

# Quality control of automatically defined cancer cases by the automated registration system of the Venetian Tumour Registry

## Quality control of cancer cases automatically registered

S. Tognazzo<sup>1</sup>, A. Andolfo<sup>1</sup>, E. Bovo<sup>1</sup>, A.R. Fiore<sup>1</sup>, A. Greco<sup>1</sup>, S. Guzzinati<sup>1</sup>,  
D. Monetti<sup>1</sup>, C.F. Stocco<sup>1</sup>, P. Zambon<sup>2</sup>

**Background:** In the Venetian Tumour Registry a substantial quota of cases (55%) is accepted using an algorithm that automatically evaluates diagnostic evidence: this study aims at assessing the reliability of the information produced in this way. **Methods:** A reabstraction study was conducted, which put a stratified sample of 1539 automatically accepted cases through a double-blind manual revision. **Results:** A significantly higher proportion of prevalent cases were found among breast, prostate and larynx cancer cases without microscopic confirmation, while there is a clear strong inverse relationship between the number of concordant diagnostic sources and the proportions of discordant diagnoses: cases based only on a single cytology record are particularly unreliable. A small number of multiple cancers are not detected because of one of the rules applied. **Conclusion:** The overall proportion of incorrect decisions is not high and similar to those reported by other registries, but errors are correlated to the diagnostic evidence pattern. As a further check, we decided to revise clinical cases for the three sites mentioned manually, in order to reduce the numbers proportion of both prevalent cases, and all cytology-based diagnoses, so as to reduce the number of 'false positives'. Coverage of hospital discharge source has been extended in order to decrease the proportion of cases based only on pathology records.

**Keywords:** automated registration, cancer incidence, quality control

The primary function of a population-based cancer registry is both to collect a set of essential informative items about all cases of cancer occurring in a defined population and to follow up registered patients over time. It represents a fundamental resource for research activity and for health service evaluation, planning and monitoring.<sup>1</sup>

In a very schematic way, the routine process to register a cancer case involves the following steps:

- collection of available information about the case;
- evaluation of the collected data and the decision about whether to register the case or not;
- updating of the archive of all cases for each required item (data storage).

In principle, automatic data processing may play a role in all the steps outlined. In practice, although electronic data storage is widely adopted in cancer registries, the automatic processing tools used for data collection vary considerably in both type and extension: indeed, there are only a few examples of automatic data evaluation in existence.<sup>1–3</sup>

The Venetian Tumour Registry (Registro Tumori del Veneto, RTV) was the first cancer registry in Europe to make extensive use of automatic data processing tools throughout the 'registration cycle'.<sup>2,4</sup>

Using a data evaluation program, 52 064 individuals were automatically accepted as incident cases, corresponding to 55% of the total number of cases registered as primary cancer for the period 1987–1996 (not melanotic skin cancer excluded); the remaining 43 351 cases (45%) were defined manually.

Our analysis has sought:

- to assess the level of concordance between automatic and manual tumour definition, in terms of site (ICD-IX three digit code) and morphology (ICD-O-1 code);<sup>5</sup>
- to evaluate the proportion of prevalent cases incorrectly classified as incident;
- to evaluate the rate of undetected second primary cancers.

## Registration system

The RTV registration system (figure 1) is based on coded data generated on a routine basis for purposes other than cancer registration and relating to three sources.

