National Suicide Rates a Century after Durkheim: Do We Know Enough to Estimate Error?

Cynthia A. Claassen, PhD, Paul S. Yip, PhD, Paul Corcoran, PhD, Robert M. Bossarte, PhD, Bruce A. Lawrence, PhD, and Glenn W. Currier, MD, MPH

Durkheim's nineteenth-century analysis of national suicide rates dismissed prior concerns about mortality data fidelity. Over the intervening century, however, evidence documenting various types of error in suicide data has only mounted, and surprising levels of such error continue to be routinely uncovered. Yet the annual suicide rate remains the most widely used population-level suicide metric today. After reviewing the unique sources of bias incurred during stages of suicide data collection and concatenation, we propose a model designed to uniformly estimate error in future studies. A standardized method of error estimation uniformly applied to mortality data could produce data capable of promoting high quality analyses of cross-national research questions.

National rates of suicide are used by an ever-expanding part of the global community as a crude public health measure (e.g., Lopez & Mathers, 2006; Pritchard & Hansen, 2005)—a practice based on the assumption that these national rates accurately reflect the number of suicides which actually occur. Unfortunately, labeling and registering a death as a suicide is anything but a straightforward process, and appropriate use of official suicide data therefore requires a thorough understanding of both the sources and magnitude of error contained within this public health measurement tool (Cantor, McTaggart, &

Cynthia Claassen, Robert Bossarte, and Glenn Currier are with the VISN 2 Center of Excellence for Suicide Prevention, U.S. Veterans Administration, and affiliated with the University of Rochester Medical Center in Rochester, NY; Paul Yip is Director, Centre for Suicide Research and Prevention at the University of Hong Kong; Paul Corcoran is Deputy Director, National Suicide Research Foundation in Cork, Ireland; and Bruce Lawrence is Associate Research Scientist at the Pacific Institute for Research and Evaluation in Calverton, MD.

The authors wish to express their deep gratitude to Kenneth D. Kochanek of the Mortality Statistics Branch, Division of Vital Statistics, U.S. National Center for Health Statistics, and to Joseph A. Prahlow, Chairman, Board of Directors, National Association of Medical Examiners, for their generosity in lending their substantial expertise to this project.

The authors were supported, in part, by grant P20MH071897 to the Center for Public Health and Population Interventions for Preventing Suicide (E. Caine, PI—Drs. Currier and Bossarte), the American Foundation for Suicide Foundation (Drs. Bossarte and Claassen), and the Timberlawn Foundation (Dr. Claassen).

Address correspondence to Cynthia A. Claassen, PhD, c/o Center of Excellence, Canandaigua VAMC, 400 Fort Hill Blvd, Canandaigua, NY 14472; E-mail: Cindy.Claassen@va.gov

De Leo, 2001; Groholt & Ekeberg, 2003; Jobes, Casey, Berman, & Wright, 1991; Mathers, Ma Fat, Inoue, Rao, & Lopez, 2006; O'Carroll, 1989; Stengel & Farberow, 1967).

Over a century ago, Durkheim (1897/1951) first suggested that national suicide rates were one reflection of a nation's sociocultural climate. He was familiar with prior work suggesting that these rates likely contained some degree of error, but assumed that their relative stability from year to year was nonetheless evidence of sufficient reliability to permit sociological analyses (Douglas, 1967). Only four decades after publication of his seminal work on suicide, however, the rates in some European countries were found to be differentially underreported by method of injury, suggesting that these suicide data contained not only random error but substantial systematic error as well (Zilborg, 1935). This discovery re-energized earlier debates about the fidelity of suicide mortality data, a debate which has yet to be completely resolved (Atkinson, Kessel, & Dalgaard, 1975; Breiding & Wiersema, 2006; Brooke, 1974; Brugha & Walsh, 1978; Douglas, 1967; Farberow & Neuringer, 1971; Hesso, 1987; Jennings & Barraclough, 1980; Jobes et al., 1991; Monk, 1987; Naughton, Doyle, Melina, & Barry, 1996; Peck & Warner, 1995; Platt, 1986; Platt, Hawton, Kreitman, Fagg, & Foster, 1988; Ross & Kreitman, 1975; Stengel & Farberow, 1967; Whitt, 2006a, 2006b).

Even in relatively recent studies, surprising levels and patterns of suicide rate error continue to be uncovered. Early results from the United Kingdom's (UK's) National Confidential Inquiry into Suicide and Homicide by People with Mental Illness were of limited use because of a disappointing case ascertainment rate (Appleby, Shaw, & Amos, 1997; House, 1997; Royal College of Psychiatrists, 1996). Widespread underreporting also appears to be the norm in some Islamic countries (Pritchard & Amanullah, 2006). In the Australian state of Queensland, official rates have trended downward over the past several years while a 2007 Queensland Suicide Monitoring Registry study does not reflect the same trend (De Leo, 2007). Finally, annual suicide rates were underestimated in Hong Kong by as much as 18% in one year, and by 6% per year in Ireland over more than a decade's time. Both countries' error was attributed to reporting deadlines set too narrowly, leaving suicides occurring near year's end unreported because of inadequate time to complete death investigations (Corcoran, Arensman, & O'Mahoney, 2006; Cui, Yip, & Chau, 2004).

Regardless of data pool, decade, or geographic location, twentieth-century studies demonstrate a markedly consistent finding of *underreporting* in local and national rates of suicide, suggesting that a region's "true" suicide rate is almost always higher than the officially reported rate. However, these studies also reveal that level of underreporting may vary widely by source of error (Birkhead, Galvin, Meehan, O'Carroll, & Mercy, 1993; Kleck, 1988; Warshauer & Monk, 1978). The relevant question for any given mortality data set, then, is not *whether* there is error in suicide rates, but *how much* error, *coming from what sources*, impacting rates *to what degree*, and *trending how* over time (K. Kochanek, personal communication, April 2007; Neelman & Wessley, 1997)?

In countries with vital event registration systems, annual cause-of-death statistics are often developed through a complex extended reporting chain. The Irish death reporting system (illustrated in Figure 1), for instance, can include as many as eleven separate steps prior to death registration—each representing a potential source of error. To date, national death reporting systems have rarely been evaluated in a manner comprehensive enough to quantify total suicide error rate incurred across steps in the mortality reporting process. A staged framework within which common types of error can be characterized might therefore be of assistance.

While reporting steps vary by mortality system, the known sources of suicide measurement error can largely be encompassed within a three-stage framework (Table 1). In this model, Stage One error occurs during

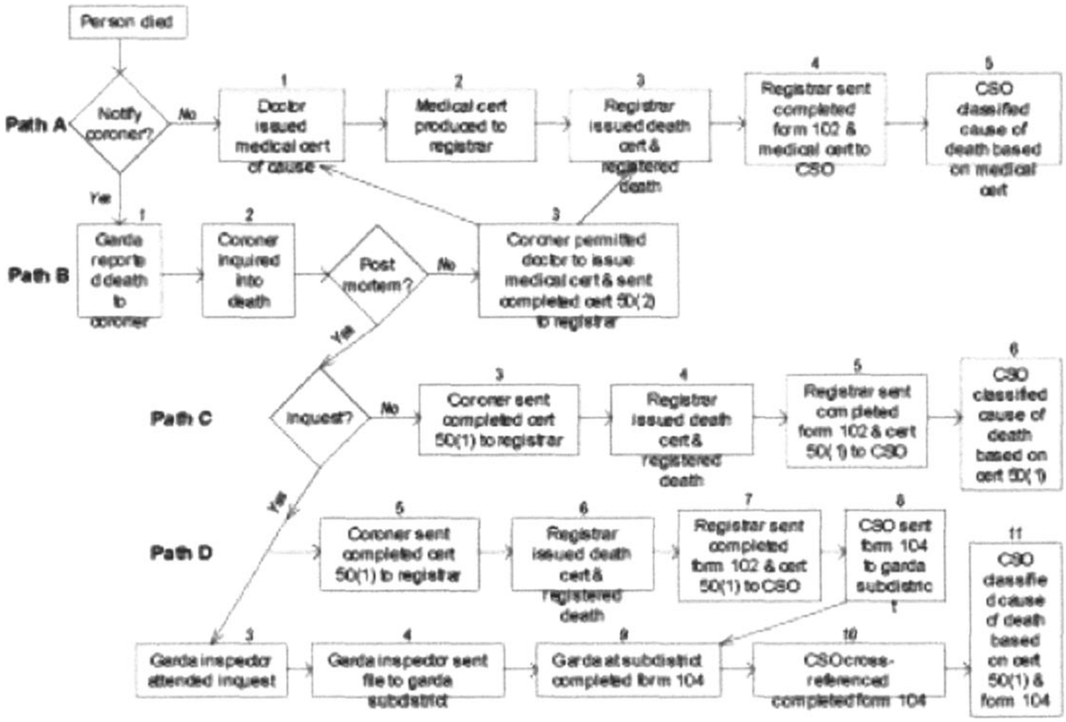

Figure 1. Schematic diagram describing the process of death registration and cause of death determination in Ireland
Source: National Suicide Research Foundation (2007). *Inquested deaths in Ireland: A study of routine data and recording procedures.* Cork: Author.

the death classification process when establishing official cause(s) of death. Stage Two, errors of death codification, involve technical problems associated with death certificate completion, cause-of-death coding, and identification of the underlying cause of death for official reporting purposes. Stage Three error occurs during annual, all-cause mortality report preparation, and is associated with missing or vague cause data and/or inaccurate estimation of denominator populations. In this paper, we describe the nature of the error incurred within each stage. Then, after modeling interrelationships between stages, we conclude the discussion by identifying study design issues that impact the scientific utility of research within this genre. Throughout the discussion, causes of death are defined using the World Health Organization's (WHO's) *International Statistical Classification of Disease* manuals (*ICD*; WHO, 1992), and the WHO's International Collaborative Effort for Injury Statistics Committee's recommended case categories (Injury-ICE; Centers for Disease Control and Prevention, 1997; Injury-ICE, 2002; Table 2). In the WHO classification system, ill-defined cause deaths refer to deaths coded as (*ICD*-10) R99, (*ICD*-9) 799.9: "Other ill-defined and unspecified causes of mortality," or (*ICD*-9) 799.99: "Other unknown and unspecified causes."

