty instructions will be locked in an “on” position and unable to respond appropriately to environmental cues. Instead, that cell will signal for growth all the time and a tumor will result. This type of mutation is commonly seen in colon cancer, which is why it is a tumor marker for that disease.

**Conclusion**

We have provided a brief introduction to the world of molecular biology, although in practice that world is more complex than the picture we have presented. Two basic principles are dependably true in this world, however. First, targets must be labeled or amplified in some way for the cellular information to be observable. In most cases, this means that a probe is used to mark the target and that probe is then labeled. Second, DNA targets require nucleic acid probes to selectively bind them and, likewise, protein targets require special proteins called antibodies to selectively bind them. This knowledge informs the type of approach that must be used to evaluate a given target. This overview includes the most common application of the techniques described and we hope that it will prove to be a useful foundation for CTRs.

As CTRs are asked to record more and more non-anatomical factors during the standard abstracting process, it will be more important to understand techniques such as FISH, IHC, ELISA, flow cytometry, and RT-PCR. These techniques yield predictive information about the cancers we document. Careful reporting on biomarkers enhances the role of CTRs in the research process as cancer databases become richer and more useful.

**References**

Raising the Bar: 5 Steps to Recover From a Mistake
Michele Webb, CTR

Have you ever said something or made a mistake at work that made you want to crawl into a hole or hide under your desk? Of course—we all have! Just in case the mere thought of a mistake makes you cringe, let’s consider a recent, well-publicized blunder.

If you watched Super Bowl XLVIII on February 2, 2014, you saw the Denver Broncos and Seattle Seahawks get into position for the first, all-important play of the game. Denver won the toss and Peyton Manning was stepping up to the line as he began calling the audible. Then, something happened. The ball was snapped and we saw it sail over Manning’s shoulder into the end zone. The Seahawks scored on a safety that set the tone for the rest of the game. In a post-game interview, Manning reported that the noise in MetLife Stadium was so loud that his team couldn’t hear him making the audible.1 To Manning’s credit, he went on to say that the mistake was not the fault of any one person on the team. It was unfortunate that the mistake was witnessed by 111.5 million viewers2 and will be talked about for decades to come.

For most of us, our mistakes are far less public or grandiose compared to what we saw on the Super Bowl. Nonetheless, any mistake made by a cancer registry professional has the potential to be far-reaching, and to cause embarrassment, loss of data integrity, personal credibility, or worse.

The media was ruthless in their criticism of Peyton Manning and we can only imagine how much worse it would have been if he had blamed his teammates for the error in order to protect his own reputation. While some mistakes may result in serious penalty or damage to our credibility, others may be less severe, but still painful to endure. Regardless of the magnitude of the mistake, it is our reaction to it and how we conduct ourselves after that will or will not set us apart as a health care professional.

Because we are human, we will inevitably mess things up at some point. The best recourse is to manage the problem as quickly as you can once you know the mistake has taken place. Here are 5 steps you can use to begin repairing the error:

1. **Admit your mistake.** Immediately talk to your boss and, with your emotions in control, state the facts as you know them to be. The only exception to this would be if you made an insignificant mistake that did not affect the work, process, data, or anyone in the workplace around you. Otherwise, hiding, staying silent about it, or trying to cover it up will only make things worse.

2. **Be prepared to offer a plan or solution to fix the mistake.** Take a few minutes before you talk with your boss to jot down your notes and best recommendations for resolving the problem. Present a clear, thought-out plan in short, succinct steps. Explain how long it will take to resolve the problem and what resources or financial considerations will be involved.

3. **Be accountable for your own actions—do not blame anyone else.** Remember the old adage that when we point a finger at someone else, 3 fingers are pointing back at you. If others were involved with the error, you can encourage them to follow suit, but you ultimately are responsible for your own actions. The worst thing a cancer registrar can do in this situation is to be apathetic, get defensive, or to convey through words or body language that they do not understand the impact of their error, or are not concerned about the possible effect it may have on the registry or organization.

4. **Apologize for your mistake and move on.** Be sincere without expecting anything in return. Do not continue to stomp around the problem or beat yourself up privately or publicly.

5. **Be willing to work overtime to correct your mistake.** Regardless of whether it takes a few minutes or a few hours, your action plan should include an offer to put in extra hours to fix the mistake. You can come early or stay late to get the job done.