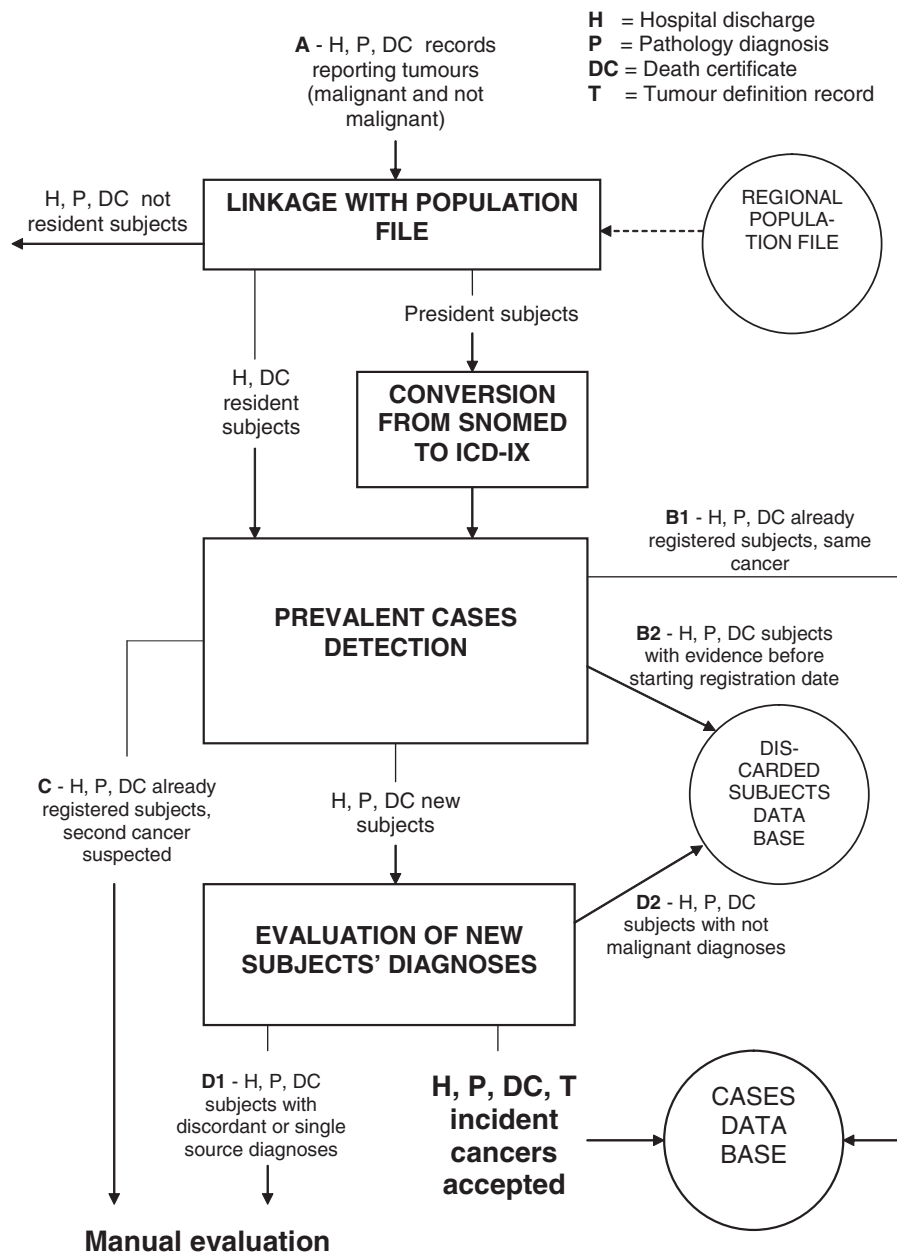
- Hospital discharge records (H): each record reports a maximum of six diagnoses, coded according to the ICD-IX classification.<sup>6</sup> Currently, files are transmitted to the RTV by the Regional Department of Social Security and Health on an annual basis and cover all the hospitals in the Veneto region. However, initially these data were transmitted by the Local Health Units (LHU) whose population was under registration (about 1 970 000 persons corresponding to 45% of the regional population).
- Pathology records (P): all the 15 pathology laboratories in the registered areas code the diagnoses using the SNOMED<sup>7</sup> classification, then send the reports to the RTV, these are then converted to ICD-IX by the Registry, which uses a specific program to do so.
- Death certificates (DC): magnetic files are transmitted by the Regional Department of Health and Social Security on an annual basis, and include all deaths that have occurred among residents in the region. The main cause of death is coded using ICD-IX classification by physicians of the Public Health Departments in each LHU.