ESTIMATING STAGE 1 ERROR: QUANTIFYING RATES OF UNDERASCERTAINMENT AND MISCLASSIFICATION OF SUICIDES

The majority of all suicide rate error studies have focused exclusively on biases occurring during the first stage of mortality rate production and involve misclassification of intent in injury-related deaths. Prior work

TABLE 1
Sources of Error in Suicide Rates by Stage in Rate Development Process and Common Methods of Error Estimation

Stage of Rate Development	Source of Error	Methods of Estimating Level of Error
I. Death Classification Process	Case underascertainment and misclassification	1. Expert case review 2. Indirect, upper limit estimating
II. Death Codification	A. Death certificate completion error B. Inadequate cause listings	(Both sources of error) Multiple cause of death automated file subroutine analysis)
III. Rate Calculations	A. Error caused by reporting delays B. Error caused by inaccurate population estimates C. Error due to data loss	A-1. Horvitz-Thompson-type model establishing number of suicides omitted under various cut-off dates for annual reporting. A-2. Comparison of trends in counts of suicides based on date of occurrence vs. numbers in preliminary annual reports, assessed for trend via Pearson's correlation coefficient. B. Error of closure C. Mortality system data loss indices; population completeness and coverage indices

demonstrates that accurate classification of intent in a self-inflicted, fatal injury can be a formidable challenge (Cantor et al., 2001; Farberow, MacKinnon, & Nelson, 1977; Fauveau, 2006; Gist & Welch, 1989; Litman, Curphey, Shneidman, Farberow, & Tabachnick, 1963). Central to the process of officially labeling a suicide is the need to confirm that (a) the death was deliberately self-initiated and (b) the decedent intended to induce death via the specific injury act that ended life (Litman et al., 1963; O'Carroll, 1989). Typically, under half of all suicides contain explicit (objective) markers of the decedent's intent, suggesting that some degree of inference is required to classify most of these deaths (McCarthy & Walsh, 1975). Even in the absence of social pressure, the intent of some injury deaths (e.g., unobserved falls, defensive police shootings, self-poisoning in the substance dependent) may be difficult to establish (Douglas, 1967; Groholt & Ekeberg, 2003; Kleck, 1988; Litman et al., 1963).

The amount of evidence legally required to support an inference of suicide varies both within and across countries, and national suicide rates are substantially impacted by the specific standard of proof required for cause-of-death determination (Cooper & Milroy, 1995; Farberow et al., 1977; Nelson, Farberow, & MacKinnon, 1978; Stanistreet, Taylor, Jeffrey, & Gabbay, 2001; Walsh, Walsh, & Whelan, 1975). The two most common standards in use today are the legal and medical standards (Brooke, 1974; Fingerhut, Cox, & Warner, 1998; Gist & Welch, 1989; O'Carroll, 1989). A *legal* verdict of suicide is rendered by a coroner, based on a requirement that suicidal intent is established beyond reasonable doubt, while *medical* cause-of-death is rendered by medically-trained personnel who assign a suicide diagnosis in accordance with reasonable medical certainty—a criterion often employing a lower (more subjective) standard of proof (McCarthy & Walsh, 1975; O'Carroll, 1989). The trend in western countries has been away from the legal and toward the medical standard, but the legal standard remains in effect in some places, notably Great Britain and parts of the United States (Hanzlick & Combs, 1998).

Two basic methods have been used to quantify Stage One underascertainment within a given mortality data set. The most direct of these is the expert case review method, but the indirect, upper limit estimating procedure has perhaps been used for a broader range of research questions.

Expert Case Review and Reclassification

This method involves reclassification of a sample of medical examiner/coroner case files to quantify level of underreported suicides. The scope of cases selected for reclassification varies by study, but at minimum includes all injury deaths of undetermined intent. The stringency of case inclusion/exclusion criteria also varies, but more rigorous investigations tend to use case exclusion criteria that address completeness of record as well as presence or absence of specific elements necessary for classification. Some studies ask expert reviewers to both classify deaths and rate their level of confidence in that classification using methods such as categorical "confidence ratings" (i.e., probable suicide; possible suicide; suicide unlikely) (Cantor et al., 2001).

In most studies, the requirement of evidence of intent within the act itself (i.e., evidence that the individual was trying to die via the act that precipitated death) is inferred (see Table 3). Conversely, explicit criteria were developed by Ohberg and Lonnqvist (1998), who used guidelines created by a working group from the Centers for Disease Control and Prevention (Rosenberg et al., 1988), and by Brent, Joshua, Perper, and Allman (1987), who rated intent using items derived from the Suicide Intent (Beck, Resnik, & Lettieri, 1974) and Risk Rescue Rating scales (Weisman & Worden, 1972).

Study investigators often regard their expert panel's decisions as a kind of gold standard against which rates of case underascertainment can be calculated, but this assumption is almost always unsubstantiated. The exception is a study by Clarke-Finnegan and Fahy (1983), who examined the concurrent validity of their review panel's classifications via secondary peer review of the primary panel's final classifications. Results of this concurrent validation exercise were analyzed for agreement in both cause of death and decision-making certainty.

Indirect, Upper Limit Estimation of Suicide Rates

Although no true gold standard exists by which to assess the validity of suicide classification, O'Carroll (1989) has noted that these data generally have adequate *specificity* because nonsuicide deaths are rarely if ever classified as suicides. In contrast, the *sensitivity* of suicide mortality data (e.g., the degree to which true suicides are correctly identified) is of much greater concern. When data sensitivity is the major concern, the upper limit estimating procedure can be used to establish the theoretical maximum number of suicides that may have occurred, as well as the theoretical maximum number misclassified.

The choice of mortality categories assumed to contain misclassified suicides heavily influences upper level estimates, and studies using this methodology vary widely in terms of the nonsuicide mortality categories analyzed (Table 4). A theoretical upper limit estimate that includes all undetermined intent deaths plus all accidental poisoning deaths has been the standard in several Belgian analyses (e.g., Moens, 1985). Rockett, Samora, and Coben's (2006) lower limit estimate combined all suicides and undetermined intent deaths in the United States, while their upper limit estimate added all unintentional poisonings and drownings. In contrast, the Kolmos and Bach (1987) Scan-

TABLE 2
Injury-Ice External Cause of Injury Case Definitions: ICD-9 and ICD-10 Equivalent Codes

	ICD-9			ICD-10		
Mechanism/Cause	Self-inflicted	Undetermined	Accidental	Self-inflicted	Undetermined	Accidental
Cut/pierce	E956	E986	E920.0-.0	X78	Y28	W25-W29, W45
Drowning/submersion	E954	E984	E830.0-.9 E832.0-.9 E910.0-.9	X71	Y21	W65-W74
Fall	E957.0-.9	E987.0-.9	E880.0-E886.9, E888	X80	Y30	W00-W19
Fire/burn	E958.1,.2,.7	E988.1,.2,.7	E890.0-E899, E924.0-.9	X76-X77	Y26-Y27	X00-X19
Fire/flame	E958.1	E988.1	E890.0-E899	X76	Y26	X00-X09
Hot object/substance	E958.2,.7	E988.2,.7	E924.0-.9	X77	Y27	X10-X19
Firearm	E955.0-.4	E985.0-.4	E922.0-.9	X72-X74	Y22-Y24	W32-34
All Transport[1]	E958.5-6	E988.5-6	E800-E807 E810-E819 E820-E825 E826-E829.1	X82	Y32	V01-V99
Poisoning	E950.0-E952.9	E980.0-E982.9	E850-E869.9	X60-X69	Y10-Y19	X40-X49
Suffocation	E953.0-.9	E983.0-.9	E911-E913.9	X70	Y20	W75-W84
Other specified and classifiable	E955.5,.6, .7,.9 E958.0,.4	E985.5,.6,.7 E988.0,.4	E846-E848, E914-E915, E918, E921.0-.9, E923.0-.9, E925.0-E926.9, E929.0-.5	X75, X81, *U03.0	Y25, Y31	W23, W35-41, W44, W49, W85-W91, Y85
Other specified, not elsewhere classifiable	E958.8, E959	E988.8, E989	E928.8, E929.8	X83, Y87.0	Y33, Y87.2	X58, Y86
Unspecified	E958.9	E988.9	E887, E928.9, E929.9	X84, *U03.9	Y34, Y89.9	X59
All injury	E950-E959	E980-E989	E800-E869, E880-E929	X60-X84, Y87.0 *U03	Y10-Y34,Y87.2, Y89.9	Y10-Y34,Y87.2, Y89.9
ILL-DEFINED CAUSE CODES						
All Diseases and Conditions	799.9			R99		

ICE—International Collaborative Effort for Injury Statistics
ICD-9 International Classification of Diseases and Injuries, 9th edition
ICD-10 International Classification of Diseases and Injuries, 10th edition
Source: Centers for Disease Control and Prevention (1997); Injury-ICE (2002).

dinavian analysis constructed upper limit estimates that included all suicides, deaths of undetermined intent and ill-defined cause, accidental poisonings, shootings, stabbings, falls where the decedent was over 15 years of age, and drowning (except drownings occurring on vessels at sea).

The potential suicide misclassification rate for a given mortality data set is assumed to be related to the magnitude of the difference between lower and upper rate estimates (Rockett & Smith, 1995). However, not all deaths in a given nonsuicide category are misclassified suicides, and the method has been challenged on the basis of the additional error it therefore introduces into the rate estimation process (Sainsbury & Jenkins, 1982). An alternative method by which to estimate the rate of misclassification within nonsuicide mortality categories is the "peak time" analysis. Phillips and Smith (1991) suggest that rates of suicide seem to peak at symbolic ages across the lifespan (e.g., when one turns 30, 40, etc.), presumably because such ages represent a time to take stock of one's life, its purpose, and trajectory. By averaging the number of deaths occurring among individuals one year older and one year younger than a symbolic age in a nonsuicide cause category assumed to contain misclassified suicides, the expected number of deaths in that cause category among symbolic age decedents can be estimated. The number of deaths exceeding the expected number is taken as the number of misclassified suicides (Mohler & Earls, 2001; Phillips & Ruth, 1993).

