RERF TR 15-91

Technical Report Series

Combining Diagnostic Categories to Improve Agreement between Death Certificate and Autopsy Classifications of Cause of Death for Atomic Bomb Survivors, 1950–87

Randolph L. Carter, Ph.D.; Elaine Ron, Ph.D.; Kiyohiko Mabuchi, M.D.

放影研 RERF
Radiation Effects Research Foundation
A Cooperative Japan–United States Research Organization

RERF Technical Report Series

Technical reports are the basic medium for reporting of original research carried out at the Radiation Effects Research Foundation. Reports in this series receive both internal and external peer review and may serve, in part or in toto, as the basis for publication in the open scientific literature. Although they may be quoted and cited, these reports are considered to be internal publications of the Foundation. Copies are available upon request from Publication and Documentation Center, RERF, 5-2 Hijiyama Park, Minami-ku, Hiroshima, 732 Japan.

The Radiation Effects Research Foundation (formerly ABCC) was established in April 1975 as a private nonprofit Japanese foundation, supported equally by the Government of Japan through the Ministry of Health and Welfare, and the Government of the United States through the National Academy of Sciences under contract with the Department of Energy.

業績報告書シリーズ

原爆被爆者における死亡診断書および剖検死因分類の一致率向上のための診断分類の結合，1950−87年§

Combining Diagnostic Categories to Improve Agreement between Death Certificate and Autopsy Classifications of Cause of Death for Atomic Bomb Survivors, 1950−87

Randolph L. Carter[a]　Elaine Ron[b,c]　馬淵清彦[b]

要　約

死亡診断書による診断と剖検診断とは，ほとんどの特定死因についてだけでなく，それらの死因を主要疾患群にまとめた場合でさえも，さほど一致しないことがいくつかの研究で認められている。広島・長崎の原爆被爆者の寿命調査集団で，1987年9月以前に剖検を受けた5,130人のデータを改めて検討した結果も，同様に思わしくなかった。観察例数が10例以上であった疾患についてみると，確認率は13％から97％，検出率は6％から90％の範囲であった。確認率および検出率がともに70％以上であったのは，調査した60種類の疾患中わずか6疾患であり，国際疾病分類（ICD）の大分類別にまとめた16の疾患分類中わずか1分類群（新生物）だけであった。このような低い一致率は，死亡診断書の診断に基づく調査結果の解釈にあたって十分な注意が必要であることを示唆するものである。診断をどのような分類群にまとめればまずまずの精度が得られるかを決定するために，対象者5,130人のデータに段階的集積法を適用した。その結果，分類方法として，乳癌；女性のその他の癌；消化器系癌；喉頭癌；白血病；鼻と耳と副鼻腔の癌；舌癌；外因死；血管疾患；およびその他のすべての死因の10種類の分類を得た。これらの疾患分類群の確認率および検出率は少なくとも66％であった。この分類群の幅は，特に非腫瘍性の疾患については広範囲にわたっているが，これ以上細分類すると精度が下がるものもでてくる。このようにして得た分類方法を用いたときの死亡診断書と剖検診断との一致率が全体として72％であったのに対し，主要疾患分類に基づく別の厳密な分類方法による一致率は53％であった。血管疾患

§本報告にはこの要約以外に訳文はない。承認1991年10月29日。印刷1993年5月。

[a]放影研統計部　現在，フロリダ州ゲインズビル フロリダ大学生物統計部門統計部　[b]放影研疫学部
[c]現在，メリーランド州ベテスダ米国国立癌研究所放射線疫学部

を他のすべての非腫瘍性疾患にまとめた同様の分類方法では，87％の一致率が得られた。これ以上に大きくまとめても一致率の向上はほとんど見られなかった。上記の10種類の診断分類群の中には，各種の共変量によって精度が変化するものがあった。例えば，これらの分類群のほとんどは，死亡年齢の増加とともに精度の低下を示した。したがって，母集団全体に比べて精度が良かったり，悪かったりするような副次集団が存在する。以上の結果は，今後の死因別死亡調査においてどの疾患もしくは集団を調査対象とすべきかを必ずしも示すものではないが，調査の企画，データの分析および結果の解釈に有用な情報を研究者に提供するものである。

Technical Report Series

Combining Diagnostic Categories to Improve Agreement between Death Certificate and Autopsy Classifications of Cause of Death for Atomic Bomb Survivors, 1950–87[§]

Randolph L. Carter, Ph.D.[a]; Elaine Ron, Ph.D.[b,c]; Kiyohiko Mabuchi, M.D.[b]

Summary

Several investigators have observed less-than-desirable agreement between death certificate diagnoses and autopsy diagnoses for most specific causes of death, and even for some causes grouped by major disease category. Our results from data on 5130 autopsies of members of the Life Span Study cohort of atomic bomb survivors in Hiroshima and Nagasaki conducted prior to September 1987 were equally discouraging. Among diseases with more than 10 cases observed, confirmation rates ranged from 13% to 97% and detection rates from 6% to 90%. Both rates were greater than 70% for only 6 of 60 disease categories studied and for only 1 of 16 categories defined by major International Classification of Disease categories (neoplasms). This deficiency suggests cautious interpretation of results from studies based on death certificate diagnoses. To determine whether any groupings of diagnoses might meet acceptable accuracy requirements, we applied a hierarchical clustering method to data from these 5130 cohort members. The resulting classification system had 10 categories: breast cancer; other female cancers; cancers of the digestive organs; cancer of the larynx; leukemia; nasal, ear, or sinus cancer; tongue cancer; external causes; vascular disease; and all other causes. Confirmation and detection rates for each of these categories were at least 66%. Although the categories are broad, particularly for nonneoplastic diseases, further divisions led to unacceptable

[§]*The complete text of this report will not be available in Japanese. Approved 29 October 1991; printed May 1993.*

[a]*Department of Statistics, RERF, and presently at the Department of Statistics, Division of Biostatistics, University of Florida, Gainesville, Florida;* [b]*Department of Epidemiology, RERF, and* [c]*presently at the Radiation Epidemiology Branch, National Cancer Institute, Bethesda, Maryland.*

accuracy rates for some of the resulting diagnostic groups. Using the derived classification system, there was 72% agreement overall between death certificate and autopsy diagnoses compared to 53% agreement for a second system obtained by grouping strictly by major disease category. Eighty-seven percent agreement was observed for a similar classification system with vascular disease grouped with all other nonneoplastic diseases. Further agglomeration achieved very little additional improvement. Accuracy rates for some of the categories of the 10-category diagnostic system defined above varied with various covariates. For example, accuracy decreased with increasing age at death for most of these categories. Thus, subpopulations exist for which accuracy rates can be expected to be either better or worse than for the whole population. Although these results do not necessarily dictate which diseases and/or populations should be studied in future cause-specific mortality investigations, they do provide investigators with useful information pertinent to the planning of their study, analysis of the data, and interpretation of the results.

