

Using an epidemiological approach to examine outcomes affecting young children with Down syndrome and their families

Robert M. Hodapp^{1,2}, Richard C. Urbano^{1,3}, and Stephanie A. So^{1,3,4}

¹Vanderbilt Kennedy Center for Research on Human Development

Departments of Special Education², Pediatrics³, and Economics⁴, Vanderbilt University, Nashville, TN, USA

Abstract – In this paper, we utilise an approach drawn from the field of epidemiology to explore what is known and unknown about young children with Down syndrome and their families. After describing what we mean by an epidemiological approach, we review basic findings for children with intellectual disabilities, as well as challenges to performing such research. In considering the epidemiology of Down syndrome, we note that most studies to date have focused on prevalence, mortality-life expectancy, and rates of diseases and syndrome-related health-physical problems, while neglecting many other important issues. In considering potential advances in the epidemiology of Down syndrome, then, we first overview the process of linking two or more separate administrative records, before reviewing several of our own recent studies. We end this paper by discussing four challenges to future epidemiological studies of children with Down syndrome and their families.

Keywords: Down syndrome, epidemiology, population, health-related outcomes, case identification and definition, linkage

Within the world of disability research, Down syndrome is unique in that it is both known and unknown. On one hand, Down syndrome features a long, illustrious research history. From the late 19th and early 20th centuries, workers such as Langdon Down and Lionel Penrose studied physical, psychological, and other aspects of individuals with this syndrome. Almost 30 years ago, enough was known that David Gibson (1978) could devote one or more chapters apiece of his masterwork on the syndrome to early psychological development, intelligence, personality, socialisation, learning, speech-language, and behaviour management. Since the late 1970s, research has continued, with new emphases on how persons with the syndrome are affected throughout the life cycle. To this day, the syndrome has probably been the subject of more research than all other genetic intellectual disability conditions combined, with over 1,200 studies appearing on behavioural aspects of Down syndrome during the 1990s alone (Hodapp & Dykens, 2004).

As more is learned about the syndrome, however, large and important gaps remain. Consider the fundamental issue of how many persons have Down syndrome. Cur-

rently in the United States, approximately 350,000 people are thought to have Down syndrome. But this number is based on extrapolations, given estimates of the U.S. population (nearing 300 million), of incidence rates among all live births (1/800 to 1/1000), and of life-expectancy. Each of these estimates, in turn, depends on its own sampling design and set of assumptions. Other key demographic findings, especially those that depend on or interact with age, family, or health status, are harder to pin down. Even within the context of a 140-year research history, then, we continue to lack data about many of the basic demographic, health, family, and other characteristics that are associated with having Down syndrome.

This article presents one approach that may help to fill such gaps. This approach involves using large-scale, epidemiological databases to attain basic information about individuals with Down syndrome. First, we define what we mean by an epidemiological approach and discuss the strengths and challenges of using such an approach to examine both intellectual disabilities and Down syndrome. We then describe recent computer and methodological advances, before summarising a series of recent and ongoing studies that out-

line the possibilities of this method. In the final section, we reflect on epidemiological methods more generally and their potential for future advances.

Epidemiology

Before describing the main characteristics of epidemiology, it is instructive to recount the well-known story of John Snow and the Broad Street pump, arguably the beginning of the field itself. This story begins in the summer of 1854, when what Snow called “the most terrible outbreak of cholera which ever occurred in the kingdom” broke out in the Soho neighbourhood of London (Summers, 1989). Dr. John Snow, a successful surgeon and anaesthesiologist, had earlier published a paper speculating that cholera was transmitted through contaminated water, a view discounted by most within the medical profession. But in a painstaking study in which he interviewed the families of over 600 victims, Snow demonstrated that every death could be linked to drinking water from a popular water pump on the corner of Broad and Cambridge streets. Moreover, removing the handle from that water pump (along with the evacuation from the neighbourhood by many of the residents) resulted in the end of the epidemic (Cameron & Jones, 1983; Paneth, 2004).

Definitional issues

Given the story of John Snow and the Broad Street pump, the definitions provided in Table 1 nicely illustrate the main ideas in the field of epidemiology. Granted, the field has changed somewhat over the years, and epidemiology now focuses on both health and disease, and increasingly examines environments that are more social in nature (Susser & Susser, 1996). Still, the definitions provided in Table 1, along with the story of John Snow, highlight the following tenets of epidemiology.

1) Population focus. Epidemiology examines outcomes from a population perspective. In many epidemiological studies, geographic populations are the unit of interest, whether these involve populations of a country, state, city, or neighbourhood (e.g., Snow's study of Soho). But as epidemiology is concerned with the occurrence of illness in populations, the concept of population can be interpreted as any group at risk: females, children, or persons with Down syndrome. Because the concept of risk within a population is essential both in defining and interpreting epidemiological studies and results, enormous attention is paid to the gender, ethnic, racial, familial, socioeconomic, urban-suburban-rural, and other characteristics of the sample under study.

Contrast this population-based strategy to the approach usually adopted in psychological studies. In most such studies, researchers examine small numbers (20-40) of children or adults. Such small numbers of subjects, while acceptable for answering many types of ‘main effects’ questions, are usually inadequate to examine various interaction effects. Such studies also generally involve so-called samples-of-convenience, individuals who, on the basis of

Table 1. Definitions of epidemiology

“Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control health problems (Last, 1995).” (Yeargin-Allsopp & Boyle, 2002, p. 113).

“Mental disorder epidemiology is the quantitative study of the distribution and causes of mental disorder in human populations. This definition has several important components, including a population focus and a reliance on statistical methods to assess significant differences among population groups in their risk for developing mental disorders” (Reiger & Burke, 2000, p. 500).