While upper limit estimating techniques have demonstrated utility, the method suffers from the absence of a gold standard against which to validate findings. The demonstrated responsiveness of upper limit estimates to events that *should* impact suicide rates—such as the 1968 introduction of *ICD-8*'s undetermined intent death category—has been used as indirect evidence of the validity of these estimates (Sainsbury & Jenkins, 1982; Speechley & Stavraky, 1991; Warshauer & Monk, 1978). Some researchers have assumed that when upper and lower limit estimates trend together, large numbers of misclassified suicides must be influencing upper limit rate performance. However, as Mohler and Earls (2001) have pointed out, individuals with similar patterns of risk die by accident, suicide, and homicide, such that these comparisons are therefore confounded by these similar risk characteristics.

ESTIMATING STAGE TWO ERROR: MEASURING CODIFICATION PROBLEMS

Error also occurs with regularity during Stage Two suicide rate development process which includes the steps involved in death certification and codification; that is, completion of the official death certificate, application of *ICD* codes to multiple cause death certificate data, and identification of underlying cause of death for official reporting purposes (Bradshaw et al., 2006; Mathers, Ma Fat, Rao, & Lopez, 2005; Shibuya, Scheele, & Boerma, 2005). While significant error is known to occur at these points in mortality data processing, Stage Two error has not often been quantified.

The multiple-cause death certificate protocol advocated by the WHO is used by almost all nations (WHO, 1989). It requires death examiners to list the chain of medical events leading to death, beginning with the immediate cause of life cessation and working backwards sequentially. The last, or least proximal, cause listed is identified as the "underlying" cause, defined by WHO as the "the disease or injury event which initiated the train of morbid events leading directly to death; the circumstances of the accident or violence which produced the fatal injury" (WHO, 1989). Addressing the initiating link in the causal chain (i.e., the underlying cause of death)—either directly or through prevention of its known precipitants—is assumed to have the greatest public health preventive value (K. Kochanek, personal communication, September 2008). Most annual national mortality reports, therefore, focus exclusively on rates of deaths by underlying cause, rather than by multiple- or immediate-cause listings.

TABLE 3

Examples of Stage One Expert Case Review Studies

Author/Number of cases examined	Classification of cases examined: Suicides	Classification of cases examined: Undetermined/Open	Classification of cases examined: Other	Case Inclusion Criteria	Official/Study Standard of Proof for Suicide Classification	Underestimation Rate
I. Ireland						
McCarthy, 1975 (McCarthy & Walsh, 1975) *n* = unreported	All	All	All	No stated inclusion/exclusion criteria for cases examined.	Official Standard of Proof—Coroner, with mandatory inquest. Study conducted when suicide was a criminal offense in Ireland. Study Standard of Proof—(a) Direct evidence in the form of communication of intent, written or verbal or via inference based on the manner of death; (b) Indirect evidence in the form of recent mental health problems (esp. depression) as reported by family or medical records or past history of suicide attempts and death under suspicious circumstances	All deaths investigated by Dublin county or city coroners, 1964–1968. Suicides *underestimated by 379%* [For case review, official suicide rate: 1.4/100,000 (58 cases identified as suicides out of unreported total *n* reviewed); expert case review rate: 5.3/100,000 (210 total cases identified as suicides)]

Clarke-Finnegan, 1983 (Clarke-Finnegan & Fahy, 1983) $n = 410$ cases (County Galway), plus replication using 72 cases (County Kerry)	All	All		Only records with a postmortem examination were evaluated, but this apparently included all suicides and all undetermined intent deaths.	Official Standard of Proof—Coroner in a system where registering deaths is not mandated (approximately 6.1% of all Galway deaths during the study period were neither registered or certified as to cause). Study Standard of Proof—Clinical judgment of a research psychiatrist for selection of cases. Suicide categorization based on: (1) self-poisoning or self-injury recorded by pathologist as cause of death; (2) history of mental illness with or without suicidal behavior in association with self-poisoning/injury; (3) "clinically suspicious circumstances with initial paucity of information. " Validation of selection procedure done using all identified suicides plus 7 'dummy' cases where death was due to violent injury or drowning but not classified as suicides. These were rated independently by additional pathologists and psychiatrists. Validation done using a visual analogue scale, divided into quadrants. High agreement described.	All Galway city and county 1978 cases with a postmortem examination; attempted replication using all 1978 cases from County Kerry with a postmortem examination. *Suicides underestimated by 285% for Galway. under the official standard.* [For case review, official suicide rate: <1% (2/410) expert case review rate: 5.4% (22/410 cases).] *Suicides underestimated by 300% for Kerry under the official standard.* [For case review, official suicide rate: 4.2% (3/71) expert case review rate: 12.7% (9/71 cases).]

(continued)

TABLE 3
Continued

Author/Number of cases examined	Classification of cases examined: Suicides	Undetermined/ Open	Other	Case Inclusion Criteria	Official/Study Standard of Proof for Suicide Classification	Underestimation Rate
II. Scotland						
Ovenstone, 1973 (Ovenstone, 1973) $n = 214$ cases	All	All	Some accidental	Records for all decedents aged 12 and over where the decedent had a resident address in Edinburgh at the time of death and in which the area Procurator Fiscal sent a précis of findings to the Crown Office. Included some accidental deaths, chosen as follows: (1) All where there was a record of prior psychiatric treatment or suicidal communications; (2) all carbon monoxide poisonings; (3) all where researcher considered suicide a possibility; (4) random sample beyond those chosen by criteria 1–3.	Official Standard of Proof—Hybrid Legal System Study Standard of Proof—(a) Direct evidence: written or verbal communication; manner of death "e.g., hanging, shooting, and . . . size of overdose judged from post mortem evidence, tablet hoarding, etc" (Ovenstone, 1973). Suicide was classified with evidence of deliberate initiation of an act of self-harm irrespective of whether there was evidence of intent to die.	All Edinburgh (Scotland) deaths for which a précis was available in the Crown Office, October 1969–March 1971. Suicides *underestimated by 50% by Crown Counsel* [For case review, official suicide rate: 30.8% of examined cases (66/214 cases); expert case review rate: 49.5% (106 / 214).]
III. England						
Cooper, 1995 (Cooper & Milroy, 1995) $n = 653$ cases	All	All		No stated inclusion/exclusion criteria for cases.	Official Standard of Proof—Coroner System Study Standard of Proof—"a) suicide note, b) 'prima facie' evidence of intent; c) circumstantial evidence as to the most reasonable and probable explanation for the death" (p. 320).	All South Yorkshire (UK) suicides and open verdicts, 1985–1999. *Suicides underestimated by 39.7%* [For case review, official suicide rate: 57% of examined cases (323/653 cases); expert case review rate: 82% (536/653).]

Linsley, 2001 (Linsley, Schapira, & Kelly, 2001) $n = 373$ cases	All	All	No stated inclusion/exclusion criteria for cases examined.	Official Standard of Proof—Coroner System Study Standard of Proof—Not explicitly stated. Categorization was to "probable suicide," "suicide unlikely," or "suicide impossible." Examples given of unlikely group classification were: 'a young man who in a state of solvent intoxication fell from a window;' 'an elderly lady with dementia who was prone to wander, found drowned in a small stream.'	All Newcastle upon Tyne (UK) suicides and open verdicts, 1985–1994. *Suicides underestimated by 42.6%*.[For case review, official suicide rate: 50.4% of examined cases (188/373 cases); expert case review rate: 93.0% (347/373).]
IV. Finland					
Ohberg, 1998 (Ohberg & Lonnqvist, 1998) $n = 190$ cases	All	All	No case inclusion/exclusion criteria stated.	Official Standard of Proof—Medical Examiner (Forensic Pathologist) System Study Standard of Proof—Indirect evidence of intent, defined as any of the following, per Rosenberg, 1988: preparations for death; unusual farewell; desire to die; hopelessness; learning about means of death; avoiding rescue; previous suicide attempt; previous suicide threat; significant losses; depression, alcohol dependence or abuse, or other mental disorders as indicated by a diagnosis for which the patient had received treatment by a specialist.	All undetermined intent deaths in Finland, April 1987–March 1988. Overall national *suicide rate possibly underestimated by 10% due to omission of undetermined intent deaths with indirect evidence of intent*. [For case review, official suicide rate: 0% of examined cases (0/190 cases); expert case review rate: 27.9% (53/190).]

(continued)

TABLE 3
Continued

Author/Number of cases examined	Classification of cases examined: Suicides	Undetermined/ Open	Other	Case Inclusion Criteria	Official/Study Standard of Proof for Suicide Classification	Underestimation Rate
V. Australia						
Cantor, 2001 (Cantor, McTaggart, & De Leo, 2001) n = unknown	All	Some (unk)	Some (unk)	All incoming cases analyzed; included those accidental deaths that "could be suicides."	Official Standard of Proof—(Local) Coroner System Study Standard of Proof—Counted as suicides cases classified as 1) "beyond reasonable doubt"—i.e., "strong indicators (one or more required): verbal or other suicide communication, behaviors displaying suicidal intent, measures taken to ensure death; and weak indicators (two or more required): expression of severe distress, past suicide attempts or threats, recent negative life events, history of mental illness, elevated drug levels at post-mortem" OR 2) "probable" suicides—between 50–90% rater certainty that the death was a suicide.	All Queensland (Australia) deaths referred to the coroner as possible suicides, 1990–1995. *Suicides underestimated by 5.5%* (total number of cases reviewed is not reported).