Introduction

Previous reports on agreement between death certificate (DC) and autopsy diagnoses have not been encouraging. Yamamoto et al[1] observed confirmation and detection rates both over 70% for only 4 of 46 cause-of-death categories studied in the Atomic Bomb Casualty Commission (ABCC)/Radiation Effects Research Foundation (RERF) Life Span Study (LSS) cohort of atomic bomb survivors in Hiroshima and Nagasaki, Japan. A number of other authors also have noted inaccuracies in DC diagnoses.[2–4]

It is well understood that improved agreement can be obtained by combining specific causes of death into coarser diagnostic groups. Agreement rates that have been reported for broadly defined disease types, however, are still generally not satisfactory. In a study of autopsied deaths among LSS cohort members, Jablon et al[5] reported overall detection rates of over 70% for only one (digestive system) of five groups of specific causes of death due to cancer (digestive, respiratory, genitourinary, female genital, and hematologic excluding leukemia). The prospects of grouping to attain satisfactory accuracy for noncancer causes appear even bleaker in light of relatively poor accuracy rates for noncancers.[1,4] Nevertheless, the possibility of defining mutually exclusive and exhaustive classifications for which there is satisfactory agreement should not be ruled out until rigorous attempts to do so have failed.

We utilized statistical clustering methods to derive diagnostic categories with acceptable agreement between DC and autopsy diagnoses.

Methods

The LSS cohort of atomic bomb survivors was identified from census records in 1950 (see Beebe and Usagawa[6] for details). In 1961, a comprehensive autopsy procurement program focusing on this cohort was initiated. Prior to this program, autopsies were performed selectively with a bias toward highly exposed individuals and those who were thought to have died from cancer.[1] Overall, autopsies were performed on 5130 cases (11.1%) of the 46,331 deaths among the LSS cohort prior to September 1987, of which 652 (12.7%) occurred before 1961.

Table 1. International Classification of Diseases (ICD) cause-of-death codes used for agreement analysis

Classification	RERF codes	ICD codes		
		7th	8th	9th
I. Infectious and parasitic diseases		001–138	000–136	001–139, 511.9
Tuberculosis	01	001–019	010–019	010–018, 137, 511.9
Other	02	020–138	000–009, 020–136	001–009, 020–136, 138, 139
II. Neoplasms		140–204+	140–209+	140–208+
A. Lip, oral cavity, and pharynx		140–148	140–149	140–149
Lip	03	140	140	140
Tongue	04	141	141	141
Salivary glands	05	142	142	142
Other	06	143–148, 210	143–149, 210	143–149, 210
B. Digestive organs		150–159	150–159	150–159
Esophagus	07	150	150	150
Stomach	08	151	151	151
Colon	09	153	153	153
Rectum	10	154	154	154
Liver	11	155, 155.0, 155.8, 156	155.0, 197.8	155.0, 155.2
Gallbladder, biliary	12	155.1	155.1, 156	155.1, 156
Pancreas	13	157	157	157
Other	14	152, 158, 159, 211	152, 158, 159, 211	152, 158, 159, 211
C. Respiratory		160–165	160–163	160–165
Nasal, ear, sinuses	15	160	160	160
Larynx	16	161	161	161
Trachea, bronchus, lung	17	162, 163	162	162
Other	18	164, 165, 212	163, 212	163–165, 212
D. Bone	19	196, 225	170, 213	170, 213
E. Soft tissues	20	197	171, 192.4, 192.5	171
F. Skin		190, 191	172, 173	172, 173
Melanoma	21	190	172	172

Continued

Table 1. Continued

Classification	RERF codes	ICD codes		
		7th	8th	9th
Other skin	22	191, 220	173.0–173.4, 173.6–173.9, 216	173, 216
G. Breast	23	170, 213	174, 217	174, 175, 217
H. Female genital		171–176	179–184	179–184
Cervix uteri	24	171	180	180
Corpus uteri	25	172	182.0	182.0
Uterus, NOS	26	174, 215	182.9, 219	179, 219
Ovary	27	175.0, 175, 216	183.0, 220	183.0, 220
Other female genital	28	173, 175.1, 175.8, 175.9, 176, 214, 217	181, 183.1, 183.9, 184, 218, 221	181, 183.2–183.9, 184, 218, 221
I. Male genital		177–179	173.5, 185–187	185–187
Prostate	29	177	185	185
Testis	30	178	186	186
Other	31	179, 218	173.5, 187, 222	187, 222
J. Urinary		180, 181.0, 181.7, 181.8	188, 189.0–189.2, 189.9	188, 189.0–189.4, 189.8, 189.9
Urinary bladder	32	181.0, 181	188	188
Kidney	33	180	189.0	189.0
Other	34	181.7, 181.8, 219, 232–236	189.1, 189.2, 189.9, 223, 233–237	189.1–189.4, 189.8, 189.9, 223, 233, 236
K. Eye	35	192	190, 224	190, 224
L. CNS	36	193, 223	191, 192.0–192.3, 192.9, 225	191, 192, 225
M. Thyroid	37	194, 251	193, 241	193, 226
N. Other endocrine	38	195, 224	194, 226	194, 227
O. Hematopoietic		200–204	200–207	200–208
Lymphoma	39	200–202	200–202	200–202
Multiple myeloma	40	203	203	203
Leukemia	41	204	204–207	204–208