These 5 steps seem fairly simple when we put them to paper. But it takes strong individuals who are passionate about their lives and work to admit that they were wrong and be accountable for their actions. Life is full of challenges that each of us must face, and accountability for ourselves provides many opportunities to learn and grow from the mistakes we make.

Maya Angelou said, “You may encounter many defeats, but you must not be defeated. In fact, it may be necessary to encounter the defeats, so you can know who you are, what you can rise from, how you can still come out of it.”3
The work of the cancer registrar is not easy and we are held to a very high standard. This is necessary if we are to deliver quality data and facilitate the outcomes that are needed to improve patient care. As you go about your work each day, may you have the courage and confidence to be accountable for your words and actions and be empowered to serve your communities and organizations with humility and excellence.

References

Journal of Registry Management Continuing Education Quiz—SPRING 2014

ASSESSING THE QUALITY OF RACE/ETHNICITY, TUMOR, AND BREAST CANCER TREATMENT INFORMATION IN A NON-SEER STATE REGISTRY

Quiz Instructions: The multiple choice or true/false quiz below is provided as an alternative method of earning CE credit hours. Refer to the article for the ONE best answer to each question. The questions are based solely on the content of the article. Answer the questions and send the original quiz answer sheet and fee to the NCRA Executive Office before the processing date listed on the answer sheet. Quizzes may not be retaken nor can NCRA staff respond to questions regarding answers. Allow 4–6 weeks for processing following the submission deadline to receive return notification of your completion of the CE process. The CE hour will be dated when it is submitted for grading; that date will determine the CE cycle year.

After reading this article and taking the quiz, the participants will be able to:
• Define the elements that contribute to quality data in a population-based cancer registry
• Compare and contrast the quality of data in Surveillance, Epidemiology, and End Results (SEER) cancer registries and non-SEER population-based state cancer registries
• Identify the cancer registry data elements that are most reliable

1. Which type of registry is more likely to receive funding from the National Cancer Institute and/or the Centers for Disease Control and Prevention?
   a) Commission on Cancer (CoC)–accredited hospital cancer registry
   b) Surveillance, Epidemiology, and End Results (SEER) cancer registry
   c) Non–CoC accredited registry
   d) Non-SEER registry

2. This study included patients from the Breast Cancer Care in Chicago (BCCC) study who:
   a) were over 30 and under 80 years old
   b) were diagnosed between 2009 and 2013
   c) refused medical record consent
   d) had recurrent breast cancer

3. Which of the following is included in the Hispanic identification algorithm (NHIA)?
   a) Religious affiliation
   b) Insurance status
   c) First name
   d) Maiden name

4. According to Table 2, Medical Consent Rate by Selected Characteristics, which characteristic was most likely to affect the medical record consent rate?
   a) Primary language
   b) Household income
   c) Education
   d) Ethnicity

5. According to Table 3, Quality of Race/Ethnicity and Insurance Information in the Illinois State Cancer Registry, the group of patients most likely to have insurance status correctly coded were those with:
   a) no insurance
   b) military insurance
   c) private insurance
   d) public insurance

6. According to Table 4, Quality of Breast Cancer Tumor Characteristics in the Illinois State Cancer Registry, sensitivity was lowest for which of the following variables?
   a) Later stage
   b) Lower grade
   c) ER/PR status positive
   d) Lymph node status negative

7. According to Table 5, Potential Level of Treatment Underdocumentation in the Illinois State Cancer Registry, which type of treatment was least likely to be documented?
   a) Chemotherapy
   b) Immunotherapy
   c) Hormone therapy
   d) Radiation therapy

8. Study results on race/ethnicity were consistent with other similar studies, showing underclassification for which group?
   a) Black
   b) Hispanic
   c) Asian
   d) Indian

9. According to the study authors, documentation of hormonal treatment is particularly questionable because it typically:
   a) takes place within the hospital setting
   b) takes place outside the hospital setting
   c) is recommended immediately following diagnosis
   d) is recommended in the neoadjuvant setting

10. According to the study authors, potential sources of errors were decreased through the use of:
    a) telephone interviews
    b) written interviews
    c) standardized collection forms
    d) nonstandardized collection forms

The JRM Quiz and answers are now available through NCRA’s Center for Cancer Registry Education (CCRE). For your convenience, the JRM article and quiz can be accessed online at www.CancerRegistryEducation.org/jrm-quizzes. Download the article, complete the quiz and claim CE credit all online.
Please print clearly in black ballpoint pen.

First Name

M.I.