VI. United States						
Litman, 1963 (Litman, Curphey, Shneidman, Farberow, & Tabachnick, 1963) $n = 100$ cases	Some	Some		Cases reviewed were called "equivocal cases," defined as cases "in which suicide is a possibility but in which there could be more than one interpretation and, therefore, the decision is uncertain and doubtful."	Official Standard of Proof—Combined Medical Examiner-Coroner System Study Standard of Proof—Clarity of intention and decisiveness of the lethal action, established via psychological autopsy-based decision made during "Death investigation review," where details such as the following were reviewed: level of substance in system from tox/labs; history of depression; recent life circumstances; circumstances surrounding the death obtained by interview from survivors.	100 consecutive "equivocal" (ambiguous intent) cases from Los Angeles County, referred by the county Medical Examiner-Coroner in the latter half of 1959 and the first half of 1960. *"Equivocal" Suicides underestimated by 3%*. [For case review, official suicide rate: 52% of referred cases (52/100 cases); expert case review rate: 55% (55/100, but 19 cases were reclassified by death investigation review).]
Brent, 1987 (Brent, Joshua, Perper, & Allman, 1987) $n = 314$ cases	All	All (45%)	Some (9%)	Included accidental death cases that were "questionable—i.e., those certified as accidents, but in which the circumstances of death did not exclude suicide." Excluded cases where description of death circumstances contained insufficient description of death with reference to intent to die (e.g., absence or presence of a suicide note) or evidence of prior preparation and at least one of the following: 1) any history of stressors, 2) psychiatric history and report of psychological state prior to death, 3) description of prior suicidal behavior and threats.	Official Standard of Proof—Coroner System Study Standard of Proof—"1) [Suicide could be achieved by] comparable method [of injury], 2) circumstances suggesting intent to die, 3) presence of one of following: a) recent stressor, b) psych history, c) past suicidal behavior."	All Allegheny County, Pennsylvania, adolescent decedents, ages 10–19. *Adolescent suicides underestimated by an average of 23.9%/year*. [For case review only, official suicide rate: 50.6% of qualifying cases (159/314 cases); expert case review rate: 62.7% (197/314).]

(continued)

TABLE 3
Continued

Author/Number of cases examined	Classification of cases examined: Suicides	Undetermined/ Open	Other	Case Inclusion Criteria	Official/Study Standard of Proof for Suicide Classification	Underestimation Rate
Donaldson, 2006 (Donaldson, Larsen, Fullerton-Gleason, & Olson, 2006) $n = 84$ cases reviewed and modeled	Poisonings only	Poisonings only (21%)	Poisonings only	Used all cases reported to the National Violent Death Reporting System—Standardized variables extracted from centralized toxicology reports, narratives, and other evidence collected by medical examiner and coroner's offices (>20 variables collected).	Official Standard of Proof—State-level Medical Examiner; County Coroner System Study Standard of Proof—Statistical modeling. Used classification and regression tree (CART) analysis to create classification rule by recursive partitioning of predictor variables. Random forests classification was used to evaluate the stability of CART results.	Random selection of 50% of undetermined and unintentional Utah poisoning deaths occurring in 2002. *Poisoning suicides (overall) underestimated by 10%; intentional self-poisonings underestimated by 30%.* Probable suicides among undetermined intent poisoning deaths underestimated by 21% [For case review only, official suicide rate: 0% of examined cases (0/84 cases); expert case review rate: 21.4% (18/84).]
Carr, 2004 (Carr, Hoge, Gardner, & Potter, 2004) $n = 124$ cases	All	All (60% of convenience sample)	All homicide; Some accidental cases. (6% of convenience sample)	(Expert case review only) Inclusion 1) Accidental death caused by handgun, overdose, asphyxia, drowning, or falling, 2) homicide, 3) undetermined intent deaths; and 4) all deaths of any category occurring within 30 days of a service member's retirement to determine whether death was directly related to an illness or injury sustained while on active duty. Exclusion—Previously identified suicides ($n = 286$); motor vehicle accident deaths.	Official Standard of Proof—Medical Examiner System Study Standard of Proof—(from Brent, 1987, as above) 1) specific evidence of intent communicated verbally or in writing, or 2) past psychiatric history, or 3) both. Also ultimately developed from.	All active duty U.S. military decedents, 1998–99. *Active duty military suicides underestimated by 16.8%*, with an additional 4% of deaths "suspicious for suicide." [For case review only, official rate: 17.8% of all cases (286/1,604 cases; expert case review of some cases, plus Department of Defense Medical Mortality Registry additional suicides after close of death investigations: 18.1% (334/1,844).]

Effective January 1, 2003, WHO coding guideline 4.2.2 (o) of the "Rules and Guidelines for Mortality and Morbidity Coding," required that suicide always be considered the underlying cause of a death due to intentional self-harm (Update Reference Committee, 2000; WHO, 2004). In other words, even if substantial psychiatric disability precedes a suicide and treatment of this disability might have averted the suicide, it is not acceptable to list a psychiatric diagnosis as the underlying cause of a suicide. Likewise, if a person with brain cancer (e.g., *ICD*-10 code D43.2—neoplasm of uncertain or unknown behavior of brain, unspecified) dies by firearm-related suicide and the intent of the firearm injury has been verified as suicidal in an individual with capacity, the underlying cause of death will be listed as suicide.

Unfortunately, there is substantial evidence to suggest that some portion of the death certifiers in almost all countries are poorly trained in the process of conceptualizing multiple-cause mortality data (Hill, 2006; Koehler et al., 2006; Mathers et al., 2005) and death certificates are therefore often completed incorrectly. When certificate data are incomplete, implausible, or reported out of sequence, cause listings must be reworked during the official death registration and codification process. This reworking requires medical expertise, sound judgment, and a thorough knowledge of *ICD* coding rules. It is typically done manually by professional nosologists. Large numbers of nosologists using idiosyncratic approaches to correct ambiguous cause listings represent an additional source of error within mortality statistics.

Instead of manual coding, the vital statistics systems in nineteen countries (Australia, Brazil, Canada, England, Estonia, France, Germany, Hong Kong, Hungary, Ireland, Mexico, Scotland, South Africa, Spain [Barcelona], Sweden, Taiwan, United States, Wales, and part of Russia) now use the same set of computer software programs to code death certificate information (Anderson, 2006; Kochanek, personal communication, April 2007; Rosenberg & Kochanek, 1994). Called the Mortality Medical Data System (MMDS; seee Figure 2), these linked software packages have been developed and refined by the U.S. National Center for Health Statistics over a period of two decades. The software was donated to the WHO's International Collaborative Effort for Automating Mortality Statistics, and is offered to WHO member states at no cost. MMDS has the ability to process and correct word-for-word information taken from WHO-compatible death certificates in a standardized manner across mortality data sets.

Analysis of Death Certificate Completion and Processing Errors

All original death certificate information is retained in the final data set produced during MMDS processing, permitting comparisons to quantify both death certificate completion errors and errors created during processing and correction of data (Lawrence, Miller, & Spicer, unpublished manuscript; Lawrence, Miller, & Weiss, 2002). Analysis begins with the identification of every death classified as a suicide by either of two criteria: (1) when listed as the underlying cause of death before processing (i.e., as listed originally by the medical examiner/coroner on the death certificate), and (2) when listed as the underlying cause of death after automated processing. A line-by-line inspection of sequentially ordered multiple cause listings for cases with discrepant underlying causes (suicide, nonsuicide) from these two sources can be used to determine the most likely "correct" underlying cause for each case (see Lawrence et al., 2002, for a description of the preliminary reordering technique required for this analysis). The number of probable suicides identified as such on the original death certificate but changed during automated processing represents the rate of *automated processing error*. The number of probable suicides identified as such by the automated system but not listed as suicides on the original death certificate represents the rate of *death certificate completion error*. Both error rates can be examined by year to characterize trends in error rates.

TABLE 4
Examples of Stage One Upper Limit Estimate Studies

	Classification of cases examined								
Author/Year	Suicides	Undetermined Open	Accidents	Homicides	Ill-defined and Unknown Cause'	Categories (if any) of Accidental Deaths studied	Composition of Lower/Upper Limit Estimates	Study Standard of Proof	Underestimation Rate
I. Eleven Country Analysis									
Fingerhut, for the International Collaboration Effort on Injury Statistics (Fingerhut, Cox, & Warner, 1998)		All				Percent of undetermined intent deaths as a percentage of all injury deaths.	One rate only	Degree to which a country utilizes "undetermined intent" as the underlying cause of death	Eleven European and North American countries; 1–10 yrs data each. *Undetermined intent deaths constituted 1–2% of all injury deaths in six participating countries* (Australia, Canada, The Netherlands, New Zealand, Norway, and United States). *Undetermined intent deaths constituted 5–10% of all injury deaths in five countries* (Israel, Denmark, France, Scotland, and England & Wales).

II. England and Scotland								
Barraclough, 1973 (Barraclough, 1973)	All	All				Lower Limit—Rates of suicide for each country. Final Upper Limit—Assumed undetermined deaths that were really suicide would be in proportion to the ratio of suicide to accidental death. Used simple Proportion methods to add to the suicide rate that portion of the undetermined rate represented by this ratio.	Assumed that the rates of suicide for the two countries would be more similar than different, and that the differences in official rates were largely a function of differences in the reporting systems. So degree to which rates matched after adjustment was standard of proof in study.	Suicide statistics published by the Registrars General of Scotland and England, 1968—all deaths for persons aged 15 and older. *Theoretical upper limits suggest that suicide rates underestimated in Scotland in 1968 by as much as 30.7%.* *Theoretical upper limits suggest that suicide rates underestimated in England in 1968 by as much as 15.8%.*
III. Scandinavian Countries								
Kolmos, 1987 (Kolmos & Bach, 1987)		All	Some		Accidental poisonings, drowning, shooting, stab wounds, jumps/falls, cutting, firearm, and hanging deaths.	*Lower Limit*—Rate of suicide. *Upper Limit*—Rate of suicide and drowning plus "ill-defined signs and symptoms."	Comparison of international rate differences before and after upper limit correction	Four Scandinavian countries—Norway (1978 data), Denmark (1978), Finland (1976), and Sweden (1978). *Theoretical upper limits, after adjustment, suggest that suicides were not significantly underestimated for any country* after examination of alternative rate trends and extant literature.