Table 1. Continued

Classification	RERF codes	ICD codes 7th	8th	9th
P. Other, ill-defined	42	198, 199, 221, 222, 226–231, 237–239	195–196, 197.0–197.7, 197.9, 198, 199, 208, 209, 214, 215, 227, 228, 230–232, 238, 239	195, 196, 197.0–197.8, 198, 199, 214, 215, 228, 229–232, 234, 235, 237–239
III. Endocrine, nutritional, etc.		240–289	240–279	240–279
Diabetes	43	260	250	250
Other	44	240–259, 261–289	240–249, 251–279	240–249, 251–279
IV. Blood	45	290–299	280–289	280–289
V. Mental	46	300–326	290–315	290–319
VI. Nervous system	47	340–398	320–389	320–389
VII. Cerebral vascular disease	48	330–334	430–438	430–438
VIII. All other cardiovascular diseases		400–468	390–429, 440–458	390–429, 440–459
Ischemic heart disease	49	420–422	410–414	410–414
Other	50	400–419, 423–468	390–409, 415–429, 440–458	390–409, 415–429, 440–459
IX. Respiratory system		470–529	460–519	460–510, 511.0, 511.1, 511.8, 512–519
Pneumonia	51	490–493	480–486	480–486
Other	52	470–489, 494–529	460–479, 487–519	460–479, 487–510, 511.0, 511.1, 511.8, 512–519
X. Digestive system		530–587	520–577	520–579
Cirrhosis	53	581	571	571
Other	54	530–580, 582–587	520–570, 572–577	520–570, 572–579
XI. Urinary disease	55	590–609	580–599	580–599

Continued

Table 1. Continued

Classification	RERF codes	ICD codes		
		7th	8th	9th
XII. Genital disease	56	610–637	600–629	600–629
XIII. Ill-defined diseases	57	780–795	780–796	780–799
XIV. All other diseases	58	638–779, 796–799	630–779, 797–799	630–779
XV. External causes excluding suicide	59	800–969, 980–999	800–949, 960–999	800–949, 960–999
XVI. Suicide	60	970–979	950–959	950–959

Note: NOS = not otherwise specified; CNS = central nervous system.

The underlying cause of death from the DC and the principal cause from autopsy were originally recorded in the form of International Classification of Disease (ICD) codes, ICD7,[7] ICD8,[8] or ICD9,[9] depending on the date of death. For our study, these causes were grouped into 60 categories for analysis (Table 1). The primary goals of analysis were to obtain a grouping of categories that optimized agreement between DC and autopsy diagnoses and to compare the resulting diagnostic classification system with that defined by major ICD categories.

The statistics used for the comparison of alternative classification systems were overall percentage agreement, kappa[10] (adjusted for percentage agreement by chance), and confirmation and detection rates, defined by:

$$p = \frac{\text{no. of DC and autopsy agreements}}{\text{no. of autopsies performed}} \times 100\% \ ,$$

$$\kappa = \frac{p - p_c}{100 - p_c} \ ,$$

$$\text{confirmation rate} = \frac{\text{no. of DC diagnoses confirmed by autopsy}}{\text{no. of DC diagnoses among autopsied cases}} \ ,$$

and

$$\text{detection rate} = \frac{\text{no. of autopsy diagnoses detected by DC}}{\text{no. of autopsy diagnoses}} \ ,$$

respectively, where p_c is the percentage agreement expected by chance. For example, suppose we wish to evaluate the two-category diagnostic system: can-

cer, noncancer. Then the calculation of these statistics may be described best in terms of the elements of the following 2 × 2 table:

Death certificate diagnosis	Autopsy diagnosis: Cancer	Noncancer	Row totals
Cancer	A	C	$A + C$
Noncancer	B	D	$B + D$
Column totals	$A + B$	$C + D$	T

where $p = (A + D)/T$, $p_c = [(A + C)(A + B) + (B + D)(C + D)]/T^2$, the confirmation rate for cancer = $100A/(A + C)$, and the detection rate for cancer = $100A/(A + B)$.

Our attempt to define a classification system for which there is optimal agreement between the DC and autopsy classifications involved the following hierarchical partitioning strategy: First, all sex-specific cancers were excluded (RERF codes 23–31 in Table 1); then the set of causes remaining was partitioned into two groups in a way designed to maximize the percentage agreement; the confirmation and detection rates for each group were calculated; each group for which both rates were greater than 70% was partitioned again. This process was repeated subject to the following stopping rules: If neither of the two resulting groups at any step of the process had confirmation *and* detection rates greater than 70%, then they were recombined to form a final group. If only one of the resulting groups failed the 70% criteria, then it was searched for individual diagnoses with confirmation *and* detection rates over 70%, which were then separated out to form single-cause categories, after which the partitioning stopped.

The final categories resulting from our iterative partitioning strategy that did not meet the 70% rule were grouped together with male cancers to form an "all other causes" group. Breast cancer and other female cancers met or nearly met the 70% requirement and were therefore retained as categories of a final diagnostic classification system. The overall percentage agreement and the kappa coefficient were calculated and compared with those of a second classification system with the same number of categories that was based solely on grouping by major disease category.

Multiple logistic regression analyses[11] were performed to test the effects of city (Hiroshima, Nagasaki), sex, place of death (hospital, home, or clinic), Adult Health Study (AHS)* participation (yes, no), age at death (<60, 60–74, and ≥75 years), radiation dose (not in city, 0–9, 10–499, 500–999, and ≥1000 mGy), and period of death (before 1961, 1961–65, 1966–70, 1971–75, and after 1975) on confirmation, and detection rates for several categories of the diagnostic system

*The AHS cohort consists of a subsample of 20,000 LSS members who have been invited to participate in biennial physical examinations conducted by ABCC/RERF since 1958.

defined by the above-mentioned clustering procedure. The CATMOD procedure of the Statistical Analysis System (SAS) package was used for computations.[12]

Ideally, each dichotomization of diagnosis groups in the hierarchical partitioning strategy described above should be performed in a way that produces the largest possible percentage agreement between autopsies and DCs. The following hypothetical example illustrates such a partitioning.