“Epidemiology: The study of the distribution and size of disease problems in human populations, in particular to identify aetiological factors in the pathogenesis of disease, and to provide the data essential for the management, evaluation and planning of services for the prevention, control and treatment of disease” (Everitt, 1995, p. 88)

“Epidemiology is the fundamental medical science that focuses on the distribution and determinants of disease frequencies in human populations.” (Greenberg, Daniels, Flanders, Eley & Boring, 2001, p. 1).

“An emerging concept of epidemiology presents this discipline as the study of health and disease as a full spectrum across the human life span with a population approach, including etiological factors, phenomenology, comorbidities, and the uses and outcomes of clinical care.” (Mezzich & Ustin, 2005, p. 656).

advertisements or word of mouth, volunteer to participate in any given study. Of particular concern to epidemiologists (with their population-based focus) is whether these small samples truly represent the larger population in terms of their subject or family characteristics. If not, then epidemiologists cannot later infer from this sample the actual risk of occurrence of illness or other health-related events in the larger population.

2) Health-related outcomes. In contrast to psychological studies of stress, coping, or other more nebulous outcomes, most epidemiological studies examine the presence and predictors of more concrete events. Many epidemiological studies observe the occurrence of illnesses, deaths, hospitalisations, or other health-related outcomes. Recent studies also examine divorce, employment, and other outcomes that involve real-life events, albeit events that do not necessarily involve physical health.

Within this focus on health-related outcomes, researchers expend large amounts of energy and thought to providing exact definitions of a case. Exact definitions and diagnostic criteria are derived for various disease and other health-related outcomes. Snow, for example, was interested only in the cases of Soho deaths caused by cholera; deaths from other causes were outside of his scope of interest. More recently, researchers have worked hard to define low-birth-weight in newborns; disease states such as pneumonia, AIDS, or cancer; or conditions such as intellectual disabilities or autism. Although issues involving the consistent use and application of ‘case-ness’ arise in all epidemiological

studies, they are especially noteworthy when examining intellectual disabilities.

3) Focus on causes or probable causes. Epidemiological studies are designed to describe, explain, and predict the occurrences of the outcomes; the ultimate goal is to connect outcomes with their predictors. The epidemiologist, however, is not focused specifically on the outcome of a particular individual. Instead, epidemiologists think about outcomes in population terms: how to reduce risk across the population so that the proportion of cases in a population diminishes. Thus, predictors may include individual risk behaviours (e.g., riding in a car without seatbelts) or more broadly-based social risk factors (e.g., low socio-economic status, lack of access to health care). In all cases, epidemiology has as its goal connections between outcomes and predictors that provide clues as to possible cause(s) of the final outcome.

Such clues sometimes relate to already-specified disease mechanisms, sometimes to less well understood societal processes. In the example of the Broad Street pump, John Snow had specifically rejected the miasma view of disease – that disease was somehow ‘in the air’ – and was instead convinced that contaminated well-water was responsible for the outbreak of cholera. His subsequent study supported this view. In other instances, however, researchers begin with so-called ‘risk indicators’ and then attempt to proceed to the ‘risk mechanisms’ by which such indicators produce outcomes (Rutter, Pickles, Murray & Eaves, 2001).

A good example might involve one’s socioeconomic status, or SES. Many conditions and diseases occur more often in low SES individuals. Within this sample of low-income people, an epidemiologist would test hypotheses to identify which specific characteristics might indicate a direct pathway, or risk mechanism, for the outcome. For example, the epidemiologist might examine diet, environment, low education, or lack of health insurance for each factor’s relationship to the cause or progression of a disease. It is in these studies that the epidemiologist’s population-based data and toolkit of statistical techniques make up a useful – and different – approach from the small-scale studies.

4) Focus on intervention and public health. The Snow story ends with officials removing the handle from the Broad Street pump, thereby ensuring that no other Soho citizens would die from drinking contaminated water. In a similar way, epidemiology’s focus on determining amounts and correlates of diseases within populations has as its goal the prevention, amelioration, or treatment of those diseases.

When considered in this sense, epidemiology is one among many ‘mixed’ disciplines. Like the field of child development – which historically mixes basic research with applications of that research to children’s development (Sears, 1975) – so too is epidemiology both basic and applied. Epidemiology has a basic-science orientation in that its studies examine the prevalence and correlates of disease in different human populations. Its methods, which often lead the way among the biological and social sciences, involve compli-

cated statistical and methodological procedures to describe, explain, and predict health events. At the same time, however, the outcomes are inherently human. People die, get sick, are hospitalised or experience any number of events that affect their well-being. Epidemiological results often produce knowledge that reduces or prevents the numbers of people adversely affected. In epidemiology, as in other mixed disciplines, the distance between basic research and application of research results is very short indeed.

Epidemiology within developmental disabilities

Given this sense of epidemiology overall, it is instructive to examine the sub-field that has applied epidemiology to developmental disabilities. Over the past 10-20 years, numerous studies have examined the prevalence, distribution, and correlates of developmental disabilities. For the most part, these studies have been performed in Great Britain, the United States, the Scandinavian countries, and Australia and New Zealand. Most studies have considered intellectual disabilities as the outcome; correlates of intellectual disabilities (ID) have included the child’s gender, age, and race-ethnicity, as well as the parents’ educational levels and family socio-economic status (SES).

For example, many studies have now been performed concerning the prevalence rates of severe (usually, IQ < 50) and mild (IQ 50-69) intellectual disabilities (Leonard & Wen, 2002). For individuals with severe intellectual disabilities, levels mostly converge on 3-4 children per 1000; for mild ID, rates range wildly from 5.4 to 10.6 children per 1000 (see also Roeleveld, Zielhuis & Gabreels, 1997). Studies also examine such correlates as gender, age, and SES. More boys than girls have intellectual disabilities and rates of ID are generally low in the early years, peak at around 10-14 years, and decrease slightly in the late-school years and markedly during adulthood. Individuals of lower SES and of ethnic minority groups (in several cultures; e.g., Kearns, 2000) also show higher-than-expected rates of intellectual disability.