(continued)

TABLE 4
Continued

Author/Year	Classification of cases examined: Suicides	Undetermined Open	Accidents	Homicides	Ill-defined and Unknown Cause'	Categories (if any) of Accidental Deaths studied	Composition of Lower/Upper Limit Estimates	Study Standard of Proof	Underestimation Rate
IV. Belguim									
Moens, 1984 (Moens, 1985)	All	All	Some		All	Accidental poisonings	*Lower Limit*—Rate of suicide *Upper Limits*—Rate of suicide, plus undetermined intentional deaths and accidental poisonings.	Estimates of "potential misclassification" derived by calculating % difference between upper and lower limit estimates.	Belgium, 1968–72; 1989–81 *Theoretical upper limits suggest that suicides underreported by as much as 16–19% for males and 25–37% for females.*
V. United States									
Rockett, 2006 (Rockett, Samora, & Cohen,**Au: Coben in refs** 2006)	All	All	Some			Unintentional poisonings and drownings.	*Lower Limit*—All suicides, plus all undetermined intentional deaths. *Upper Limit*—Lower limit deaths, plus all unintentional poisonings and drownings.	Ratio of lower to upper limit estimates by racial/gender groups. Assumed that comparable ratios should occur if differential misclassification were not occurring.	U.S. suicides for all individuals, aged 15 and older, 1999–2002. *Theoretical upper limits suggest that suicide rates underestimated by as much as 30% for Black males vs. White males, and 38% for Black females compared to White females.*

Mohler, 2001 (Mohler & Earls, 2001)	All	All	Some	All	Accidental barbiturate poisoning, single-car crashes, pedestrian deaths.	*Lower Limit*—Official rate. *First upper limit estimate* used the Phillips (1993) method with suicides, and five accidental death causes, plus undetermined and ill-defined cause deaths. *Second upper limit estimate* used only suicides plus undetermined intent deaths.	As per Phillips (1993): Excess deaths in any category at "symbolic" ages (e.g., 20, 30. 40, etc.) are considered suicides. Estimates of misclassification are based on comparison of observed "excess" death rates, and expected rates (the average of deaths at the age before and the age after the symbolic age).	U.S. Adolescents, ages 10–24, 1979–1994. *Theoretical upper limits suggest that suicide rates under first upper limit estimate underestimated by as much as 18% for Black females, 11% for White females, 10% for Black males, and 6% for White males.*
Kleck, 1998 (Kleck, 1988)	All	All	All			Lower Limit—Rate of suicide *First Upper Limit*—All deaths with causes representing same mechanism of death as the suicide. *Final Upper Limit*—Rate of suicide, plus the most likely misclassified deaths, based on literature review, and adjusted for overcounting of suicides in some categories.	Validation from broad-based literature review and/or ratios between rates that implicated misclassification.	U.S. Suicides, 1980. *Theoretical upper limits suggest that suicides underreported by a maximum of approximately 7.2%, after adjustment for most likely under- and over-counting in various categories.*

(continued)

TABLE 4
Continued

Author/Year	Classification of cases examined: Suicides	Undetermined Open	Accidents	Homicides	Ill-defined and Unknown Cause'	Categories (if any) of Accidental Deaths studied	Composition of Lower/Upper Limit Estimates	Study Standard of Proof	Underestimation Rate
Phillips, 1993 (Phillips & Ruth, 1993)	All	All	Some		All	Accidental barbiturate poisoning, single-car crashes, pedestrian deaths.	*Expected number of deaths*—Average number of deaths for year older and younger than "symbolic" age (e.g., for symbolic age "turning 30," expected number of deaths is the average of deaths for 29- and 31-year-olds. *Observed number of deaths*—Rate of all suicides and undetermined intent deaths, plus all accidental poisonings and all drownings.	Excess deaths in any category at "symbolic" ages (e.g., 20, 30. 40, etc.) are considered suicides. Estimates of misclassification are based on comparison of observed "excess" death rates, and expected rates (the average of deaths at the age before and the age after the "symbolic" age).	California, 1966–1990. *Theoretical upper limits suggest that 24-year suicide rates underreported by as much as 7.9% for Whites; 11.8% for Blacks, 6.3% for males; 11.0% for females.*

Warshauer, 1978 (Warshauer & Monk, 1978)	All	Compared published "official" rates of suicide by race with the suicide records of the medical examiner (ME) who originally had analyzed information about each case. ME cases included cases considered to be suicides, but not meeting official criteria or cases overlooked because final disposition was not requested.	*Lower Limit* estimate consisted of the official rates, broken down by rage/gender. *Upper Limit* estimates consisted of rates so classified using medical examiner case files. Rates were compared immediately before and after introduction of *ICD-8*	Medical examiner classification rates were considered the gold standard in this study. Changes in the ratios of Black to White official suicide rates before and after introduction of *ICD-8*'s undetermined intent category in 1969.	Suicides among decedents 18 years and older occurring in four health districts in New York City, including East Harlem in Manhattan and three districts in the South Bronx. *For the first two years of* ICD-*8 use (i.e., in 1969–70), theoretical upper limits suggest that the combined official rates of suicide in four NYC health districts were underestimated by as much as 45% for Whites and by as much as 80% for Blacks*. These differences appeared to be primarily related to differences in preferred methods of suicide.

(continued)

TABLE 4
Continued

Author/Year	Classification of cases examined: Suicides	Undetermined Open	Accidents	Homicides	Ill-defined and Unknown Cause'	Categories (if any) of Accidental Deaths studied	Composition of Lower/Upper Limit Estimates	Study Standard of Proof	Underestimation Rate
VI. Canada									
Speechley, 1991 (Speechley & Stavraky, 1991)	All	All (after 1968)	Some	All	All	Drowning, jumping/falling, firearms, poisonings and "all others."	*Lower Limit*—Rate of suicide, by year. *Upper Limit*—Rate of suicide, plus (after 1968) rate of suicide plus all undetermined intent deaths.	Estimates of "potential unevenness" of rates due to misclassification derived by calculating % difference between suicide rates and suicide plus all undetermined intent deaths after 1968.	Canada, 1950–1982. Underreporting calculated only for years after 1968. *Theoretical estimates of the potential unevenness of suicide rates suggest that suicide rates may be underestimated by as much as 17.5% for females and 12% for males annually between 1969 and 1982.*

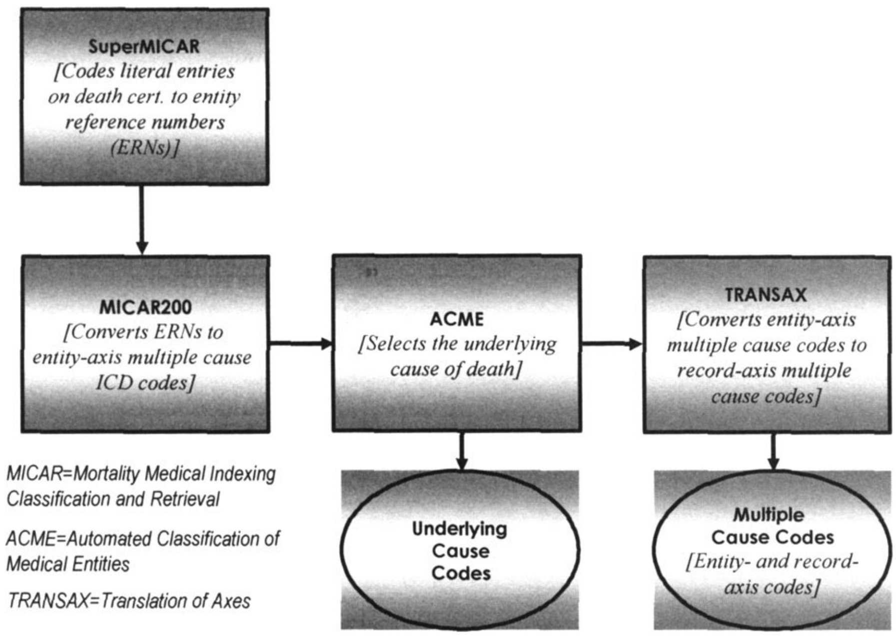

Figure 2. Mortality medical data system (MMDS)

Source: Anderson (2006).

For example, when applied to 1999–2004 U.S. mortality data, the Lawrence et al. methods reveal that, in 4,625 cases, MMDS software correctly changed a death certificate's nonsuicide underlying cause listing to suicide, suggesting that there would have been an average annual 2.5% underreporting of U.S. suicides for those years based on uncorrected death certificate cause listings (death certificate completion error). All misclassified cases were corrected during MMDS automated processing. In 324 other cases the software changed deaths originally attributed to suicide on the death certificate to another underlying cause. In 90 cases, the underlying cause was changed from suicide to a mental illness diagnosis (63 of these were depressive disorder, seven of which were made after the January 1, 2003, rule change addressing this issue). In 61 other cases the MMDS system changed an underlying cause of suicide to brain hemorrhage or stroke (58 of these errors were likely firearm suicides). This second pattern remained stable across study years. The average automated coding error would, therefore, be 0.15% (6-year average = 54 cases/year, trending downward across time; 1999–2001 = average 75 cases/year; 2002–2004 average = 33 cases/year). No other studies of this type are available. The validity of the Lawrence et al. method is tied closely to the specificity and quality of the decision rules used to determine cause during line-by-line inspection of discrepant cases, and further description of these rules has yet to be published.

ESTIMATING STAGE THREE ERROR: QUANTIFYING ERROR OCCURRING DURING PREPARATION OF OFFICIAL MORTALITY STATISTICS

Three types of error commonly occur during preparation of all-cause, annual mortality reports.