Suppose there are four possible diagnoses, denoted by 1, 2, 3, 4, and that autopsy and DC reports for 80 deaths are summarized as follows:

		Autopsy			
		1	2	3	4
DC	1	9	1	0	11
	2	0	18	2	0
	3	1	5	15	1
	4	8	0	2	7

Now, there are seven ways to partition the set of four diagnoses into two sets, {1, (2, 3, 4)}, {2, (1, 3, 4)}, {3, (1, 2, 4)}, {4, (1, 2, 3)}, {(1, 2), (3, 4)}, {(1, 3), (2, 4)}, and {(1, 4), (2, 3)}. The optimal dichotomization is the one that results in maximum agreement between autopsies and DCs; that is, the one that minimizes the number of misclassifications between resulting groups of diagnoses. In this example it is easy to verify by enumeration that the optimal partitioning is {(1, 4), (2, 3)}. The correspondingly rearranged table is

		Autopsy			
		1	4	2	3
DC	1	9	11	1	0
	4	8	7	0	2
	2	0	0	18	2
	3	1	1	5	1

This table includes 35 agreements in the first block and 40 in the second and is the most nearly block diagonal among the seven tables obtained by arranging rows and columns in correspondence to the seven dichotomous partitionings listed above.

We would like to perform such a partitioning at each stage of the hierarchical strategy described above. Unfortunately, with 60 diagnoses, it would be computationally burdensome to do so. Therefore, we applied methods of factor analysis[13] as an approximating alternative. The remainder of this section is devoted to the details and justification of these methods.

Let k be the number of individual causes in a set of causes to be partitioned, and define the variables

$$y_i = (a_i + d_i)/[\sum (a_i + d_i)^2]^{1/2} , \quad i = 1, 2, \ldots, k,$$

where

$a_i = \begin{cases} 1 \text{ if the } i\text{th diagnosis was given on the autopsy report} \\ 0 \text{ otherwise,} \end{cases}$

$d_i = \begin{cases} 1 \text{ if the } i\text{th diagnosis was reported on the death certificate} \\ 0 \text{ otherwise,} \end{cases}$

and the summation is over all 5130 deaths autopsied at RERF. Note that the subscript for individuals has been suppressed in y_i, a_i, and d_i. Further, let

n_{ij} = the number of deaths classified as type i by DC and type j by autopsy,

$n_{i\cdot}$ = the number of type-i diagnoses by DC,

$n_{\cdot j}$ = the number of type-j diagnoses by autopsy,

$n_{\cdot\cdot}$ = the total number of deaths studied, and

$m_{ij} = n_{ij} + n_{ji}$, $i \neq j$.

Note that m_{ij} is the total number of disagreements between DCs and autopsies that involve both diagnoses i and j.

Let S denote the uncorrected sum of squares and cross-products matrix of the y_i variables, with the $(i, j\text{th})$ element denoted by s_{ij}. We see that

$$\begin{aligned} s_{ij} &= \sum y_i y_j \\ &= m_{ij} / [\Sigma (a_i + d_i)^2 \, \Sigma (a_j + d_j)^2]^{1/2} \quad , \quad i \neq j \\ &= m_{ij} / [n_{\cdot i} + 2n_{ii} + n_{i\cdot})\,(n_{\cdot j} + 2n_{jj} + n_{j\cdot})]^{1/2} \, , \end{aligned}$$

and

$$\begin{aligned} s_{ii} &= \Sigma y_i^2 \\ &= 1 \, . \end{aligned}$$

The best two-group split of k diagnoses is that which minimizes the sum of the values of m_{ij} over i in one group and j in the second. Alternatively, we attempted to minimize the sum of the values of s_{ij} over i in one group and j in the other. This alternative is reasonable because the numerator of s_{ij} is m_{ij} and the sum of the values of m_{ij} is zero if and only if the sum of the values of s_{ij} is zero. The alternative was chosen for computational convenience, as it required only the use of standard programs for factor analysis (e.g., SAS PROC FACTOR[12]).

Stated differently, our alternative objective was to find a reordering of rows and corresponding columns of $\boldsymbol{S}$ that would produce a new matrix, which was as nearly block diagonal as possible. To this end, note that

$$\begin{aligned} \boldsymbol{S} &= \beta_1' \boldsymbol{X'X} \beta_1 + \boldsymbol{E'E} \\ &= (\Lambda_1^{1/2} \beta_1)'(\Lambda_1^{1/2} \beta_1) + (\Lambda_2^{1/2} \beta_2)'(\Lambda_2^{1/2} \beta_2) \\ &\overset{(def)}{=} \boldsymbol{L'L} + \boldsymbol{E'E} \quad , \end{aligned} \tag{1}$$

where β_1' is the $k \times 2$ matrix formed by the first two eigenvectors of **S**; **X** is the $n.. \times 2$ matrix of scores on the first two principal components of y_i; $i = 1, 2, \ldots, k$; β_2' is the $k \times (k-2)$ matrix formed by the last $k-2$ eigenvectors of S; $\Lambda_1^{1/2}$ is the diagonal matrix formed from the square roots of the two largest eigenvalues of S; and $\Lambda_2^{1/2}$ is the diagonal matrix of the remaining $k-2$ eigenvalues of S. Thus, to the extent that the elements of **E** are small and, hence $E'E$ is approximately **0**, our objective can be achieved by reordering the rows and corresponding columns of $L'L$ to approximate a block-diagonal matrix.

Note that $L'O'OL = L'L$ for any 2×2 matrix O satisfying $O'O = I_2$ (i.e., for any 2×2 orthonormal matrix O). Thus, our problem of reordering the rows and columns of $L'L$ can be solved simply by choosing O so that $L'O'$ approximates the matrix $(u, 1-u)$ as closely as possible, where u is a $k \times 1$ vector of zeros and ones. Readers familiar with factor analysis will recognize $(u, 1-u)$ as representing "simple structure" and the problem of choosing O as the factor rotation problem (see ref. 13, pp 118–36). The most commonly applied solution, and the one used in this report, is the varimax rotation (see ref. 13, pp 129–34). Varimax rotation was accomplished by choosing O to maximize the pooled variance of squared elements of $L'O'$ within columns. Each row of the resulting $L'O'$contained one element that was large relative to the other. When the first element was larger than the second, the diagnostic category was put in one group, otherwise it was put in the second. Thus, the desired reordering of $L'L$ was achieved by ordering the k y_i variables with respect to corresponding entries in the first column of $L'O'$. Subsequently, we will denote the resulting reordered $L'L$ by $\hat{S}^*$.