Last Name

Address

Address

City

State/Province

ZIP Code/Postal Code

NCRA Membership Number (MUST complete if member fee submitted)

CTR Number

Instructions: Mark your answers clearly by filling in the correct answer, like this ■ not like this ✗. Passing score of 70% entitles one (1) CE clock hour per quiz.

Please use black ballpoint pen.

1

2

3

4

5

6

7

8

9

10

Submit the original quiz answer sheet only! No photocopies will be accepted.

This original quiz answer sheet will not be graded, no CE credit will be awarded, and the processing fee will be forfeited unless postmarked by:

June 30, 2015

Quiz Identification Number:

JRM Quiz Article:

ASSESSING THE QUALITY OF RACE/ETHNICITY, TUMOR, AND BREAST CANCER TREATMENT INFORMATION IN A NON-SEER STATE REGISTRY

Processing Fee: Member $25 Nonmember $35

Payment is due with submission of answer sheet. Make check or money order payable to NCRA. US currency only. Do not send cash. No refund under any circumstances. Please allow 4–6 weeks following the submission deadline for processing.

Please check one:

Enclosed is check #__________ (payable to NCRA)

Charge to the following card:

MasterCard (16 digits) Visa (13 or 16 digits) American Express

Card Number________________________________ Exp. Date _______

Signature ________________________________________________

Print Cardholder’s Name____________________________________

Telephone #______________________________________________

Mail to: NCRA Executive Office

JRM CE Quiz

1330 Braddock Place

#520

Alexandria, VA 22314

For Internal Use Only

Date Received: _______________

Amount Received: ____________

Notification Mailed: ___________

Available online at www.cancerregistryeducation.org/jrm-quizzes
National Cancer Registrars Association
CALL FOR PAPERS

Vicki G. Nelson, MPH, RHIT, CTR | EDITOR-IN-CHIEF, JRM

The Journal of Registry Management, official journal of the National Cancer Registrars Association (NCRA), announces a call for original manuscripts on registry methodology or research findings related to the 5 subjects listed below and related topics.

Topic:
1. Birth Defects Registries
2. Cancer Registries
   - Cancer Collaborative Stage
   - Cancer and Socioeconomic Status
   - History
3. Trauma Registries
4. Recruitment, Training, and Retention
5. Public Relations

Contributed manuscripts are peer-reviewed prior to publication.

Manuscripts of the following types may be submitted for publication:
1. Methodology Articles addressing topics of broad interest and appeal to the readership, including methodological aspects of registry organization and operation.
2. Research articles reporting findings of original, reviewed, data-based research.
3. Primers providing basic and comprehensive tutorials on relevant subjects.
4. “How I Do It” Articles describe tips, techniques, or procedures for an aspect of registry operations that the author does particularly well. The “How I Do It” feature in the Journal provides registrars with an informal forum for sharing strategies with colleagues in all types of registries.
5. Opinion papers/editorials including position papers, commentaries, essays, and interviews that analyze current or controversial issues and provide creative, reflective treatments of topics related to registry management.
6. Bibliographies which are specifically targeted and of significant interest will be considered.
7. Letters to the Editor are also invited.

Address all manuscripts to: Vicki G. Nelson, MPH, RHIT, CTR, Editor-in-Chief, Journal of Registry Management, (770) 488-6490, vnelson@cdc.gov.

Manuscript submission requirements are given in “Information for Authors” found on the inside back cover of each Journal and on the NCRA Web site at http://www.ncra-usa.org/jrm.
INFORMATION FOR AUTHORS

Manuscripts may be submitted for publication in the following categories: Articles addressing topics of broad interest and appeal to the readership, including Methodology papers about registry organization and operation; Research papers reporting findings of original, reviewed, data-based research; Primers providing tutorials on relevant subjects; and “How I Do It” papers are also solicited. Opinion papers/editorials including position papers, commentaries, and essays that analyze current or controversial issues and provide creative, reflective treatments of topics related to registry management; Letters to the Editor; and specifically-targeted Bibliographies of significant interest are invited.

The following guidelines are provided to assist prospective authors in preparing manuscripts for the Journal, and to facilitate technical processing of submissions. Failure to follow the guidelines may delay consideration of your manuscript. Authors who are unfamiliar with preparation and submission of manuscripts for publication are encouraged to contact the Editor for clarification or additional assistance.