Error Due to Short Reporting Deadlines

In the developed world, suicides are among the most highly investigated of all deaths, and final-cause decisions are commonly withheld until death investigations are completed (Corcoran et al., 2006; Cui et al., 2004). This is one reason that publication of national mortality reports is often delayed for a substantial amount of time after the close of a given reporting year—a practice that is not helpful for monitoring and surveillance purposes. However, as noted earlier, the impact of short reporting deadlines on suicide statistics can also be profound.

To quantify the theoretical relationship between reporting deadline and relative completeness of annual suicide counts in Hong Kong, Cui et al. (2004) developed the following analytic technique. First, a reporting delay distribution for a given mortality data set is derived using retrohazard regression modeling (Gross & Huber-Carol, 1992). The suicide incidence (representing the "true" number of suicides occurring within a given time unit as classified by death investigators) is estimated from the reporting delay distribution, under the premise that individuals who kill themselves by the same method within the same coroners' district during a given time unit will have nearly the same reporting delay distribution. Estimates of data completeness by length of time to reporting deadline are modeled using a Horvitz-Thompson-type estimator (Horvitz & Thompson, 1952; Kalbfleisch & Lawless, 1988). (Horvitz-Thompson techniques estimate variance in samples drawn from a finite universe of cases when unequal selection probabilities are present.) The total estimated number of suicides occurring during the reporting time period is also established, based on the number of reported cases, using an adjustment to the Horvitz-Thompson procedure. Accurate use of the model requires that both the number of medical examiners practicing in any given year and the rate by method of injury be considered, as both parameters impact reporting delay (Cui et al., 2004). In ongoing analyses, estimates of delay functions also need to be updated periodically to reflect changes occurring within the reporting system.

For example, for 1999 Hong Kong suicides, a year 2000 reporting deadline set 4 months after the end of the year was associated with 77% suicide case registration, while a 5-month deadline resulted in 85% case registration, and a 6-month deadline was associated with 91% case registration (Cui et al., 2004). Data were essentially complete with an 11-month reporting deadline. No study other than Cui et al.'s work has been published using the model, and no reliability/validity procedures have been developed to assess the method.

Error Due to Quality of Mortality System Data

Missing and ambiguous data pose a substantial challenge to national mortality data quality, and several data loss indices have been used to characterize this Stage Three error. A *population completeness index* reflects the extent to which events (births, deaths) occurring within the vital events registration system's geographic coverage area are actually represented in the data set. A *population coverage index* is used to identify the percent of the population within the vital event system's coverage area that does not routinely participate in the reporting process, and a *mortality data quality index* is used to characterize the relative precision of *ICD*-based mortality diagnoses.

Although this article focuses exclusively on the suicide rate error found within countries with established vital events registration systems, these resource-intensive systems are essentially complete in only 64 of the world's 193 countries (Mathers et al., 2005). As a result, only one third of the approximately 60 million deaths occurring worldwide each year are systematically recorded (Soleman, Chandrimohan, & Shibuya, 2006). Even in countries that do have comprehensive vital events registration systems, highly mobile groups such as the Tanka of China, Hispanic field workers of the southwestern United States, and Roma populations of Eastern Europe pose a substantial challenge to the accuracy of population-based statistics. The census of relevant migratory groups not consistently reporting vital events is used to calculate the *population coverage index*.

In addition, the responsibility for reporting deaths not occurring within medical facilities falls to bereaved family members in some countries, with minimal consequences for not completing the process. Where the cause of death is stigmatizing, as in the case of suicide, there is little incentive to carry through with reporting. The estimated incompleteness of vital events reporting created under such conditions is used to calculate a vital events *population completeness index*.

Finally, deaths classified as of "ill-defined cause" (i.e., *ICD*-10 R99, *ICD*-9 799.9: "Other ill-defined and unspecified causes of mortality") represent a group of imprecise cause listings with poor data quality. These *ICD*-10 R-codes are included in the nosological framework to designate deaths for which cause is poorly understood. Some R-code diagnoses are so vague that inferences about the actual underlying cause of death cannot be made. In regions where a death certificate has to be completed before burial, some R-codes are sometimes used as a "placeholder" diagnosis on death certificates while forensic investigation is underway (Kochanek, April 2007; Warshauer & Monk, 1978). If these placeholders are not updated in the official mortality database after investigation is complete, they become the official underlying cause for reporting purposes, thereby limiting the public health utility of the data (Hlady & Middaugh, 1988). A recent Stage One expert case review study in France found that 25% of all ill-defined cause deaths were actually suicides (Andriessen, 2006; Jougla, Pequignot, Chappert, Rossollin, Le Toullec, & Pavillon, 2002). If this were also true in the United States, the impact on national suicide rates would be substantial. Between 1999 and 2005, 105,944 deaths (17,657/year average) were classified as *ICD*-10 R99 deaths (NCHS, 1999–2005). If 25% of these deaths were actually "hidden" suicides, an average increase of 12.2% per year (3,784 additional suicides) would be added to the national count for each of these 7 years.

Stage Three suicide rate error is commonly a function of the estimated proportion of a nation's resident population not covered by its vital events registration system, the estimated portion of unrecorded deaths in the covered population, and plus the proportion of deaths for which cause listings are ambiguous (Johansson et al., 2006). Mathers et al. (2005) have suggested that data loss statistics reported alongside annual mortality rates would serve as a measure of the quality of the death data upon which the rates were based.

Error Inherent in Inter- and Postcensal Population Estimates

For ease of comparison, official suicide statistics are often reported as population-based rates such as the crude rate per 100,000 citizens, or the total estimated years of life lost due to suicide within a specific population group. In non-census years, this convention requires the use of an estimated (denominator) base population. Postcensal population estimates reflect the estimated population for years occurring after the last official census, while intercensal population estimates apply to years between two completed censuses. Postcensal estimates contain the higher error. The difference between the actual count during a census year and the postcensal estimate for that year is known as the "error of closure." The error of closure is not likely to be distributed uniformly by age, race, ethnicity, or gender, and reflects the degree to which mortality rates calculated with postcensal denominator population estimates are incorrect (Best & Wakefield, 1999). In the year 2000, for instance, 6.8 million more citizens were living in the United States than had been estimated for that year based on 1990 postcensal projections. When suicide rates are based on postcensal estimates, this should be noted, along with the relevant error of closure.

DESCRIBING OVERALL LEVEL OF ERROR BY STAGE OF RATE DEVELOPMENT: A THEORETICAL MODEL

The staged approach to modeling multistage rate error can be represented theoretically as:

$$bias = A + C^{pa} + P + R^{pc}$$

where

A = *Error within the death classification process (Stage One),*

C^{pa} = *Error within the death codification process, given the probability of any given set of classification processes (Stage Two),*

P = *Error within population estimates (Stage Three), and*

R^{pc} = *Error related to annual cause of death reporting procedures (Stage Three), given the probability of any (Stage Two) death codification procedure, and set of (Stage One) classification processes*

This model expresses the total suicide rate error found within a given mortality data set (*bias*) as a function of multiple processes occurring across the phases of data collection and data streams. Relevant case review and upper limit estimates form the basis for Stage One case underascertainment and misclassification calculations (A), and the degree of Stage Two certificate and/or automated processing error (C^{pa}) is related to the probability of these underascertainment/misclassification estimates for Stage One. Stage Three postcensal population estimates (P) represent an independent source of error, as do uncovered populations and incomplete reporting (R^{pr}).

DISCUSSION

Perhaps the most scathing critique of official suicide data published in the last century was written by Douglas (1967), who compared the "complex, varied, inconsistent and changing intellectual" definitions of suicide to the "common-sense level of thought at which most doctors, coroners . . . etc., work in deciding whether or not a death is a suicide." He concluded that, because of widespread and systematic bias, "official statistics are so greatly in error that they cannot be used for the scientific study of suicide" (pp. 229–230). Responding in part to Douglas's argument 20 years later, Pesconsolido and Mendelsohn (1986) developed county-level statistical models to assess the degree to which the social construction of suicide impacted U.S. suicide rates. While they did identify patterns consistent with systematic misreporting of suicide, they also demonstrated that this misreporting had "little discernible impact" on variables commonly used in sociological analyses. Publication of this landmark work was followed just 2 years later by development of operationalized criteria for use by field investigators in identifying and classifying U.S. suicides (Rosenberg et al., 1988). Programming of MMDS computer software routines was initiated around the same time. Although there is much left to do, these advances substantially increase confidence in the fidelity of suicide mortality data, and place the goal of accurate suicide rate estimation within reach.

Where should we go from here? To date, no country in the world has consistently produced high quality mortality data, and progress toward this goal over the past century has been slow at best. Therefore, in public health research, application of multistage, multisource error estimation models may represent an intermediate step toward increased measurement precision. The staged framework presented above is one such model, incorporating case underascertainment, misclassification, codification error, mortality system data loss, and inaccurate census estimates. This kind of working statistical model is increasingly feasible, and the public health import is substantial. Once developed, error-estimation tools could be applied annually in national reports in much the same manner as current postcensal population estimating models. Over time, reports of standardized estimates of error published alongside population-based suicide statistics would inform increasingly precise modeling strategies. Alternatively, calculating standard error rates and making them available for meta-analyses, even if not published, would likewise contribute to the creation of a knowledge base that could support ongoing improvement in population-based suicide metrics.

Apart from post hoc modeling strategies, our literature review also identified key study procedures that substantially impact the utility of findings. Adequate sample size is essential, and, as Geddes (1999) has pointed out, definitive studies of suicide phenomena need to be "at least an order of magnitude larger" than was the case in much early work. Three other methodological decisions also appear to be critical.