The extent to which the corresponding reordering of S, say S^*, differs from a block-diagonal matrix depends on the frequency of misdiagnoses by DC between the two newly defined categories. If there are no cross-diagnoses between these categories, then one element in each row of $L'O'$ would be zero and both S^* and $\hat{S}^*$ would be block diagonal. Hence, the optimal solution is obtained in such cases. If there are cross-diagnoses, then the extent to which S^* differs from $\hat{S}^*$ also contributes to variation of S^* from block diagonality. From the theory of principal components, however, we know that both the trace and determinant of E'E in Equation (1) are smaller than for any other X and $\beta 1$ (see ref. 12, p 622). Thus, in this sense, $\hat{S}^*$ is the "best" estimator of S^* among estimators of rank 2.

Results

The frequency of occurrence of DC and autopsy diagnoses and confirmation and detection rates for each of the 60 specific causes of death are presented in Table 2. Detection rates ranged from 0% to 90% and were above 70% for only 7 causes. Confirmation rates ranged from 0% to 100%, and only 15 were above 70%. Confirmation rates and detection rates were greater than 70% for only 6 causes of death.

Results of our hierarchical clustering of causes of death in Table 1 are summarized in the Figure. Confirmation and detection rates are presented there for each diagnostic category at each stage of the hierarchical procedure.

Table 2. Frequency, detection rate, and confirmation rate for each diagnosis among deaths in the Life Span Study cohort that were autopsied at RERF

Diagnosis	Autopsy (*n*)	Death cert. (*n*)	Agreed (*n*)	Detect. rate (%)	Confirm. rate (%)
I. Infectious and parasitic diseases					
Tuberculosis	320	264	171	53	65
Other	58	55	12	21	22
II. Neoplasms					
A. Lip, oral cavity, and pharynx					
Lip	0	0	0	—	—
Tongue	8	7	6	75	86
Salivary glands	1	0	0	0	—
Other	4	6	2	50	33
B. Digestive organs					
Esophagus	50	42	29	58	69
Stomach	503	431	355	71	82
Colon	52	36	23	44	64
Rectum	42	44	29	69	66
Liver	60	96	33	55	34
Gallbladder, biliary	83	18	10	12	56
Pancreas	56	34	19	34	56
Other	8	24	1	13	4
C. Respiratory					
Nasal, ear, sinuses	11	10	8	80	73
Larynx	13	13	11	85	85
Trachea, bronchus, lung	202	129	103	51	80
Other	2	1	0	0	0
D. Bone	2	9	1	50	11
E. Soft tissues	4	0	0	0	—
F. Skin					
Melanoma	1	0	0	0	—
Other skin	6	6	2	33	33
G. Breast	37	29	28	76	97
H. Female genital					
Cervix uteri	61	11	8	13	73
Corpus uteri	2	0	0	0	—
Uterus, NOS	16	62	8	50	13
Ovary	31	15	7	23	47
Other female genital	3	6	1	33	17
I. Male genital					
Prostate	23	7	3	13	43
Testis	0	0	0	—	—

Continued

Table 2. Continued

Diagnosis	Autopsy (*n*)	Death cert. (*n*)	Agreed (*n*)	Detect. rate (%)	Confirm. rate (%)
Other	2	1	1	50	100
J. Urinary					
Urinary bladder	31	20	15	48	75
Kidney	16	5	2	13	40
Other	3	13	1	33	8
K. Eye	0	0	0	—	—
L. CNS	18	1	1	6	100
M. Thyroid	19	8	7	37	88
N. Other endocrine	10	1	1	10	100
O. Hematopoietic					
Lymphoma	41	21	17	42	81
Multiple myeloma	8	7	5	63	71
Leukemia	49	55	44	90	80
P. Other, ill-defined	11	72	0	0	0
III. Endocrine, nutritional, etc.					
Diabetes	31	61	19	61	31
Other	29	58	6	21	10
IV. Blood	16	32	7	44	22
V. Mental	24	46	6	25	13
VI. Nervous system	62	55	21	34	38
VII. Cerebral vascular disease	688	1158	461	67	40
VIII. All other cardiovascular diseases				34	
Ischemic heart disease	244	375	83	34	22
Other	992	430	176	18	41
IX. Respiratory system					
Pneumonia	216	209	39	18	19
Other	192	126	29	15	23
X. Digestive system					
Cirrhosis	134	131	63	47	48
Other	224	260	92	41	35
XI. Urinary disease	105	84	20	19	24
XII. Genital disease	22	10	4	18	40
XIII. Ill-defined	46	327	9	20	3
XIV. All other diseases	83	55	19	23	35
XV. External causes excluding suicide	148	114	76	51	67
XVI. Suicide	7	40	5	71	13

Note: NOS = not otherwise specified; CNS = central nervous system.