Although researchers justifiably highlight such findings (Yeargin-Alsopp & Boyle, 2002), the sub-field of epidemiology of intellectual disabilities also confronts several difficult problems. The most salient involves the definition of a case. As mentioned above, in order for epidemiologists to decide how many persons have a particular condition – or which child, parent, or family characteristics are associated with the condition’s occurrence – one needs an explicit definition of the outcome. But what constitutes intellectual disabilities?

Answering this question is more difficult than it seems. In the United States, the American Association on Mental Retardation (1992, 2002) has promulgated recent definitions of mental retardation (the American term for intellectual disabilities) that combine IQ levels (below 70 or 75), adaptive deficits in numerous areas, and onset during the childhood years. The American Psychiatric Association’s (2000) *Diagnostic and Statistical Manual, 4th edition Text*

Revised (DSM-IV-TR) also provides diagnostic criteria based on IQ, adaptive deficits (in slightly different areas), and childhood onset. The *International Statistical Classification of Diseases and Related Health Problems, 10th edition* (ICD-10; World Health Organisation, 1992) provides similar but not quite identical criteria. Given that IQ cut-off scores and criteria for adaptive deficits vary from definition to definition, wide discrepancies may exist in prevalence rates depending on which definition is used (MacMillan, Gresham & Siperstein, 1993). In reviewing this issue, Leonard and Wen (2002) concluded that "Taxonomy in this field is particularly difficult because professionals and consumers come from a range of backgrounds and have different purposes such as advocacy, education, medical care and service provision" (p. 120).

In epidemiological studies, intellectual disabilities have been diagnosed mostly using an 'IQ-only definition.' In most studies, all individuals who have IQs below 70 are considered to have an intellectual disability. In a well-known study, Yeargin-Allsopp, Murphy, Oakley and Sikes (1992) used only the IQ-criterion (i.e., IQ < 70) to identify those Atlanta schoolchildren considered to have intellectual disabilities. Were Yeargin-Allsopp et al., (1992) to have added adaptive deficits, fewer children would have been diagnosed, as children with IQs below 70 but who showed higher levels of adaptive behaviour would not have qualified for the ID diagnosis.

Another noteworthy issue involves whether those with intellectual disabilities are considered as the outcome or the population. In most existing epidemiological studies, intellectual disabilities constitute the outcome. Questions then centre on how many persons have intellectual disabilities, how many new cases occur each year, and what the prevalence rates are within different age groups, genders, ethnic, or minority groups. Increasingly, however, epidemiological researchers are turning the equation around. Instead of considering intellectual disabilities as the outcome, these researchers consider the group with intellectual disabilities as the population. They then examine other diseases as outcomes within the ID population.

Two examples illustrate this changed approach. First, over the past 10 years, various attempts have been made to examine the prevalence of psychiatric disorders within groups with intellectual disabilities (Einfeld & Tonge, 1996; Tonge & Einfeld, 2003). Granted, such studies are difficult to perform, as one again encounters the problem of deciding what constitutes a 'true case' of depression, schizophrenia, conduct disorders, or any other psychiatric condition among individuals with ID (see Dykens, 2000 for a discussion). Nevertheless, different studies find that large percentages of persons with intellectual disabilities – from 30% to 40% – also have significant emotional and behavioural problems.

A second example concerns specific diseases such as Alzheimer's disease. Partly as an outgrowth of studies of Alzheimer's disease in adults with Down syndrome, Zigman et al., (2004) have recently examined rates of dementia in

adults with intellectual disabilities who do not have Down syndrome. Their findings generally mirror those of Alzheimer's disease in the general population, with persons with (non-Down syndrome) intellectual disabilities showing similar rates of Alzheimer's disease during middle- and old-age periods. Lower levels of IQ, per se, do not seem to increase one's risk for Alzheimer's disease.

In considering epidemiological studies in intellectual disabilities, then, the field features many achievements and many challenges. Achievements involve the documentation of differences in prevalence rates across genders, age-periods, racial, and ethnic groups. Challenges mostly involve case identification and definition.

Epidemiology of Down syndrome

Down syndrome is a condition that occurs fairly frequently, is generally diagnosed at or shortly after birth, and involves a condition familiar to professionals and laypersons alike. Diagnostic criteria involve trisomy 21 and diagnostic tests involve karyotypes available in most hospitals. Most paediatricians and obstetricians have seen one or more newborns with the syndrome and are unlikely to overlook these children. In Down syndrome studies, then, problems of case definition and diagnosis seem less problematic than in epidemiological studies of persons with intellectual disabilities in general.

And yet, even given the major definitional advantages involved in epidemiological studies of Down syndrome, epidemiological research in the syndrome remains in its infancy. To date, only the following few restricted topics have been addressed.

1) Prevalence. Although Down syndrome has long been thought to occur once in every 800 to 1,000 births (National Down Syndrome Society, 2005), many studies examining these rates have appeared in recent years. Such studies have generally been aimed at understanding if rates have changed over the years, differ in different populations, or vary among women of different ages.

Although a complete review is beyond the scope of this article, prevalence rates of Down syndrome continue to hover in the 1/800 to 1/1000 range. Rates range from highs of 1.17 per 1000 births (=1 per 854 births; Stoll, Alembik, Dott & Roth, 1990) to lows of slightly less than 1 per 1000 (Forrester & Merz, 2002). In all cases, the amount of selective termination – particularly among women above 35 years of age – must be taken into account. Such analyses require diagnoses from both hospitals and perinatal offices (Siffel, Correa, Cragan & Alverson, 2004). Since trisomy 21 occurs more often in women aged 35 and older, any secular changes relating to women's delayed childbearing may also affect rates of Down syndrome (particularly among women who are White and well-educated; Siffel et al., 2004).