(1) Use of standardized suicide and nonsuicide cause category case definitions (in the data pool from which potentially misclassified suicides are to be identified) is essential to cross-study comparisons and meta-analyses. A related concern involves selection of the cause-specific mortality categories that should be routinely included in suicide case-finding studies. A minimum standard data set to be included in Stage One and Two suicide rate error studies would be of value. Presently, most Stage One and Two studies uniformly include undetermined intent cause category deaths (*ICD*-8 & *ICD*-9 E980–989; *ICD*-10 Y10–Y34), and consistent reporting of numbers of suicides found within this single category would be beneficial. Further research on suicides within the ill-defined cause of death category (*ICD*-9 780–799; *ICD*-10 R00–R99, excluding R95) also appears warranted (Kochanek, personal communication, April 2007; Phillips & Ruth, 1993; Warshauer & Monk, 1978). Finally, although additional research would be needed to develop a standardized list, past evidence suggests that certain accidental injury categories may contain substantial numbers of misclassified suicides, potentially including single-vehicle fatalities and unintentional poisonings, drownings, and falls (Adelstein & Mardon, 1975; Eaton, Messer, Garvey Wilson, & Hoge, 2006; Kelleher, Corcoran, Keeley, Dennehy, & O'Donnell, 1996; Kolmos & Bach, 1987; Mohler & Earls, 2001; Phillips & Ruth, 1993; Rockett et al., 2006).

(2) Designing and applying methods to test reliability and validity in Stage One underascertainment studies, as noted above, is necessary to establish the credibility of underascertainment estimates.

(3) Reporting standardized measures of data fidelity alongside population-based suicide rates is a time-honored practice for annual national mortality publications (e.g., Best & Wakefield, 1999; Kochanek & Smith, 2004), and the inclusion of such information in research studies examining population-based suicide rates would provide similar information about data quality. Candidate measures include indices of mortality system data loss (e.g., percent uncovered population; percent ambiguous cause coding), estimates of data completeness and coverage, annual reporting deadlines, and the relevant error of closure statistic from the most recent intracensal/postcensal population comparison.

CONCLUSIONS

To meet the challenges associated with global increases in rates of suicide, well-developed, credible population-based suicide metrics must be available both for surveillance purposes and to measure public health outcomes. Our review suggests that rate error may be characterized using a phased model that can accommodate the multisource, multistage error sources inherent in these data. Compared to current estimating methods, the logical framework underlying such a model permits more precise error quantification and enhanced measurement precision. If a standardized methodology based on such a framework were routinely applied to annual mortality data, the effort would eventually produce a lingua franca of comparator data. The result would be an improved platform upon which to conduct comparative analyses, moving the international suicide research community toward a time when sociological analyses could finally begin to generate high quality answers to those cross-national questions first raised by Durkheim so long ago.

REFERENCES

ADELSTEIN, A., & MARDON, C. (1975). Suicides 1961–74. *Population Trends*, 2, 13–64.

ANDERSON, R. (2006). *The international collaborative effort on automating mortality statistics*. Retrieved July 2007 from www.cdc.gov/nchs/ppt/duc2006/anderson_39.ppt.

ANDRIESSEN, K. (2006). Do we need to be cautious in evaluating suicide statistics? *European Journal of Public Health, 16*, 445.

APPLEBY, L., SHAW, J., & AMOS, T. (1997). National confidential inquiry into suicide and homicide by people with mental illness. *British Journal of Psychiatry, 170*, 101–102.

ATKINSON, M., KESSEL, N., & DALGAARD, J. (1975). The comparability of suicide rates. *British Journal of Psychiatry, 127*, 247–256.

BARRACLOUGH, B. (1973). Differences between national suicide rates. *British Journal of Psychiatry, 122*, 95–96.

BECK, A., RESNIK, H., & LETTIERI, D. (1974). *The prediction of suicide.* Philadelphia: Charles Press.

BEST, N., & WAKEFIELD, J. (1999). Accounting for inaccuracies in population counts and case registration in cancer mapping studies. *Journal of the Royal Statistical Society, Series A (Statistics in Society), 162*, 363–382.

BIRKHEAD, G., GALVIN, V., MEEHAN, P., O'CARROLL, P., & MERCY, J. (1993). The emergency department in surveillance of attempted suicide: Findings and methodologic considerations. *Public Health Reports, 108*, 323–331.

BRADSHAW, D., GROENEWALD, P., BOURNE, D., MAHOMED, H., NOJILANA, B., DANIELS, J., ET AL. (2006). Making COD statistics useful for public health at local level in the city of Cape Town. *Bulletin of the World Health Organization, 84*, 211–217.

BREIDING, M., & WIERSEMA, B. (2006). Variability of undetermined manner of death classification in the US. *Injury Prevention, 12*, 49–54.

BRENT, D., JOSHUA, A., PERPER, M., & ALLMAN, C. (1987). Alcohol, firearms and suicide among youth: Temporal trends in Allegheny County, Pennsylvania, 1960–83. *Journal of the American Medical Association, 257*, 3369–3372.

BROOKE, E. (1974). Suicide and attempted suicide. (*Public Health Paper*), Geneva: World Health Organization, pp. 5–127.

BRUGHA, T., & WALSH, D. (1978). Suicide past and present: The temporal constancy of underreporting. *British Journal of Psychiatry, 132*, 177–179.

CANTOR, C., MCTAGGART, P., & DE LEO, D. (2001). Misclassification of suicide: The contribution of opiates. *Psychopathology, 34*, 140–146.

CARR, J., HOGE, C., GARDNER, J., & POTTER, R. (2004). Suicide surveillance in the U.S. military: Reporting and classification biases in rate calculations. *Suicide and Life-Threatening Behavior, 34*, 233–241.

CENTERS FOR DISEASE CONTROL AND PREVENTION. (1997). Recommended framework for presenting injury mortality data. *Morbidity and Mortality Weekly Review, 46(RR14)*, 1–30.

CLARKE-FINNEGAN, M., & FAHY, T. (1983). Suicide rates in Ireland. *Psychological Medicine, 13*, 385–391.

COOPER, P., & MILROY, C. (1995). The coroner's system and under-reporting of suicide. *Medicine, Science and the Law, 35*, 319–326.

CORCORAN, P., ARENSMAN, E., & O'MAHONY, D. (2006). Suicide and other external-cause mortality statistics in Ireland: A comparison of registration and occurrence data. *Crisis, 27*, 130–134.

CUI, J., YIP, P., & CHAU, P. (2004). Estimation of reporting delay and suicide incidence in Hong Kong. *Statistics in Medicine, 23*, 467–476.

DE LEO, D. (2007). *Suicide in the world: An overview.* Killarney, Ireland: International Association of Suicide Prevention.

DONALDSON, A., LARSEN, G., FULLERTON-GLEASON, L., & OLSON, L. (2006). Classifying undetermined poisoning deaths. *Injury Prevention, 12*, 338–343.

DOUGLAS, J. (1967). *The social meaning of suicide.* Princeton: Princeton University Press.

DURKHEIM, E. (1951). *Suicide.* New York: Free Press. (Original work published 1897)

EATON, K., MESSER, S., GARVEY WILSON, A., & HOGE, C. (2006). Strengthening the validity of population-based suicide rate comparisons: An illustration using U.S. military and civilian data. *Suicide and Life-Threatening Behavior, 36*, 182–191.

FARBEROW, N., MACKINNON, D., & NELSON, F. (1977). Suicide: Who's counting? *Public Health Reports, 92*, 223–232.

FARBEROW, N., & NEURINGER, C. (1971). The social scientist as coroner's deputy. *Journal of Forensic Science, 16*, 15–39.

FAUVEAU, V. (2006). Assessing probable causes of death without death registration or certificates: A new science? *Bulletin of the World Health Organization, 84*, 246–247.

FINGERHUT, L., ANNEST, J., & MCLOUGHLIN, E. (1997). Recommendations and reports. *Morbidity and Mortality Weekly Review*, Vol 16/ NoRR14, pages 1–31.

FINGERHUT, L., COX, C., & WARNER, M. (1998). International comparative analysis of injury mortality: Findings from the international collaborative effort on injury statistics. *Advance Data, 303*, 1–20.

GEDDES, J. (1999). Suicide and homicide by people with mental illness: We still don't know how to prevent most of these deaths. *British Medical Journal, 318*, 1225–1226.

GIST, R., & WELCH, Q. (1989). Certification change versus actual behavior change in teenage suicide rates, 1955–1979. *Suicide and Life-Threatening Behavior, 19*, 277–288.

GROHOLT, B., & EKEBERG, O. (2003). Suicide in young people under 15 years: Problems of classification. *Nordic Journal of Psychiatry, 57*, 411–417.

GROSS, T., & HUBER-CAROL, C. (1992).

Regression models for truncated data. *Scandinavian Journal of Statistics, 19,* 193–213.

HANZLICK, R., & COMBS, D. (1998). Medical examiner and coroner systems: History and trends. *Journal of the American Medical Association, 279,* 870–874.

HESSO, R. (1987). Scandinavian routines and practices in the registration of suicide. *Acta Psychiatrica Scandinavica, 336*(Suppl.), 17–21.

HILL, K. (2006). Making deaths count. *Bulletin of the World Health Organization, 84,* 162.

HLADY, W., & MIDDAUGH, J. (1988). The underrecording of suicides in state and national records, Alaska 1983–1984. *Suicide and Life-Threatening Behavior, 18,* 237–244.

HORVITZ, D., & THOMPSON, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association, 47,* 663–685.

HOUSE, A. (1997). Homicides and suicides by mentally ill people. *British Medical Journal, 312,* 135–136.

INJURY-ICE. (2002). *ICD framework: External cause of injury mortality matrix.* Retrieved September 2008 from http://www.cdc.gov/nchs/about/otheract/ice/matrix10.htm.

JENNINGS, C., & BARRACLOUGH, B. (1980). Legal and administrative influences on the English suicide rate since 1900. *Psychological Medicine, 10,* 407–418.

JOBES, D., CASEY, J., BERMAN, A., & WRIGHT, D. (1991). Empirical criteria for the determination of suicide manner of death. *Journal of Forensic Science, 36,* 244–256.

JOHANSSON, L., PAVILLON, G., ANDERSON, R., GLENN, D., GRIFFITHS, C., HOYERT, D., ET AL. (2006). Counting the dead and what they died of. *Bulletin of the World Health Organization, 84,* 254.