Furthermore, all the LHUs of the region transmit a copy of the population files each year; these are pooled and put into a single

1 Venetian Tumour Registry, Azienda Ospedaliera di Padova, Padua, Italy

2 Department of Oncology, University of Padua, Padua, Italy

**Correspondence:** Sandro Tognazzo, Venetian Tumour Registry, Via Gattamelata 64, 35128 Padua, Italy, tel: +39 049 8076412, fax: +39 049 8076789, e-mail: sandro.tognazzo@unipd.it



**Figure 1** Venetian Tumour Registry data processing flow-chart

archive by the RTV. The population file is used as a reference for linking source records: if no exact match by regional health code, name, date of birth and sex can be found, the record is either discarded or manually compared with similar population records.

Data processing is carried out in the following steps:

- diagnostic source records (A) are linked with the population archive, in order to identify the diagnoses related to resident subjects and to give each one a distinct identification code (Regional Health code);
- resident subjects whose data relate to the period before registration had started are *automatically discarded* (B2), while those already registered with the same or a similar diagnoses are *automatically considered as prevalent cases* (B1);
- subjects with evidence from both before and after the starting date of registration, as well as cases already registered but with a different diagnosis, are *manually defined* (C);
- the remaining individuals, with evidence which falls into the current period, are run through a decision program that applies a set of rules drawn up to define incident

cases automatically: those subjects that do not meet such rules have to be defined manually if a diagnosis indicates a malignant tumour (D1), otherwise they remain undefined (D2). In particular, cases reported either only in hospital discharge records or only in death certificates, or in pathology records which indicate different, well-defined, primary sites, are always defined manually.

The diagnostic evidence of those cases which are automatically accepted in step D are:

- *pathology records, hospital discharge records and/or death certificate*: when the primary cancer registered is taken from pathology records and hospital discharge diagnoses or cause of death
  - give the same three-digit ICD-IX code indicated by the pathology diagnosis or
  - give only metastasis or an ill defined or unknown primary site or
  - give a ‘compatible’ primary site, close to that indicated by pathology (i.e., colon is compatible with rectum);

**Table 1** Stratification of cases automatically accepted in RTV—Incidence period 1990–1994

Diagnostic evidence: stratum	Number	of	Sample size (n)	Sampling fraction (100 × n/N)
	cases (N)	%		
Pathology report, hospital discharge and/or death certificate	19 136	75.1		
1. All concordant diagnoses	10 390	40.8	205	2.0
2. Pathology diagnoses + metastasis, ill defined or unknown site	4925	19.3	206	4.2
3. Two concordant pathology diagnoses + compatible sites	2683	10.5	201	7.5
4. Single pathology diagnosis + compatible sites	1138	4.5	183	16.1
Hospital discharge and death certificate	3248	12.7		
5. Discharge records before terminal hospitalization	2126	8.3	193	9.1
6. Discharge record from terminal hospitalization only	1122	4.4	183	16.3
Pathology report only	3093	12.1		
7. Two or more concordant diagnoses	721	2.8	167	23.2
8. Single diagnosis	2372	9.3	201	8.5
Total	25 477	100	1539	6.0

- *pathology records only*, reporting a single cancer with a well defined primary site;
- *hospital discharge records and death certificate*: the registered cancer is reported in the first source; the second is concordant or reports metastasis.

## Methods

A stratified sampling scheme was adopted, partitioning the universe of automatically accepted cases with a diagnosis of primary cancer in 1990–1994 (not melanotic skin cancers excluded) by diagnostic evidence (table 1). Eight strata were derived from the categories outlined at the end of the previous section, on the basis of the fact that different evidence leads to different probabilities of an incorrect decision:<sup>8</sup> when compatible sites are reported or acceptance is based on a single pathology record, or no pathology record is available, it is reasonable to presume that errors regarding tumour topography, morphology and date of incidence are more likely to occur. For each stratum and sex, sample size was determined so as to ensure a 95% confidence interval with maximum width 0.2 when estimating the specific error rate.

A sample of 1539 cases was drawn and put through a double-blind manual revision, based on the original sources (clinical records, pathology records in verbal form, population file records, etc.): the two revisers, a physician and a registration technician, both working at RTV and with long experience in manual evaluation of cases did not know the diagnosis that had been assigned automatically.