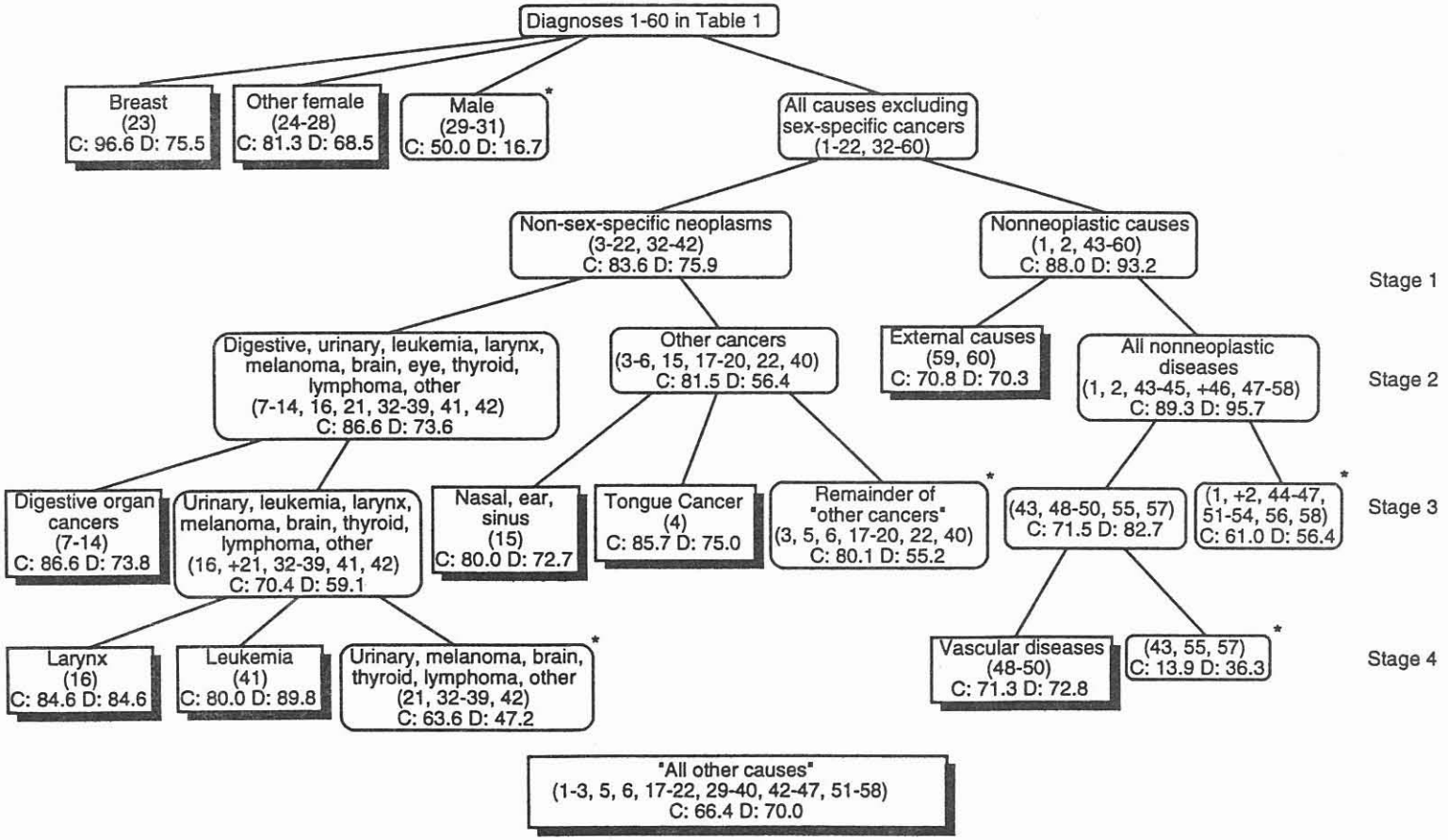

Figure. The shadowed boxes form categories in the diagnostic system from our hierarchical clustering strategy. The diagnosis groups marked with asterisks were eventually combined in the "all other causes" box. Codes preceded by a "+" in Stages 2 through 4 belonged in the "other factor" group but were switched to improve the interpretability of the groups. All switches made at Stage 1 are described in the text and are not marked here.

The initial partitioning split the 49 specific causes of death (excluding sex-specific cancers and lip and eye cancers, which were not diagnosed among the 5130 deaths studied) into two groups, one of which was dominated by neoplastic and the other by nonneoplastic diseases. Three neoplastic causes were included in the nonneoplastic group: endocrine tumors excluding thyroid (code 38), multiple myeloma (code 40), and salivary gland cancer (code 5). All were weakly associated with the nonneoplastic group and therefore were combined with other neoplasms to improve interpretability. Three nonneoplastic causes were included initially in the neoplasm group: other gastrointestinal (code 54), cirrhosis (code 53), and blood disease (code 45). Their associations with the nonneoplastic group were nearly as strong, however, and they were changed to the nonneoplastic group to improve interpretability. The first partition, therefore, resulted in two diagnostic groups: neoplasms and nonneoplasms (Stage 1 in the Figure).

The methods described in the preceding section resulted in a classification system with 10 categories of specific causes shown in the shadowed boxes of the Figure. This system includes these categories: female breast cancer; other female cancer; cancer of the digestive organs; cancer of the larynx; leukemia; nasal, ear, or sinus cancer; tongue cancer; external causes; vascular diseases; and all other causes of death. Modifications made for the sake of interpretation in Stages 2–4 of the hierarchical clustering procedure are marked by a "+" in the Figure.

Given these 10 diagnostic categories, the overall percentage agreement between DCs and autopsies was 72%, kappa = 0.59. An approximate 95% confidence interval for kappa is 0.56, 0.61. In contrast, the percentage agreement and kappa were 53% and 0.44 (approximate 95% CI = 0.43, 0.45) when we used a 10-category system based on the more conventional classification system defined by the 16 major categories in Table 1, with blood, mental, nervous system, genital, ill-defined, and all other diseases combined and suicide combined with other external causes.

In addition, we considered several broader classification systems that were motivated by the results at different stages of our hierarchical clustering method (see the Figure). Pertinent statistics are presented in Table 3. Overall agreement was very high (87%–88%) when nonneoplastic diseases were grouped together and remained moderately high (72%) when vascular diseases were separated from other nonneoplasms. Any finer classification, however, resulted in unacceptably low accuracy rates.

Results from the logistic regression analyses to assess the effects of various covariates (city, sex, place of death, AHS participation, age at death, radiation dose, and period of death) on confirmation and detection rates for the 10-category classification system defined in the Figure are presented in Table 4. Because of the small sample sizes in the breast, leukemia, tongue, larynx, and nasal, ear, or sinus cancer categories, these were grouped into a "small groups combined" category for this analysis.

The age-at-death effect, when significant, resulted from an observed decrease in accuracy with increased age. Both confirmation and detection of all other causes were worse for females than males. Radiation dose affected only detection rates for vascular diseases and external causes. Identification of vascular diseases as the primary cause of death improved with increasing dose, whereas that of external causes was worse in the highest-dose group than in the lower-dose

Table 3. Alternative classifications derived from the Figure

Motivating stage(s) in the Figure	Categories	Min. confirm. rate[a] (%)	Min. detect. rate[a] (%)	Overall % agree-ment	κ
3 and 4	**Neoplasms:** breast; other female[b]; digestive organs; larynx; leukemia; nasal cavity/ear/sinus; tongue **Nonneoplasms:** external causes; vascular **Mixed:** all other causes of death[c]	66.4	68.5	72	0.59
3 and 4 for neoplasms, 2 for nonneoplasms	**Neoplasms:** breast; other female; digestive organ; larynx; leukemia; tongue; nasal cavity/ear/sinus; other[b] **Nonneoplasms:** external,[c] all nonneoplastic diseases	70.8	53.6	87	0.72
1	Breast; other female; male[b,c]; other neoplasms; all nonneoplastic causes	50.0	16.0	87	0.71
1 with sex-specific cancers grouped with other neoplasms	**Neoplasms, nonneoplasms**	86.6	77.4	88	0.72

[a]Confirmation and detection rates for each classification system were at least 68% for all but the minimum category (when indicated) in each diagnostic classification system considered.