2) Mortality and life-expectancy. Over 70 years ago, Penrose (1933) estimated that the average life expectancy for a person with Down syndrome was 9 years, with early deaths

due to heart, respiratory, and other conditions. Today, such conditions are highly treatable and contribute fewer early deaths, although persons with Down syndrome continue to experience shorter life-spans than those in the general population. As a result, the average life-span of persons with Down syndrome has increased dramatically.

In one study, Yang, Rasmussen and Friedman (2002) examined the deaths of over 17,000 individuals with Down syndrome across the United States, over the period from 1983 through 1997. Overall, the median age at death increased from 25 years in 1983 to 49 years in 1997. The largest increase in age at death occurred in the early 1990s, and few differences were noted from region to region. The authors speculate that increased survival rates during the 1983-1997 period may relate to the lessening of institutionalisation (especially of children) and better medical practice, particularly to the more timely provision of cardiac surgery for children with Down syndrome.

3) Rates of various diseases. Many studies provide estimates for the proportion of persons with Down syndrome who have a wide variety of diseases or physical problems (see Cohen, 1996; Roizen, 2003; Roizen & Patterson, 2003). Such conditions include congenital heart defects, leukaemia, respiratory problems, hearing or vision problems, obesity, diabetes, seizures, obstructive sleep apnoea, coeliac disease, hypothyroidism, and, among adults, dementia.

In each case, studies generally report, as outcomes, the proportions of individuals with Down syndrome who show the relevant diseases or physical problems (e.g., poor vision). For the most part, such studies do not stratify rates by age, nor has much attention been paid to differences in rates due to gender, SES, race, or ethnicity. Moreover, for certain conditions, exceptionally wide ranges of estimates have been provided; witness Roizen and Patterson's (2003) conclusion, based on the three existing studies, that "Between 38% and 78%" of people with Down syndrome have hearing loss (p. 1283). Far more work is needed regarding the basic demographics of medical and physical conditions within this syndrome.

Considered in the aggregate, however, epidemiological studies document how often Down syndrome occurs, how long these individuals live, and how often they experience many medical and physical problems. One might argue that the field has now achieved a basic epidemiological view of Down syndrome. The time seems ripe to extend this basic information.

Advancing epidemiological studies of Down syndrome

To understand epidemiological advances concerning Down syndrome, it is first necessary to appreciate the wide variety of information that exists. Most of this information resides in already-collected records, often involving official administrative records that are routinely collected by governmental agencies. In most states, for example, Vital Statistics include records of birth, death, marriage, and divorce.

Some states have records of hospitalisations and doctor's visits, and many have educational records. In nearly every case, statewide administrative databases are now computerised.

To consider how such records can be used to answer interesting research questions, we first explore briefly some new computer-based techniques for linking a person's various records. After describing the techniques themselves, we present some examples of epidemiological studies in Down syndrome that make use of such record-linking techniques.

Technological advances involving linkage

As computing power increases, it becomes more feasible to analyse large-scale, epidemiological data on inexpensive desktop computers. Such computers have increased the number of records per unit time that can be examined in order to join together the records of the same individual. To understand how such data linkage works, we discuss separately issues of acquiring and cleaning data, matching, classifying, and achieving analytic datasets. Throughout these discussions, we rely on the early works of Howe and Lindsay (1981), Baldwin, Acheson and Graham (1987) and Newcombe (1988), who established the basic framework for record linkage, as well as more recent extensions (Boussy & Scott, 1993; Victor & Mera, 2001).

Acquiring and cleaning data

Before considering studies of this type, one must be aware of the various ethical issues and guidelines necessary to ensure confidentiality. In all large-scale epidemiological studies – particularly those involving administrative records – the greatest danger involves threats to subject confidentiality. Specific laws and protections, variable across regions, institutions, and datasets, exist to ensure that breaches do not occur. In most states and localities within the United States, for example, Birth Records are available to legitimate research organisations with approval from the researcher's Institutional Review Board (IRB). Within limits set by federal laws, school and health records are also available, albeit with significant restrictions on their uses.

In the studies described below, we utilise state-mandated records of birth, death, marriage, divorce, and hospital discharge. These computerised records, which in the state of Tennessee involve several million individuals for periods up to 13 years, constitute a vast, under-utilised resource to those interested in studying the characteristics of the population of persons with Down syndrome.

But even having acquired the data, one does not simply jump to analyses. Once data are in hand, the most time-consuming and tedious part of the process begins, as researchers must clean and check the data. Simply stated, longitudinal data are often not collected in the same way over time. Similarly, when using data from different sources, researchers must check that the coding of information is consistent. To achieve standard data from one type of record to another, computer experts expend enormous amounts of time

putting data from different databases into identical formats (Christen et al., 2004; Christen & Churches, 2005).

Matching

With multiple databases available to be linked, the next task involves matching the records of individual persons from one administrative dataset to another. One first decides what entities are to be linked and how the entities are identified in each dataset. For example, if one is interested in linking children in the Birth Records to their records in the Hospitalisation file, one would select the variables that uniquely identify those individuals in both the Birth and Hospitalisation datasets. Variables likely to occur in both records include Social Security Number (SSN), Birth Date, Last and First Names, Race, and Gender.

Ideally, each record in the Birth dataset is compared to each record in the Hospitalisation dataset. The problem is that, with many different matching variables and several million records, even the fastest, most powerful computers would still take days or even weeks to examine all subject pairs. To limit the number of potential comparison pairs, records are often grouped into blocks and only records within blocks are compared. A blocking variable is chosen to bring together records that are similar. Unique identifiers like SSNs are the ultimate blocking factor. If SSNs are recorded accurately in all records, record pairs with the same SSNs are very likely matches. By using such a unique identifier, SSNs allow one to match two or more separate records, for thousands or even millions of individuals, in seconds or minutes.