JOUGLA, E., PEQUIGNOT, F., CHAPPERT, J., ROSSOLLIN, F., LE TOULLEC, A., & PAVILLON, G. (2002). [Quality of suicide mortality data]. *Review Epidemiologic Sante Publique, 50,* 49–62.

KALBFLEISCH, J., & LAWLESS, J. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in Medicine,* 7, 149–160.

KELLEHER, M., CORCORAN, P., KEELEY, H., DENNEHY, J., & O'DONNELL, I. (1996). Improving procedures for recording suicide statistics. *Irish Medical Journal, 89,* 14–15.

KLECK, G. (1988). Miscounting suicides. *Suicide and Life-Threatening Behavior,18,* 219–236.

KOCHANEK, K., & SMITH, B. (2004). *Deaths: Preliminary data for 2002.* Hyattsville, MD: National Center for Health Statistics.

KOEHLER, S., WEISS, H., SHAKIR, A., SHAEFFER, S., LADHAM, S., ROZIN, L., ET AL. (2006). Accurately assessing elderly fall deaths using hospital discharge and vital statistics data. *American Journal of Forensic Medicine and Pathology, 27,* 30–35.

KOLMOS, L., & BACH, E. (1987). Sources of error in registering suicide. *Acta Psychiatrica Scandinivica, 76*(Suppl.), 22–43.

LAWRENCE, B., MILLER, T., & SPICER, R. (2007). *Injury-related deaths in the 1997 multiple-cause-of-death data.* Unpublished manuscript.

LAWRENCE, B., MILLER, T., & WEISS, H. (2002). *Does underlying cause of death count injury deaths accurately?* Paper presented at the American Public Health Association annual meeting, Washington, DC.

LINSLEY, K., SCHAPIRA, K., & KELLY, T. (2001). Open verdict v. suicide: Importance to research. *British Journal of Psychiatry, 178,* 465–468.

LITMAN, R., CURPHEY, T., SHNEIDMAN, E., FARBEROW, N., & TABACHNICK, N. (1963). Investigations of equivocal suicides. *Journal of the American Medical Association, 184,* 102–207.

LOPEZ, A., & MATHERS, C. (2006). Measuring the global burden of disease and epidemiological transitions: 2002–2030. *Annals of Tropical Medicine & Parasitology, 100,* 481–499.

MATHERS, C., MA FAT, D., INOUE, M., RAO, C., & LOPEZ, A. (2006). Capturing health information: A coding perspective. *Bulletin of the World Health Organization, 84,* 254–255.

MATHERS, C., MA FAT, D., RAO, C., & LOPEZ, A. (2005). Counting the dead and what they died from: An assessment of the global status of cause of death data. *Bulletin of the World Health Organization, 83,* 171–177.

MCCARTHY, P., & WALSH, D. (1975). Suicide in Dublin: I. The under-reporting of suicide and the consequences for national statistics. *British Journal of Psychiatry, 126,* 301–308.

MOENS, G. (1985). The reliability of reported suicide mortality statistics: An experience from Belgium. *International Journal of Epidemiology, 14,* 272–275.

MOHLER, B., & EARLS, F. (2001). Trends in adolescent suicide: Misclassification bias. *American Journal of Public Health, 91,* 150–153.

MONK, M. (1987). Epidemiology of Suicide. *Epidemiology Review, 9,* 51–69.

NAUGHTON, M., DOYLE, A., MELINA, P., & BARRY, D. (1996). Suicide in Kerry: A comparison of the official and a clinically-based rate. *Irish Journal of Psychological Medicine, 13,* 147–148.

NCHS. (1999-2005). *CDC WONDER online database, compiled from compressed mortality file 1999-2005 Series 20 No. 2K, 2008.* http: wonder. cdc.gov Washington, DC: Centers for Disease Control and Prevention, National Center for Health Statistics.

NEELMAN, J., & WESSLEY, S. (1997). Changes in classification of suicide in England and Wales: Time trends and associations with coroners' professional backgrounds. *Psychological Medicine, 27*, 467–472.

NELSON, F., FARBEROW, N., & MACKINNON, D. (1978). The certification of suicide in eleven western states: An inquiry into the validity of reported suicide rates. *Suicide and Life-Threatening Behavior, 8*, 75–88.

O'CARROLL, P. (1989). A consideration of the validity and reliability of suicide mortality data. *Suicide and Life-Threatening Behavior, 19*, 1–16.

OHBERG, A., & LONNQVIST, J. (1998). Suicides hidden among undetermined deaths. *Acta Psychiatrica Scandinavica, 98*, 214–218.

OVENSTONE, I. (1973). A psychiatric approach to the diagnosis of suicide and its effect on the Edinburgh statistics. *British Journal of Psychiatry, 123*, 15–21.

PECK, D., & WARNER, K. (1995). Accident or suicide? Single-vehicle car accidents and the intent hypothesis. *Adolescence, 30*, 463–472.

PESCONSOLIDO, B., & MENDELSOHN, R. (1986). Social causation or social construction of suicide? An investigation into the social organization of official rates. *American Sociological Review, 51*, 80–101.

PHILLIPS, D., & RUTH, T. (1993). Adequacy of official suicide statistics for scientific research and public policy. *Suicide and Life-Threatening Behavior, 23*, 307–319.

PHILLIPS, D., & SMITH, D. (1991). *Suicide at symbolic ages: Death on stocktaking occasions.* In A. Leenaars (Ed.), *Lifespan perspectives of suicide* (pp. 81–92). New York: Plenum.

PLATT, S. (1986). *Suicide trends in 24 European countries, 1972–1984.* Geneva: World Health Organization.

PLATT, S., HAWTON, K., KREITMAN, N., FAGG, J., & FOSTER, J. (1988). Recent clinical and epidemiological trends in parasuicide in Edinburgh and Oxford: A tale of two cities. *Psychological Medicine, 18*, 405–418.

PRITCHARD, C., & AMANULLAH, S. (2006). An analysis of suicide and undetermined deaths in 17 predominantly Islamic countries contrasted with the UK. *Psychological Medicine, 37*, 427–430.

PRITCHARD, C., & HANSEN, L. (2005). Comparison of suicide in people aged 65–74 and 75+ in England and Wales and the major Western countries 1979–1999. *International Journal of Geriatric Psychiatry, 20*, 17–25.

ROCKETT, I., SAMORA, J., & COBEN, J. (2006). The black-white suicide paradox: Possible effects of misclassification. *Social Science and Medicine, 63*, 2165–2175.

ROCKETT, I., & SMITH, G. (1995). Suicide misclassification in an international context. In *Proceedings of the International Collaborative Effort on Injury Statistics.* Hyattsville, MD: National Center for Health Statistics.

ROSENBERG, H., & KOCHANEK, K. (1994). The death certificate as a source of injury data. *Proceedings of the International Collaborative Effort on Injury Statistics.* Hyattsville, MD: National Center for Health Statistics.

ROSENBERG, M., DAVIDSON, L., SMITH, J., BERMAN, A., BUZBEE, H., GANTNER, G., ET AL. (1988). Operational criteria for the determination of suicide. *Journal of Forensic Science, 33*, 1445–1456.

ROSS, O., & KREITMAN, N. (1975). A further investigation of differences in the suicide rates of England and Wales and of Scotland. *British Journal of Psychiatry, 127*, 575–582.

ROYAL COLLEGE OF PSYCHIATRISTS. (1996). *Report of the confidential inquiry into homicides and suicides by mentally ill people.* London: Author.

SAINSBURY, P., & JENKINS, S. (1982). The accuracy of officially reported suicide statistics for purposes of epidemiologic research. *Journal of Epidemiology and Community Health, 36*, 43–48.

SHIBUYA, K., SCHEELE, S., & BOERMA, T. (2005). Health statistics: Time to get serious. *Bulletin of the World Health Organization, 83*, 722.

SOLEMAN, N., CHANDRIMOHAN, D., & SHIBUYA, K. (2006). Verbal autopsy: Current practices and challenges. *Bulletin of the World Health Organization, 84*, 239–244.

SPEECHLEY, M., & STAVRAKY, K. (1991). The adequacy of suicide statistics for use in epidemiology and public health. *Canadian Journal of Public Health, 82*, 38–42.

STANISTREET, D., TAYLOR, S., JEFFREY, V., & GABBAY, M. (2001). Accident or suicide? Predictors of coroners' decisions in suicide and accident verdicts. *Medicine, Science and the Law, 41*, 111–115.

STENGEL, E., & FARBEROW, N. (1967). *Certification of suicide around the world.* Paper presented at Fourth International Conference for Suicide Prevention, Los Angeles, California.

UPDATE REFERENCE COMMITTEE. (2000). *Official updates to* ICD-10. Rio de Janero, Brazil: World Health Organization.

WALSH, B., WALSH, D., & WHELAN, B. (1975). Suicide in Dublin: II. The influence of some social and medical factors on coroners' verdicts. *British Journal of Psychiatry, 126*, 309–312.

WARSHAUER, M., & MONK, M. (1978). Problems in suicide statistics for whites and blacks. *American Journal of Public Health, 68*, 383–388.

WEISMAN, A., & WORDEN, J. (1972). Risk-rescue rating in suicide assessment. *Archives of General Psychiatry, 26*, 553–560.

WHITT, H. (2006a). Where did the bodies go? The social construction of suicide data, New York City, 1976–1992. *Sociological Inquiry, 76*, 166–187.

WHITT, H. (2006b). Errata. *Sociological Inquiry, 76*, 429.

WHO. (1989). *International form of medical cause of death*. Retrieved August 2007 from http://www.searo.who.int/LinkFiles/Pregnancy_Childbirth_n.pdf.

WHO. (1992). *International statistical classification of diseases and related health problems* (10th rev.). Geneva, Switzerland: Author.

WHO. (2004). *International statistical classification of diseases and related health problems* (10th Rev., 2nd Ed., Vol. 2). Geneva, Switzerland: Author.

ZILBORG, G. (1935). Suicide among civilized and primitive races. *American Journal of Psychiatry, 92*, 1–34.

Manuscript Received: October 13, 2008
Revision Accepted: October 15, 2009