[b]Denotes the category with the minimum detection rate for each classification system.

[c]Denotes the category with the minimum confirmation rate.

groups. The confirmation rate for vascular disease was better for nonhospital than hospital deaths. Conversely, detection and confirmation rates for all other causes were better for hospital than nonhospital deaths. Detection rates for digestive cancers improved dramatically during 1961–65, after initiation of the comprehensive autopsy procurement program, and declined steadily in subsequent periods. Detection of vascular diseases improved with time, whereas that of all other causes worsened with time until the last period, when it improved slightly compared with the previous period. Confirmation of cancers in the small-groups-combined category improved in the last period. Period effects on accuracy rates for external causes seemed to be due primarily to high values during 1961–65. City and AHS participation had no significant effect on accuracy.

Discussion

Most cause-specific mortality studies face the problem of potential inaccuracies in DC diagnoses. The results of Yamamoto et al[1] suggested that this problem is substantial and cannot be ignored for many specific causes of death. The results of Jablon et al[5] and Ron et al[14] show that this problem is not solved by grouping of major ICD disease categories. Sposto et al[15] proposed a statistical solution to this problem. Their methods, however, require accurate estimates of misclassification rates, specialized statistical software, and substantial statistical and computing expertise. Furthermore, their methods presumably are most effective when DC diagnoses are reasonably accurate for the diagnostic categories studied. Thus, the question arises whether broader classification groups can be defined to minimize the problem of misclassified primary causes of death on the DC. An answer to this question is also of interest in that it necessarily identifies causes of death that are inadequately distinguished by DC diagnoses and thereby points

Table 4. *p* values for the effects of city, sex, and place of death, Adult Health Study (AHS) participation, age at death, and radiation dose on detection and confirmation rates for cause-of-death categories[a,b]

Category	City	Sex	Place of death	AHS	Age at death	Radiation dose	Period
Other female cancer							
Detection	0.49	NA	0.72	0.75	0.13	0.89	0.48
Confirmation	0.10	NA	0.73	0.33	0.67	0.50	0.62
Digestive organ cancer							
Detection	0.69	0.06	0.09	0.08	0.00	0.14	0.02
Confirmation	0.49	0.054	0.90	0.88	0.02	0.65	0.35
External causes							
Detection	0.99	0.68	0.56	0.38	0.001	0.02	0.04
Confirmation	0.84	0.41	0.73	0.49	0.052	0.64	0.001
Vascular disease							
Detection	0.47	0.77	0.46	0.87	0.00	0.03	0.02
Confirmation	0.77	0.33	0.01	0.16	0.13	0.33	0.52
Small groups combined[c]							
Detection	0.69	0.51	0.11	0.66	0.59	0.25	0.89
Confirmation	0.40	0.34	0.39	0.87	0.22	0.23	0.03
All other causes							
Detection	0.92	0.049	0.00	0.36	0.005	0.72	0.002
Confirmation	0.28	0.003	0.00	0.42	0.00	0.36	0.11

[a] Specific causes included in each category are indicated in the boxes of the Figure.

[b] The direction of each effect with $p < .05$ is described in the last paragraph of the Results.

[c] The small-groups-combined group was formed by combining breast, leukemia, tongue, larynx, and nasal cavity/ear/sinus cancer categories of the final classification system defined in the Figure.

to areas in which improvements in diagnoses are needed and special precautions are advisable when analyzing and interpreting DC data.

In the current study, we utilized a hierarchical partitioning clustering method in an attempt to define categories with optimal agreement between DC and autopsy diagnoses. We derived a diagnostic classification system that would be an improvement over that obtained by grouping causes of death by major disease category (agreement = 72%). Nevertheless, the kappa statistic for this improved grouping (0.59) represented only moderate agreement between DC and autopsy classifications in these groups. Thus, alternative classification systems were considered (see Table 3).

Note that the largest drop in percentage agreement and kappa occurred when the nonneoplastic diseases group was partitioned at Stage 3. The percentage agreement and kappa obtained for the categorization that does not partition this group represent good agreement (see row 2 of Table 3). Very little was gained by further combinations of diagnostic categories. Thus, we recommend consideration of either classification system in the first or second row of Table 3.

The primary substantive difference between these two systems is that the first separates vascular diseases from other nonneoplastic causes of death. Thus, the choice between them depends on a user's criterion for the adequacy of kappa and his or her interest in studying vascular diseases. In the terminology of Landis and Koch,[10] 0.59 represents "moderate" and 0.72 "substantial" agreement. It should be noted that the percentage agreement and kappa for the classification systems derived in this report are overestimates of corresponding population parameters because they were calculated from the same data used to derive their categories, which may be sample specific. Sample specificity, a well-known problem in classification studies, diminishes with increasing sample size. Because our sample is large (n = 5130), we expect the upward bias in percentage agreement and kappa to be small. The fact that our statistically based methods produced categories that, with the exception of the catchall category, "All other causes," were biologically meaningful suggests that our results are not overly sample specific.

A second potential bias of unknown magnitude and direction results from the usual assumption that autopsy reports provide the true cause of death. Unfortunately, autopsies do not identify true causes perfectly, and thus misclassifications may have produced a bias in estimated percentage agreement between the DC and true cause of death. We suspect that this bias is minimal except, perhaps, for external causes, for which the DC may be a better "gold standard" than an autopsy diagnosis. If so, however, we need not worry about how DC diagnoses of external causes agree with autopsies. If not, our results suggest that DC diagnoses are accurate enough to study external causes separately if desired. A third potential bias stems from the over- or underselection of some subpopulations for autopsy.[14] Finally, the DC coding in this series may not reflect the coding quality in other series. Many DC diagnoses in our sample were made by ABCC/RERF nosologists, who were instructed to conform to the standards of the Japanese Ministry of Health and Welfare (MHW). Representatives from the ministry visited ABCC/RERF periodically to check the coding and instruct coders. Nevertheless, conformity to MHW standards was not always achieved. Such procedural discrepancies should have little bearing on our assessment of DC accuracy at

RERF but may affect the generalizability of our detection and confirmation rate estimates. Thus, investigations similar to the current one should be conducted for different populations when possible. A contribution of this report is its demonstration that simple factor analytic methods that require only readily accessible software can be used in such studies.