Classification

The final step in linkage involves designating each pair of records as a match or not a match. Two general strategies are used for these designations: deterministic and probabilistic (Fellegi & Sunter, 1969; Gomatam, Carter, Ariet & Mitchell, 2002; Jaro, 1995). *Deterministic linkage* strategies examine the number of agreeing identifiers in a pair of records and designate the one with the most agreements (above a minimum number of agreements) as the matching pair; all other pairs are non-matches. The decision process can be a single rule (SSN only) or stepwise with multiple stopping rules (SSN = match, or SSN, then go to name, etc.). In contrast, *probabilistic linkage* processes assign weights to each pair of variables, with positive weight for an agreement and a negative weight for a disagreement. Since each individual has their own, unique social security number, SSNs constitute a variable with a very high positive weighting: when two records agree on social security number, they most likely come from the same person. In contrast, names (even last names) would receive less weight, as many individuals have the same first name (e.g., John) and, particularly in the case of common last names, the same last name as well (Jones, Smith). Taken together, the sum of the weights from the various matching variables for each pair of records constitutes the likelihood that they are a match. Scores above an upper cutoff are considered a match, those below a non-match.

Analysis datasets

Once all records are matched and decisions have been made about which ones describe the same individual, the researcher must now create datasets that both protect privacy and are easy to manage. The researcher first de-identifies all of the now-linked records, removing information that could allow any individual record to be traced back to a particular, identifiable individual. These linked, de-identified datasets are then formatted for use with SPSS, SAS, STATA, or other commonly used statistical package. From the perspective of the researcher analysing the data, the final step in the process has now arrived. The researcher can now perform analyses involving thousands – sometimes even millions – of individual subjects. Such analyses can now be performed as easily as analysing 30 subjects from a study in which every subject was tested individually.

Before proceeding, it is important to note that the general linkage procedures described above can actually be performed in three distinct ways. First, one can join together different records of the same individual. As described above, one can link the same individual's birth with hospitalisation records, or marriage with divorce records. Second, one can join together multiple records of the same type. Using all of the state's hospitalisation records, one might create 'patient profiles' of all of an individual's hospitalisations—when, where, for how long a hospital stay, and for which disease(s). Third, using the mother's or father's social security number, one can create family records. This last procedure, referred to as second-order linkage (Tu & Mason, 2004), allows one to horizontally order all children in a family. The first child might be on the left-most side of the line, with birth date and gender, followed (moving rightward) by the second child, then the third. In addition to considering these linkage applications separately, one can also join the three techniques.

Studying Down syndrome using linked administrative records: some ongoing studies

Several ongoing studies illustrate the uses of large-scale epidemiological approaches to answer questions concerning children with Down syndrome and their families. All of these studies utilise linked, administrative datasets, covering the entire population of the state of Tennessee, over periods beginning as early as 1990.

Divorce among parents of children with Down syndrome

In the first study, Urbano and Hodapp (2006) examined the amount, timing, and correlates of divorce among parents of children with Down syndrome. We first identified children with Down syndrome using the Birth records and subsequent Hospital Discharge records. On the basis of the mother's social security number (SSN), children were grouped into families. To these linked family-wide records were linked information about the mother's and father's marriage and, when applicable, divorce.

Using such individual and family linkage techniques and given the size of the state's population (5.8 million people overall), we were able to identify 918 births of children with Down syndrome during the 1990-2002 time period (given a 1/1000 prevalence rate, these constituted 86.9% of all children likely to be born with the syndrome). Of these children, 659 were born to married mothers. Families of these children were compared to 463,008 families of children who had no identified disabilities (i.e., comparison group families).

Results indicated that parents of children with Down syndrome showed slightly lower amounts of divorce, and that the timing of divorce differed markedly across the two groups. Specifically, when parents of children with Down syndrome did divorce, their divorces more often occurred early on. Among all divorces within the Down syndrome group, almost 1/3 occurred before the child turned 2 years, compared to less than 20% of such quickly-occurring divorces in the comparison group.

We were also interested in variables that made divorce more or less likely. In both groups, divorce occurred more often when the parents were younger in age, but both education and rural living status were differentially important across the two groups. Thus, when either the mother or the father had not completed high school (i.e., had less than 12 years of formal schooling), divorce in both groups was more likely to occur. The pattern, however, was much more pronounced among the mothers and fathers of children with Down syndrome.

An even more extreme finding occurred among less educated fathers who resided in rural areas. Tennessee is a predominantly rural state, with 67 of the state's 95 counties considered rural by the federal government. We therefore had many families of children with Down syndrome who resided in rural areas. Among fathers of children with Down syndrome, the combination of being less educated (i.e., non-high-school graduate) and living in a rural area led to a very high risk of divorce. Among rural fathers who had not completed high school, divorce occurred in 32% of the group. This percentage of divorces was many times higher than among high school graduates of children with Down syndrome who lived in rural areas, or of non-rural fathers of either group.

Other studies using Tennessee statewide administrative databases

In addition to our first study of parental divorce in Down syndrome, we are also examining several questions related to health and families of children with Down syndrome. Specifically, these studies are focused on the following issues:

- **Early hospitalisations of children with Down syndrome**

In a second study, So, Urbano, and Hodapp (2006) are examining the amount, correlates, timing, and causes of hospitalisation in children with Down syndrome during

their first three years of life. In this study, we are using Tennessee's hospital discharge records from 1997-2002 to examine patterns of inpatient care use in all of the state's infants with Down syndrome who were identified in their birth hospitalisation and who were born between 1997 through 1999, inclusive.

Our findings show a pattern of hospitalisation that might best be described as 'early and often.' First, half of all infants-toddlers with Down syndrome were hospitalised one or more times (not counting their birth hospitalisation) and this pattern was especially pronounced among infants who had congenital heart defects. And, although we focused on the entire 0-to-3 year period, the large majority of hospitalisations occurred within the child's first year, often within the first three months of life. In addition to congenital heart defects, the most common reasons for hospitalisation were such respiratory illnesses as pneumonia, bronchitis, and bronchiolitis.