An ordinal scale with five discordance levels was used to express the degree of agreement between automatic and manual tumour definition in terms of site and morphology:

- None: both ICD-IX 3-digit code and morphology concordant.
- Low: ICD-IX 3-digit code concordant, morphology discordant within the same histological group.
- Medium: ICD-IX 3-digit code discordant, within the same apparatus, or discordance between histological groups.
- High: ICD-IX 3-digit code discordant, different apparatus.
- Maximum: false positive, i.e. manually defined as not malignant tumour or not tumour.

By histological groups, we mean those groups of malignant neoplasms considered to be histologically ‘different’ for the purpose of defining multiple cancers.<sup>9</sup>

To estimate the proportion of prevalent cases and of discordant cases for each discordance level, as well as the corresponding confidence interval, formulae for discrete population stratified samples were applied,<sup>10</sup> estimating the variances of each stratum with their estimators for binomial and multinomial distributions.<sup>11,12</sup>

It is of primary interest to identify the categories of diagnostic evidence where errors are more frequent, by investigating the correlation between the error rates (proportions of prevalent and discordant cases) and stratification variables (gender, number of concordant sources, number of concordant diagnoses, compatible primary site indicator, basis of diagnosis) and other potentially discriminatory factors (age group and tumour site or group of sites). Sampling distribution was binomial for the proportion of prevalent cases and multinomial for the proportion of discordant cases, so the class of logit models was appropriate to the analysis.<sup>12</sup> Given the merely operative aim of the analysis, it would have been of little use to introduce interaction terms or to adopt multiplicative models, thus simple additive models were fitted.

The ordinal nature of the discordance scale was accounted for by using the cumulative logit of each discordance level as response function (‘none’ and ‘low’ levels, quite similar in many respects, were grouped together to form the baseline category). The normal logit was used when modelling the proportion of prevalent cases.

Relevant effects were identified by the stepwise selection algorithm implemented in the LOGISTIC procedure of the SAS package.<sup>13</sup>

Differences among strata in the rate of unreported second cancers were tested by the binomial exact test.<sup>11</sup>

## Results

The results of this manual revision are summarised in table 2. The proportion of cases confirmed as incident, with none or only slight differences in site and morphology, was 85.7%; among these subjects, a second primary cancer, not registered, was detected in 15 cases. Cases confirmed as incident, but with

**Table 2** Outcome of manual revision of cases

Level of discordance with site and morphology codes automatically assigned	Number of subjects				Total
	Confirmed as incident cases 1990–1994	Redefined as prevalent cases	Resident outside the registered area	False positive	
None	1206 (78.4%) <sup>a</sup>	24 (1.6%)	–	–	1230 (79.9%)
Low	113 (7.3%) <sup>b</sup>	16 (1.0%)	1 (0.1%)	–	130 (8.4%)
Medium	86 (5.6%)	2 (0.1%)	–	–	88 (5.7%)
High	51 (3.3%)	1 (0.1%)	–	–	52 (3.4%)
Maximum	–	–	–	39 (2.5%)	39 (2.5%)
Total	1456 (94.6%)	43 (2.8%)	1 (0.1%)	39 (2.5%)	1539 (100%)

a: Fourteen subjects had also a second primary cancer not registered

b: One subject had also a second primary cancer not registered

**Table 3** Distribution of prevalent cases by diagnostic evidence

Diagnostic evidence	Prevalent cases		Total cases
	N	%	
Pathology only, Pathology and other(s) source(s):	13	1.1%	1163
Female breast (174)	4	2.4%	166
Prostate (185)	–	–	65
Larynx (161)	–	–	28
Rectum, skin melanoma, uterus, bladder (154,172,179,180,182,188)	1	0.5%	204
Stomach, colon, lung, kidney (151,153,162,189)	5	1.5%	341
Other sites	3	0.8%	359
Hospital discharge and death certificate:	30	8.0%	376
Female breast (174)	14	51.9%	27
Prostate (185)	4	26.7%	15
Larynx (161)	3	100.0%	3
Rectum, skin melanoma, uterus, bladder (154,172,179,180,182,188)	5	20.0%	25
Stomach, colon, lung, kidney (151,153,162,189)	4	2.8%	145
Other sites	–	–	161
Total	43	2.8%	1539

differences in site or morphology, made up 8.9% of the sample, while subjects found to be false positive and prevalent were 2.5% and 2.8%, respectively.