Submission Requirements

Manuscripts. Articles should follow the standard format for research reporting (Introduction, Methods, Results, Discussion, References), and the submission instructions outlined above. The submission will normally include background information, and a rationale/justification as to why the subject matter is of interest. The discussion often includes a conclusion subsection. Comprehensive references are encouraged... as are appropriate combinations of tables and figures (graphs).

Manuscripts. Manuscripts should be double-spaced on white 8-1/2” x 11” paper, with margins of at least 1 inch. Use only letter-quality printers; poor quality copies will not be considered. Number the manuscript pages consecutively with the first title page as page one, followed by the abstract, text, references, and visuals. The accompanying cover letter should include the name, mailing address, e-mail address, telephone number of the corresponding author. For electronic submission, files should be IBM-compatible format in Corel WordPerfect™, Microsoft® Word for Windows®, or converted to ASCII code.

Manuscripts. Manucripts should be submitted through the Journal’s online submission system. Files should be submitted as PDF files only. Files should be IBM-compatible format in Corel WordPerfect™, Microsoft® Word for Windows®, or converted to ASCII code.

Bibliographies. Originality. Articles are reviewed for publication assuming that they have not been accepted or published previously and are not under simultaneous consideration for publication elsewhere. JRM reserves the right to edit all contributions for clarity and length. Minor changes (punctuation, spelling, grammar, syntax) will be made at the discretion of the editorial staff. Substantive changes will be verified with the author(s) prior to publication.

Peer Review. Manuscripts will be acknowledged promptly, and corresponding authors will be advised of the status of their submission as soon as possible.

Reprints. Authors receive 5 complimentary copies of the Journal in which their manuscript appears. Additional copies of reprints may be purchased from the NCRA Executive Office.
NCRA is Celebrating Its 40th Anniversary With a Special Gift to Our Members!

$100 Subscription to NCRA’s Center for Cancer Registry Education

The $100 subscription includes all Learning Modules, Journal of Registry Management quizzes, and the online immunity course. That’s over 40 CEs!


Get Started: Go to www.CancerRegistryEducation.org/NCRA-subscriptions

Questions: E-mail ccre@ncra-usa.org or call 703-299-6640 (ext. 317)

Special NCRA Anniversary Subscription to the Center for Cancer Registry Education Earn over 40 CEs for $100!
NCRA Study Guide for the CTR Exam:
2014 Edition
The new edition has been revised to reflect the updates to the 2014 exam. Five case studies have been added and answers with rationales provided. Companion CD contains additional study resources. NCRA member price: $85; non-member price: $105. Go to www.ncra-usa.org/CTRPrep to order.

Special Compilation of Collaborative Stage Data Collection System Coding Instructions (For 2014 CTR Exam Candidates Only)
NCRA has compiled the sections from the Collaborative Stage Data Collection System Coding Instructions (Version 02.04) that can be used in the open-book portion of the 2014 CTR Exam. This easy-to-use special edition includes Part 1: Sections 1&2 and the five specific sites: bladder, breast, colon, lung, and prostate. NCRA member price: $65; non-member price: $85. Go to www.ncra-usa.org/CTRPrep to order.

CTR Exam Preparation Online Workshop
NCRA has created a CTR Exam Prep Online Workshop consisting of nine, 90-minute live webinars on critical exam topics. Workshop member price: $365; non-member price: $485: Individual webinars can be purchased separately. Member price: $45; non-member price: $60. Go to the “CTR Prep” tab of the CCRE at www.CancerRegistryEducation.org for more detailed information on the online test.

Online CTR Exam Practice Test
NCRA offers a 125-question online CTR Exam Practice Test through its Center for Cancer Registry Education (CCRE). Designed to provide candidates with a tool to assess understanding and determine where further study is needed. Pricing options: One-time access: NCRA member - $65; non-member - $105. Unlimited access for 60 days: NCRA member - $180; non-member - $220. Go to the “CTR Prep” tab of the CCRE at www.CancerRegistryEducation.org for more detailed information on the online test.

2014 CTR Exam
The CTR Credential is nationally recognized in the recruitment and retention of cancer registry personnel.

TESTING DATES
June 21 - July 12, 2014
Application deadline has been extended to May 21.

October 18 - November 8, 2014
Application deadline: September 19

Learn more about the CTR Exam and download the 2014 CTR Exam Candidate’s Handbook at www.CTRExam.org.
Printed on SFI fiber sourcing paper with non-petroleum, vegetable based inks and manufactured with renewable electricity