- **Possible connections between early hospitalisation and early divorce**

So far, we have found that, when divorce does occur within the Down syndrome group, it more often occurs early on, within the child's first two years of life. In terms of the children themselves, we noted the often-occurring, serious health problems, experienced by roughly half of children with Down syndrome, which mostly began within the child's first year of life.

Is there a connection between early health problems in the children and parental marital difficulties? Stated more concretely, might children with Down syndrome who experience early, repeated, and long hospitalisations have parents who are more likely to divorce early on? As we can link together hospital discharge records of the children and divorce records of the parents, we will pursue this possibility.

- **Family correlates of early hospitalisation**

Another ongoing study concerns family correlates of early hospitalisation. It is now widely known that low SES and minority status relate to lower levels of health (Graham, 2005). This finding, which recurs across various populations in various countries, has yet to be examined for children with Down syndrome.

Again, our use of large numbers of linked, administrative records allows us to examine these issues. At present, very little is known in Down syndrome about effects of low SES, lower levels of parental education, or minority or rural status. But such variables are often routinely recorded in administrative datasets. Once records of one type (that might include information about SES, parental education, or where the family resides) have been linked to hospitalisation records (that give rich information about health), one can begin to determine if such usual correlates of health status also predict health outcomes of persons with Down syndrome.

Taken together, these various studies – some of which are recently completed, others of which are only beginning –

illustrate some of the possibilities of epidemiological studies in Down syndrome. As noted above, some of these studies link different records (e.g., birth, hospitalisation) for the same individual. Others link together records within the same dataset over time, as in the study of the early hospitalisations of children with Down syndrome over the first three years of life. Still other studies mix-and-match across people and records, as in our beginning study of whether early, recurrent, and extended hospitalisations of children with Down syndrome might relate to higher proportions of early divorce among these children's parents. In all cases, by examining linked administrative data for an entire state across multi-year periods, we are able to examine questions that are beyond the reach of most research initiatives.

Epidemiological studies of Down syndrome: prospects and challenges

In reflecting upon our own and future epidemiologically-based studies in Down syndrome, we focus on four major issues.

1) Need to be creative in conceptualising and using databases

As each of us goes about our daily business, we create an information trail documenting the events of our lives. Many of these events are recorded and saved by us, our families, insurance companies, schools, doctors, hospitals, and local, state, and federal agencies. Although available records vary from place to place, the following are just some of the types of information that we create: birth, immunisation, healthcare, school, psychological testing, income, tax, professional licensing, ownership, warranty, purchases, military service, marriage, divorce, travel, and death. Given appropriate ethical safeguards, many of these records are available to be linked and used for research.

In using such databases, we begin with research questions, then tie such questions to available or potentially available data. In our collaborative research so far, members of our team have been interested in different aspects of family functioning and health for persons with Down syndrome. In the study of divorce among families of children with Down syndrome, we linked together the child's birth and hospitalisation records to identify children with Down syndrome. Families were then constructed, and marriage and divorce indicators linked to these family records. In the early hospitalisation study, different hospitalisations were linked together to achieve patient records of each child with Down syndrome. In both cases, linking across and within different population-based datasets allowed us to answer the research questions of interest.

2) Need for studies that extend beyond usual participants and findings

In most studies of children with Down syndrome, parents and children are recruited from the local area or from regional or national parent groups. Such studies therefore feature parents and children who are middle- or upper-middle class, who are suburban or urban dwellers, or who

live close to the university or clinic out of which the study is being conducted. Given that the total numbers of subjects in many studies range from 20-40 per group, inference about the population's characteristics is limited. In large-scale, epidemiological studies, we at least know who our subjects are, and to what degree our subjects represent the entire population of the town, state, or country. Although any particular state may not perfectly represent the country as a whole, the outlines of any possible bias have been made explicit.

But also important is the related issue of missing subjects. Who, exactly, avoids participating in one's study? Here we return to our findings concerning the high rates of divorce among fathers who were less educated (<12 years formal schooling) and who lived in rural areas. Earlier studies had hinted that lower levels of parent education might relate to higher rates of troubled marriages among Down syndrome families (Gath & Gumley, 1986; Sloper, Knussen, Turner & Cunningham, 1991). Until now, however, few studies had the numbers of less educated, rural fathers to examine the effects of the interaction of the two on parental divorce. Other hard-to-access groups – including single parents, minority groups and low SES groups – might also be examined using large-scale, epidemiological approaches.

Just as we need to go beyond the 'usual suspects' in the participants in our studies, so too do we need to include hypotheses of different types. In studies of Down syndrome, it would seem that different types of hypotheses exist. The first involves the degree to which findings already shown in non-disabled children (or in their families) apply to children with Down syndrome. Among non-disabled children and adults, for example, poorer levels of health are found in persons who come from low-SES families (Graham, 2005). To what extent do low-SES children with Down syndrome also show poorer levels of health?

The second type of hypothesis relates to what has been called the "Uniqueness Question" (Pennington, O'Connor & Sudhalter, 1991). Simply stated, to what degree is any finding noted for children with intellectual disabilities in general also found among children with Down syndrome and vice-versa? Taking an example from our own studies, to what extent are children with other types of disabilities hospitalised in their first year of life, and, when children do have other disabilities, what proportion of divorces in their families occur when children are younger? Clearly, more work is needed in this area.

3) Need to go beyond disability, health, and even classical epidemiological perspectives

In Table 1, virtually every definition of epidemiology highlights disease or characterises outcomes that are related to health. However, outcomes more distantly related to health are also of interest. Large-scale, linked, administrative datasets also provide rich information on family structure, income, and service use. Many of epidemiology's statistical techniques can then be applied to study these other outcomes. In the first study, for example, we used as our outcome divorce of parents of children with versus without

Down syndrome. Similarly, one might examine as outcomes whether or not one is employed, has graduated high school, is using welfare or other social services, is living independently, or owns a car or a house. Some of these outcomes (e.g., divorce) may contribute to a broader understanding of what constitutes risks to and contributors to states of health and well-being, broadly defined. Using an epidemiological framework, then, might help us to understand how to improve the quality of life for people with Down syndrome in all of the many aspects of their well-being.