By recommending consideration of the classification systems derived in this study, we are not suggesting that causes of death should be coded only into the categories we propose nor that future investigations need be restricted to these categories. There are compelling reasons, for example, to study the effects of radiation on lung cancer regardless of whether it is included in our classification system. The results of doing so, however, should be viewed with added caution in light of the relatively low detection rate for lung cancer (51%). In such cases, the need to use statistical methods that adjust for misclassification[15,16] is magnified. Sometimes other precautions are also possible. For example, at RERF, accompanying incidence analyses based on tumor registry or AHS data can be performed.

Although the results illustrated in the Figure should not dictate what the objectives of a given study should be, they can provide useful information in some cases. An important implication of these results is that no one cause or biologically meaningful subgroup of causes in the catchall category, "All other causes," is satisfactorily identified by the DC. Thus, when there is no compelling reason to study them, it may be better not to study certain (specific) disease categories than to base an investigation of them on unreliable DC data.

Although several neoplastic diseases appear to be identified satisfactorily by DC, only external causes and, perhaps, vascular diseases are diagnosed with adequate accuracy among nonneoplastic causes of death. These findings suggest that results from studies of nonneoplastic causes of death, other than external causes or vascular diseases, that do not statistically adjust for misclassification should be viewed with caution. Thus, the findings by Shimizu et al[17] of increased mortality from vascular disease in heavily exposed individuals may be regarded as more credible than their similar conclusions concerning nonneoplastic digestive disorders and specific types of vascular diseases. We observed poor accuracy rates for DC diagnoses of nonneoplastic digestive disorders and specific vascular diseases. A question arises whether the apparent dose response for the latter may have been due to misclassification of neoplasms or vascular diseases. Studies of specific neoplastic causes that do not possess high accuracy rates should also be interpreted cautiously.

The need for caution is magnified by the fact that accuracy rates for at least some causes of death vary significantly by age at death, sex, radiation exposure, place of death, and period of death. This means that even causes with high overall accuracy rates may have poor accuracy in some groups of people. For example, special caution should be exercised in studies that focus mainly on the elderly because accuracy rates typically decline with age at death. Conversely, some causes with low overall accuracy rates may have acceptable rates in certain subpopulations.

In conclusion, the hierarchical clustering method used here identified relatively few causes of death that are well diagnosed by DC. Even among these, there may be substantial misclassification rates in some subpopulations. Using

readily accessible software for performing factor analysis, we demonstrated a relatively simple method of defining diagnostic groups with nearly optimized percentage agreement. These groups could serve as an effective starting point for cause-specific mortality studies. When objectives dictate that causes with less desirable accuracy rates be considered, the negative effects of DC inaccuracies can be reduced by applying statistical procedures that adjust for misclassification. The results presented in this report should prove useful when planning studies, although it should be noted that these results apply, directly, only to the population of atomic bomb survivors.

Acknowledgments

We thank Drs. Yukiko Shimizu, Dale Preston, and Richard Sposto for helpful comments on drafts of this report.

References

1. Yamamoto T, Moriyama IM, Asano M, Guralnick L: RERF pathology studies, Hiroshima and Nagasaki, Report 4. The autopsy program and the Life Span Study, January 1961–December 1975. RERF TR 18-78
2. Steer A, Land CE, Moriyama IM, Yamamoto T, Asano A, Sanefuji H: Accuracy of diagnosis of cancer among autopsy cases: JNIH-ABCC population for Hiroshima and Nagasaki. Gann 67:625–32, 1976
3. Moriyama IM, Baum WS, Haenszel WM, Mattison F: Inquiry into diagnostic evidence supporting medical certifications of death. Am J Public Health 48:1376–87, 1958
4. Kircher T, Nelson J, Burdo H: The autopsy as a measure of accuracy of the death certificate. N Engl J Med 313:1263–9, 1985
5. Jablon S, Thompson D, McConney M, Mabuchi K: Accuracy of cause of death certification in Hiroshima and Nagasaki, Japan. Ann NY Acad Sci 609:100–9, 1990
6. Beebe GW, Usagawa M: The major ABCC samples. ABCC TR 12-68
7. World Health Organization: *International Classification of Diseases, 1955 Revision* (Vol 1). Geneva, WHO, 1957
8. World Health Organization: *International Classification of Diseases, 1965 Revision* (Vol 1). Geneva, WHO, 1967
9. World Health Organization: *International Classification of Diseases, 1975 Revision* (Vol 1). Geneva, WHO, 1977
10. Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 33:159–74, 1977
11. Hosmer DW, Lemeshow S: *Applied Logistic Regression.* New York, Wiley, 1989
12. SAS Institute Inc: *SAS User's Guide: Statistics, Version 5 Edition.* Cary, NC, SAS Institute Inc, 1985. pp 171–253, 622
13. Overall JE, Klett CJ: *Applied Multivariate Analysis.* New York, McGraw-Hill, 1972. pp 111–36
14. Ron E, Carter RL, Jablon S, Mabuchi K: Agreement between death certificate and autopsy diagnoses among atomic bomb survivors in the LSS cohort. RERF CR 6-92

15. Sposto R, Preston DL, Shimizu Y, Mabuchi K: The effect of diagnostic misclassification on noncancer and cancer mortality dose response in the RERF Life Span Study. RERF TR 4-91

16. Hoel DG, Ron E, Carter RL, Mabuchi K: The influence of death certificate errors on cancer mortality trends. RERF CR 2-91

17. Shimizu Y, Kato H, Schull WJ, Hoel DG: Life Span Study Report 11. Part 3. Noncancer mortality, 1950–85, based on the revised doses (DS86). RERF TR 2-91