Some associations between diagnostic evidence and error rates were identified.

Subjects with clinically based diagnoses showed a significantly higher proportion of prevalent cases than subjects with histological or cytological confirmation: when modelling the logit of such proportions, only the basis of diagnosis indicator (expressed as an indicator variable clinical/not clinical) has a significant effect ( $P < 0.05$ ), with a positive regression coefficient.

Among automatically accepted cases, only subjects with diagnoses ascertained from hospital discharge records and death certificate, without pathology report, fall into this category.

As shown in table 3, the excess of prevalent cases is clearly concentrated in a few long-survival sites: female breast, prostate and larynx.

Variables significantly correlated with the proportions of discordant diagnoses detected for each level of discordance between automatic and manual definition (medium, high, maximum) are listed in table 4.

Discordant cases are more frequent when the site automatically assigned was uncertain or was referred to other sites, regarded as more compatible; discordance rates are lower when the number of concordant diagnostic sources increases, when the automatically accepted site is female breast and when the basis of diagnosis is higher. The latter association mainly reflects the difference between cytological and histological diagnoses among single source cases.

A high total number of concordant diagnoses and the automatic acceptance of a bladder cancer are associated with low discordance proportions too, but only excluded from fitting cases with three concordant sources or uncertain site.

Point estimates and confidence intervals for the proportions of discordant and prevalent cases are outlined in table 5, where the set of automatically registered cases is partitioned according to previous results. Two remarks seem appropriate:

- The quality of automatic diagnosis in terms of tumour definition is lowest in the case of a single concordant source, particularly when there are no histologically confirmed or compatible primary sites, and it is far more reliable when at

**Table 4** Variables significantly correlated with discordance levels proportions ( $\alpha = 0.01$ )

Explanatory variable	Modality	No. of cases by discordance level					Regression coefficient (a) (Standard error)
		None or low	Medium	High	Maximum	Total	
Number of concordant sources						-2.27 (0.29)	
	1	459	74	44	30	607	
		76%	12%	7%	5%	100%	
	2	740	14	7	9	770	
		96%	2%	1%	1%	100%	
	3	161	—	1	—	162	
		99%	—	1%	—	100%	
Base of diagnosis						-0.51 (0.13)	
	1. Clinical, diagnosed during terminal hospitalization	168	3	7	5	183	
		92%	2%	4%	3%	100%	
	2. Clinical, diagnosed before terminal hospitalization	188	1	2	2	193	
		97%	1%	1%	1%	100%	
	3. Cytological	59	30	22	8	119	
		50%	25%	18%	7%	100%	
	4. Histological	945	54	21	24	1,044	
		91%	5%	2%	2%	100%	
Presence of compatible primary sites						0.72 (0.20)	
	0 (No)	1,060	24	40	31	1,155	
		92%	2%	3%	3%	100%	
	1 (Yes)	300	64	12	8	384	
		78%	17%	3%	2%	100%	
Primary site uncertain (PSU) (ICD-IX 159,165,195,199)						2.02 (0.28)	
	0 (No)	1,351	65	27	37	1,480	
		91%	4%	2%	3%	100%	
	1 (Yes)	9	23	25	2	59	
		15%	39%	42%	3%	100%	
Primary site female breast						-1.46 (0.52)	
	0 (No)	1,171	86	52	37	1,346	
		87%	6%	4%	3%	100%	
	1 (Yes)	189	2	—	2	193	
		98%	1%	—	1%	100%	

(a) A positive coefficient means that high values for the explanatory variable correspond to high discordance proportions, a negative one that they correspond to low proportions

least two diagnostic sources agree. However, a certain risk of including prevalent cases exists when no pathology records are available.