4) Need to dovetail large-scale and small-scale perspectives

Given their predilection toward examining entire populations, epidemiologists have often been likened to airplane pilots. The epidemiologists' viewpoint, so the metaphor goes, is from 30,000 feet, as they notice that a mountain exists in one direction, a town or highway in another.

As with any single perspective, however, this high-in-the-sky view is also limited. In most cases, epidemiologists are limited in that they find one or another 'blip' on the screen, but cannot immediately determine why that blip occurred. In our case, for example, we know that, when they do divorce, parents of children with Down syndrome divorce proportionately more often during the first two years of the child's life. But why are early divorces disproportionately present in Down syndrome? Does early divorce relate to the shock of having a child with Down syndrome, an added stressor on an already weak marriage, or the result of a long series of the child's hospitalisations? At present, we do not know.

We now need to combine the high-above and the close-up perspectives. Granted, the perspective from 30,000 feet is necessary, in that it tells the field if indeed the phenomenon is occurring – with large numbers of subjects and throughout an entire population. But this large-scale approach is not sufficient by itself. Follow-up studies are badly needed to move from risk indicators to risk mechanisms (Rutter et al., 2001).

In some sense, then, we have returned to John Snow and his pump on Broad Street. Faced with a horrific outbreak of cholera and conflicting ideas about what was causing that outbreak, Snow engaged in what has been called 'shoe leather epidemiology.' He went out and interviewed his 600+ families and determined that each victim did indeed drink from the contaminated water-pump. So too do we need to join the high-in-the-sky and the up-close-and-personal viewpoints when examining children with Down syndrome and their families. We need the 30,000 foot view to determine many basic facts and connections that, until now, have been hard to determine, but we also need more shoe leather to identify causes of disease and other real-life outcomes. In short, only by joining a broadly based epidemiological viewpoint with our current, more microscopic perspectives can we truly make progress in understanding that unique syndrome that for 140 years has remained both known and unknown.

References

- American Association on Mental Retardation (1992). *Mental Retardation: Definition, classification, and systems of supports*. Washington, DC: Author.
- American Association on Mental Retardation (2002). *Mental Retardation: Definition, classification, and systems of supports* (10th ed.). Washington, DC: Author.
- American Psychiatric Association (2004). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed, Text Revision). Washington, DC: Author.
- Baldwin, J. A., Acheson, E.D. & Graham, W.J. (1987). *Textbook of Medical Record Linkage*. Oxford New York Toronto, Oxford University Press.
- Boussy, C. A. & K. G. Scott (1993). Use of data-base linkage methodology in epidemiologic studies of mental retardation. *International Review of Research in Mental Retardation*, 19, 135-161.
- Cameron, D. & Jones, I.G. (1983). John Snow, the Broad Street pump and modern epidemiology. *International Journal of Epidemiology*, 12, 393-396.
- Cohen, W.I. (Ed.). (1996). Health care guidelines for individuals with Down syndrome (Down syndrome preventive medical check list). *Down Syndrome Quarterly*, 1(2), 1-10.
- Christen, P., et al. (2004). *Febrl - A Parallel Open Source Data Linkage System*. Proceedings of the 8th Pacific-Asia Conference, Sydney, Australia, Springer Lecture Notes in Artificial Intelligence.
- Christen, P. & Churches, T. (2005). *Febrl - Freely Extensible Biomedical Record Linkage (Manual, release 0.3)*. Retrieved November 21, 2005, from Source Forge <http://sourceforge.net/projects/febrl>.
- Dykens, E.M. (2000). Psychopathology in children with intellectual disability. *Journal of Child Psychology and Psychiatry*, 41, 407-417.
- Einfeld, S.L. & Tonge, B.J. (1996). Population prevalence of psychopathology in children and adolescents with intellectual disability: II epidemiological findings. *Journal of Intellectual Disability Research*, 40, 99-109.
- Everitt, B.S. (1995). *The Cambridge Dictionary of Statistics in the Medical Sciences*. Cambridge, UK: Cambridge University Press.
- Fellegi, I.P. & Sunter, A.B. (1969). A theory for record linkage. *Journal of the American Statistical Association*, 64 (328), 1183-7.
- Forrester, M.B. & Merz, R.D. (2002). Epidemiology of Down syndrome (trisomy 21), Hawaii, 1987-97. *Teratology*, 65(5), 207-212.
- Gath, A. & Gumley, D. (1986). Family background of children with Down's Syndrome and of children with a similar degree of mental retardation. *British Journal of Psychiatry*, 149, 161-171.
- Gibson, D. (1978). *Down's Syndrome: The Psychology of Mongolism*. Cambridge, UK: Cambridge University Press.
- Gomatam, S., Carter, R., Ariet, M. & Mitchell, G. (2002). An empirical comparison of record linkage procedures. *Statistics in Medicine*, 21(10), 1485-1496.
- Graham, H. (2005). Intellectual disabilities and socioeconomic inequalities in health: An overview of research.