- Cancers with an uncertain or unknown primary site undoubtedly give a better definition when manually revised, but there is little risk of including false positives.

Stratified whole sample estimates are also shown in table 5. They are lower than the corresponding rough percentages, since

strata with higher error rates have a larger sampling fraction. The estimated rate of prevalent cases equals 2%: the estimated proportions of false positive and of incident subjects with medium and high discordance are 1.5%, 3.5% and 1.7%, respectively, giving a total of 6.7% of subjects for whom manual and automatic definition differ as regards tumour site or morphology.

Only single cancer cases are automatically accepted by the decision program, but manual revision revealed that 15

**Table 5.** Point estimates and 95% Confidence intervals for the proportions of discordant and prevalent cases

Subpopulation	No of cases		Proportion of cases with discordance level		
	Universe	Sample	Medium	High	Maximum
Site ill-defined or uncertain (ICD IX 159,165,195,199) automatically accepted	451	59	39%	42.4%	3.4%
			(24.9–55.2%)	(27.8–58.4%)	(0.7–15.0%)
One source, cytological base	416	44	25%	11.4%	13.6%
			(12.5–43.7%)	(4.0–28.3%)	(5.2–31.1%)
One source, histological base and presence of compatible primary diagnosis	1203	160	22.5%	6.3%	3.1%
			(15.4–31.7%)	(2.9–12.9%)	(1.1–8.7%)
One source, female breast or bladder cancer automatically accepted	802	96	–	–	3.1%
			(0–4.5%)	(0–4.5%)	(0.8–11.1%)
One source, histological base; two sources, clinical base with two concordant diagnoses	4128	508	1.8%	2.2%	4.1%
			(0.8–3.9%)	(1.0–4.5%)	(2.4–7.0%)
Two sources, histological base with less than three concordant diagnoses	8550	275	2.5%	0.4%	0.7%
			(1.0–6.2%)	(0.05–2.9%)	(0.1–3.5%)
Two sources with more than three concordant diagnoses; three sources	9919	397	0.5%	–	–
			(0.1–2.4%)	(0–1.1%)	(0–1.1%)
Stratified sample estimates:					
Point estimates			3.5%	1.7%	1.5%
Confidence intervals			2.7–4.3%	1.3–2.1%	1.1–2.2%
Subpopulation	No of cases		Proportion of prevalent cases		
	Universe	Sample			
Two concordant sources, clinical base	3180	364	8.2% (5.1%–12.9%)		
All other cases, histological or cytological base	22 289	1175	1.1% (0.5%–2.3%)		
Stratified sample estimates:					
Point estimate			2.0%		
Confidence interval			1.4%–2.6%		

individuals actually had a second cancer with first diagnosis during the incidence period, while the registered tumours (all pathology based except one) were confirmed without particular discordance. Prostate and urinary organs made up 60% of the undetected sites.

The incidence of undetected second cancers is significantly higher ( $P = 0.017$ ) among subjects who had clinical diagnoses or cause of death compatible with the pathology diagnosis accepted (eight cases out of 384, corresponding to strata 3 and 4 in table 1). The estimated rate of undetected tumours for this group, 15% of the automatically defined cases, is 2.1%, against a rate of 0.6% for the remaining subjects (seven cases out of 1155).

Overall, the rate of unreported cancers among subjects with automatic definition is 0.9%.

The acceptance rules were misleading in 10 cases: for six of these, the second cancer was actually reported in the clinical record or on the death certificate; the other four had an

ill-defined diagnosis, whose morphology, in three cases, was not compatible with the cancer accepted.

As to the remaining cases, failure to detect was due to incompleteness or errors in the diagnostic sources: four subjects had no mention of a different cancer among the computerised diagnoses, one case had a SNOMED code referring to a benign tumour.

## Discussion

Among all the studies which have sought to make a quality assessment of cancer registration, one in particular, that carried out by the Ontario Cancer Registry (OCR),<sup>2</sup> which has an established experience in automatic evaluation of diagnoses, offers excellent points for comparison with this study.