- Journal of Applied Research in Intellectual Disabilities*, 18, 101-111.
- Greenberg, R.S., Daniels, S.R., Flanders, W.D., Eley, J.W. & Boring, J.R. (2001). *Medical Epidemiology* (3rd ed.). New York, NY: Lange Medical Books/McGraw-Hill.
- Hodapp, R.M. & Dykens, E.M. (2004). Studying behavioral phenotypes: Issues, benefits, challenges. In E. Emerson, C. Hatton, T. Parmenter & T. Thompson (Eds.), *International Handbook of Applied Research in Intellectual Disabilities* (pp. 203-220). New York: John Wiley & Sons.
- Howe, G.R. & Lindsay, J. (1981). A generalized iterative record linkage computer-system for use in medical follow-up studies. *Computers and Biomedical Research*, 14, 327-340.
- Jaro, M.A. (1995). Probabilistic linkage of large public-health data files. *Statistics in Medicine*, 14, 491-498.
- Kearns, J. (2000). Children and cultural differences. In P. Dudgeon, D. Garvey & H. Pickett (Eds.), *Working with Indigenous Australians: A handbook for psychologists*. Perth, Western Australia: Gunada Press.
- Leonard, H. & Wen, X. (2002). The epidemiology of mental retardation: Challenges and opportunities in the new millennium. *Mental Retardation and Developmental Disabilities Research Reviews*, 8, 117-134.
- MacMillan, D.L., Gresham, F.M. & Siperstein, G.N. (1993). Conceptual and psychometric concerns about the 1992 AAMR definition of mental retardation. *American Journal on Mental Retardation*, 98, 325-335.
- Mezzich, J.E. & Ustin, T.B. (2005). Epidemiology. In B.J. Sadock & V.A. Sadock (Eds.), *Comprehensive Textbook of Psychiatry* (8th ed.). Vol. 1 (pp. 656-672). Philadelphia, PA: Lippincott Williams and Wilkins.
- National Down Syndrome Society (2005). Parent and professional information. Available on-line at www.ndss.org.
- Newcombe, H. B. (1988). *Handbook of Record Linkage: Methods of health and statistical studies, administration and business*. Oxford, UK: Oxford University Press.
- Paneth, N. (2004). Assessing the contributions of John Snow to epidemiology: 150 years after removal of the Broad Street pump handle. *Epidemiology*, 15, 514-516.
- Pennington, B.F., O'Connor, R. & Sudhalter, V. (1991). Toward a neuropsychology of fragile X syndrome. In R.J. Hagerman & A.C. Silverman (Eds.), *Fragile X Syndrome: Diagnosis, treatment, and research* (pp. 173-201). Baltimore: Johns Hopkins University Press.
- Penrose, L.S. (1933). *Mental Defect*. London: Sidgwick & Jackson.
- Reigier, D.A. & Burke, J.D. (2000). Epidemiology. In B.J. Sadock & V.A. Sadock (Eds.), *Kaplan & Sadock's Comprehensive Textbook of Psychiatry* (7th ed.), Vol. 1 (pp. 500-522). Philadelphia, PA: Lippincott, Williams and Wilkins.
- Roeleveld, N., Zielhuis, G.A. & Gabreels, F. (1997). The prevalence of mental retardation: A critical review of the literature. *Developmental Medicine and Child Neurology*, 39, 125-132.
- Roizen, N.J. (2003). The early interventionist and the medical problems of the child with Down syndrome. *Infants and Young Children*, 16, 88-95.
- Roizen, N.J. & Patterson, D. (2003). Down's syndrome. *The Lancet*, 361, 1281-1289.
- Rutter, M., Pickles, A., Murray, R. & Eaves, L. (2001). Testing hypotheses on specific environmental causal effects on behaviour. *Psychological Bulletin*, 127, 291-324.
- Sears, R.R. (1975). Your ancients revisited: A history of child development. In E.M. Hetherington (Ed.), *Review of Child Development Research*. Vol. 5, pp. 1-73. Chicago, IL: University of Chicago Press.
- Siffel, C., Correa, A., Cragan, J. & Alverson, C.J. (2004). Prenatal diagnosis, pregnancy terminations and prevalence of Down syndrome in Atlanta. *Birth Defects Research (Part A)*, 70, 565-571.
- Sloper, P., Knussen, C., Turner, S. & Cunningham, C. (1991). Factors related to stress and satisfaction with life in families of children with Down's syndrome. *Journal of Child Psychology and Psychiatry*, 32, 655-676.
- So, S.A., Urbano, R.C. & Hodapp, R.M. (2006). *Hospitalizations for Infants and Young Children with Down Syndrome: Evidence from person-records from a statewide administrative database*. Submitted.
- Stoll, C., Alembik, Y., Dott, B. & Roth, M.P. (1990). Epidemiology of Down syndrome in 118,265 consecutive births. *American Journal of Medical Genetics, Supplement* 7, 79-83.
- Summers, J. (1989). *Soho: A history of London's most colourful neighbourhood*. London, UK: Bloomsbury.
- Susser, M. & Susser, E. (1996). Choosing a future for epidemiology: II. From black box to Chinese boxes to eco-epidemiology. *American Journal of Public Health*, 86, 674-677.
- Tonge, B.J. & Einfeld, S.L. (2003). Psychopathology and intellectual disability: The Australian Child to Adult longitudinal study. *International Review of Research in Mental Retardation*, 26, 61-91.
- Tu, S.F. & Mason, C.A. (2004). Organizing population data into complex family pedigrees: Application of second-order data linkage to state birth defects registries. *Birth Defects Research, Part A. Clinical and Molecular Teratology*, 70, 603-608.
- Urbano, R.C. & Hodapp, R.M. (2006). *Divorce in Families of Children with Down Syndrome: A population-based study*. Submitted.
- Victor, T.W. & Mera, R.M. (2001). Record linkage of health care insurance claims. *Journal of the American Medical Informatics Association*, 8, 281-288.
- World Health Organisation (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical descriptions and diagnostic guidelines*. World Health Organisation, Geneva.
- Yang, Q., Rasmussen, S.A. & Friedman, J.M. (2002). Mortality associated with Down's syndrome in the USA from 1983 to 1997: A population-based study. *The Lancet*, 359, 1019-1025.
- Yargin-Allsopp, M., Murphy, C.C., Oakley, G.P. & Sikes, R.K. (1992). A multiple-source method for studying the prevalence of developmental disabilities in children: The Metropolitan Atlanta Developmental Disabilities Study. *Pediatrics*, 89, 624-629.
- Yargin-Allsopp, M. & Boyle, C. (2002). Overview: The epidemiology of neurodevelopmental disorders. *Special*