Their rate of discordance as to three-digit ICD-IX codes and false positive equals ours (6.7%), but with a larger quota of false



positives (OCR 3.3%, RTV 1.5%). The OCR sample includes categories defined manually in our system and, likely, more affected by errors, so we could be expected to exhibit lower figures: considering only cases with comparable evidence, our error rates are lower. If we consider only those cases with comparable evidence, our error rates are somewhat lower when registration has been based on two or more sources (OCR from 8.6% to 14.8%, RTV from 0.5% to 8.1%), but are worse for cases based on pathology only (OCR 4.7%, RTV from 3.1% to 50%). Moving the incidence date to a different, earlier, year was carried out in 3.3% of cases in the OCR, and in about 4% in the RTV. There is no mention in the OCR of undetected second primary cancers.

Other studies are only partially comparable, because of the inclusion of not melanotic skin cancer and not malignant tumours and of cases with different diagnostic data: mainly hospital only and death certificate only. As to discordance in three-digit ICD-IX codes and false positives, the following proportions may be derived: from West,<sup>14</sup> 6%, from Lapham,<sup>15</sup> 5% and from Brewster,<sup>16</sup> 5.4%, with their false positives corresponding to 0.7%, 1% and 2%, respectively. Almost all these figures are lower than ours, but since not malignant tumours and skin tumours have been excluded from numerators, because no information which permits 'cleaning' of denominators is available, 'true' proportions are certainly higher. There are 6.1% cases where the incidence date was put back in West and 4.8% in Brewster: these figures are inflated by the inclusion of skin and not malignant tumours, thus the comparable prevalence rates should be closer to ours. Undetected second primary cancers are reported by both West (2.1%) and Lapham (0.5%); our rate (0.9%) seems closer to the latter's, especially when the difference in denominators is taken into account.

No discussion about comparison with studies focused on particular sites,<sup>17–19</sup> or particular age groups,<sup>20</sup> is presented, for we deemed it was not appropriate here.

On the whole, the automatic acceptance rules which we apply have revealed a proportion of incorrect decisions not unlike those reported by other registries. However, the errors are not uniformly distributed, since their probability and seriousness is mainly correlated to the diagnostic evidence pattern.

Three main points would seem to emerge:

- Reliability of site identification increases with the number of sources reporting the same ICD-IX code and seems adequate when at least two sources agree. As other authors have also pointed out<sup>21,22</sup> identification based only on pathology records is not very accurate: there is clearly a serious problem regarding false positives when cytology is the only available evidence. The frequent misclassification of the primary site, among single source cases, occurs mainly within the same apparatus or category and thus may be regarded as a minor, though not negligible, problem.
- Inclusion of prevalent cases is important among cases with a clinical basis which are confirmed by cause of death, with particular concern for breast, prostate and larynx tumours.
- Some multiple cancers are not detected, mostly because of the acceptance rules applied.

To deal with the first point: coverage of hospital discharge sources has been extended to all hospitals in the Region, in order to reduce the proportion of pathology-only cases, and cytology-based diagnoses have to be manually revised.

Referring to the ongoing update concerning the 1997–1999 period, the share of pathology-only cases automatically accepted has dropped from 13.2% in 1990–1994 to 7.3%.

As to the second point, we have decided to manually check cases of female breast, prostate and larynx cancers which are long-survival.

There is a low incidence of unreported second cancers and about two thirds of these latter were induced by the

acceptance rules. This detection error should be reduced by the revised version of the evaluation program.<sup>23</sup> one third of these undetected cases would now be rejected by automatic definition and subjected to manual definition because we are now running a closer check on concordance between morphologies. The rule which establishes that pathology diagnoses are, under certain conditions, preferred to diagnoses referred by hospital discharge records or death certificates, unavoidably implies some measure of misdetection. However, we do not intend to change the rule because automatic decisions do not seem to be any less accurate than those achieved with the manual process.

### Key points

- Study aim: To assess the reliability of diagnoses automatically defined by the automated registration system of the Venetian Tumour Registry.
- Main results: Among cancer cases accepted in an automated way, error rates are similar to those experienced by manual registries.
- Public health implications: The validity of automated cancer registration, which allows to monitor larger populations with given resources, is confirmed by this study.

### References

- 1 Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. Cancer registration principle and methods. (*IARC Scientific Publications N. 95*), Lyon: IARC, 1991.
- 2 Black RJ, Simonato L, Storm HH, Démaret E. Automated Data Collection in Cancer Registration. *IARC Technical Reports No. 32*. Lyon: IARC, 1998.
- 3 Carrigan C. The process of automated registration. UKACR 10th Annual Conference, Sheffield, November 2001, abstract.
- 4 Simonato L, Zambon P, Rodella S, et al. A computerized cancer registration network in the Veneto region, north east of Italy: a pilot study. *Br J Cancer* 1996;73:1436–9.
- 5 World Health Organization. *Manual of the International Classification of Diseases for Oncology*, First Edition. Geneva: World Health Organization, 1976.
- 6 World Health Organization 9th Revision Conference, 1975. *Manual of the International Statistical Classification of Diseases injuries and Causes of Death*. Vol. 1. Geneva: WHO, 1997.
- 7 College of American Pathologists. *SNOMED systematized nomenclature of medicine*, Second Edition. Northfield: College of American Pathologists, 1979.
- 8 Tognazzo S, Bertolin I, Fiore AR, et al. Controllo di qualità dei criteri di decisione automatica dell'incidenza utilizzati dal Registro Tumori del Veneto: analisi descrittiva preliminare. *III Riunione Scientifica Annuale dell'Associazione Italiana Registri Tumori*. Ferrara, marzo 1999, abstract.
- 9 Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan S. Comparability and quality control in cancer registration. (*IARC Technical Reports no.19*). Lyon: IARC, 1994.
- 10 Cochran WG. *Sampling techniques*. New York: John Wiley, 1963.
- 11 Fleiss JL. *Statistical methods for rates and proportions*. New York: John Wiley, 1981.
- 12 Agresti A. *Categorical Data Analysis*. New York: John Wiley, 1990.
- 13 SAS Institute Inc. 'SAS/STAT User's Guide—The LOGISTIC Procedure' in: *SAS OnlineDoc*, Version 8. Cary, NC: SAS Institute Inc., 2000.
- 14 West RR. Accuracy of cancer registration. *Brit J Prev Soc Med* 1976;30:187–92.
- 15 Lapham R, Waugh NR. An audit of the quality of cancer registration data. *Br J Cancer* 1992;66:552–4.
- 16 Brewster DH, Crichton J, Muir C. How accurate are Scottish cancer registration data? *Br J Cancer* 1994;70:954–9.

- 17 Kyllonen LE, Teppo L, Lehtonen M. Completeness and accuracy of registration of colorectal cancer in Finland. *Ann Chir Gynaecol* 1987;76: 185–90.
- 18 Pollock AM, Vickers N. Reliability of data of the Thames cancer registry on 673 cases of colorectal cancer: effect of the registration process. *Quality Health Care* 1995;4:1984–9.
- 19 Harvei S, Tretli, Langmark F. Quality of prostate cancer data in the Cancer registry of Norway. *Eur J Cancer* 1996;32A:104–10.
- 20 Dickinson HO, Salotti JA, Birch PJ, Reid MM, Malcom A, Parker L. How complete and accurate are cancer registrations notified by the National Health Service Central Register for England and Wales? *J Epidemiol Commun Health* 2001;55:414–22.
- 21 Brewster DH, Crichton J, Harvey JC, Dawson G, Nairn ER. Benefit and limitations of pathology databases to cancer registries data. *J Clin Pathol* 1996;49:947–9.
- 22 Moss SM, Smith JAE, Nicholas DS. The quality of histopathology data in a computerised cancer registration system: implications for future audit care. *Public Health* 1997;111:101–6.
- 23 Bovo E, Tognazzo S, Monetti D, et al. RTV-Evaluate Evidence. *Società Italiana degli Autori ed Editori—Registro Pubblico speciale per i programmi per elaboratore*. Registrazione SIAE: 22/01/2003 n° 002525, D003356, 2003.

Received 10 September 2003, accepted 5 July 2004