ELSEVIER

European Journal of Cancer 40 (2004) 2794–2803

European Journal of Cancer

www.ejconline.com

Sheep and goats: separating cervix and corpus uteri from imprecisely coded uterine cancer deaths, for studies of geographical and temporal variations in mortality

A.H. Loos [a,*], F. Bray [a], P. McCarron [b], E. Weiderpass [a,c,d,e], M. Hakama [a,f], D.M. Parkin [a]

[a] *Descriptive Epidemiology Group, International Agency for Research on Cancer, 150, cours Albert-Thomas, 69372 Lyon cedex 08, France*
[b] *Queen's University, Belfast, Northern Ireland*
[c] *Finnish Cancer Registry, Helsinki, Finland*
[d] *Cancer Registry of Norway, Oslo, Norway*
[e] *Karolinska Institutet, Stockholm, Sweden*
[f] *University of Tampere, Tampere, Finland*

Received 29 July 2004; received in revised form 31 August 2004; accepted 8 September 2004

Abstract

Analysing time trends in mortality from cancers of the cervix and corpus uteri using routine data sources (such as the World Health Organistion mortality database) involves two major problems: deaths certified as ''uterus, unspecified site'', and the presence of a combined category comprising unspecified and corpus uteri cancer deaths. To avoid misleading interpretations, the unspecified and the misclassified data must be incorporated into the analysis to produce rates that allow meaningful comparisons between populations and over time. Reallocation methods based on age- and time-specific distributions of cervix and corpus uteri cancer are applied to the unspecified deaths, while for those in the combined category, the age- and time-specific distributions of unspecified and corpus uteri cancer are considered. Adjustments of the general strategies for reallocation were developed to take into account the different quality of the data. Results from eight European countries with different degrees of coding precision are presented. The reallocation methods bring the cervix and corpus uteri mortality trends more in line with the trends for countries with more precise data while keeping the country-specific characteristics. In addition, the methods ensured the availability of time trends for corpus uteri cancer in women age 50 years and older, which were completely missing without reallocation. We propose generally applicable reallocation methods that allow valid time trend analysis of cervix and corpus uteri cancer mortality using datasets of varying precision. Our results show that any sensible analysis of time trends must involve procedures for correcting for unspecified and misclassified uterine cancer deaths. The modified data are available at http://www-dep.iarc.fr/hmp/reallocation.htm.

Keywords: Cancer; Mortality; International classification of disease; Cervix uteri; Corpus uteri; Coding; Database

1. Introduction

One of the major hurdles in interpreting temporal variations in mortality from cancers of the cervix and corpus uteri is the need to consider the influence of cancer deaths coded as ''malignant neoplasm of the uterus, part not specified''. These deaths (from this point referred to as ''unspecified''), as a proportion of all uterine cancer mortality, vary considerably between countries, and for some, over 50% of all uterine cancer deaths are so coded, although there is a tendency for the proportion to decrease over calendar time. Additionally,

* Corresponding author. Tel.: +33 472 73 80 72; fax: +33 472 73 86 50.
E-mail address: loos@iarc.fr (A.H. Loos).

0959-8049/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2004.09.007

the structure of the International Statistical Classification of Diseases, Injuries, and Causes of Death (ICD) [1–4], and differing precision of mortality data transmission by the relevant national authorities, has resulted in a combined category: corpus uteri and uterine cancer, unspecified, appearing in some routine data sources, notably the World Health Organisation (WHO) mortality databank [5].

Because of these problems, time trend analyses have either examined all uterus cancers together (including unspecified cancer deaths in the total), or used a reallocation procedure that assigns the unspecified deaths to cervix or corpus, based on probable – but strictly unknown – proportions of each.

Neither of these approaches is completely satisfactory, and in this paper we describe reallocation methods for unspecified and combined data, using age- and time-dependent proportions from country-specific standards, to deal with the problem of analysis of cervix and corpus uteri cancer mortality trends. We compare the time trends before and after reallocation, and provide an adjusted dataset for the analysis of cervix and corpus uteri cancer mortality in Europe.

2. Methods

2.1. Data sources and issues

Deaths from cancer of the uterus and the corresponding population data were extracted by cause (ICD), year of diagnosis, gender, and 5-year age group from the WHO mortality database for 33 European countries. For Germany, a combination of data from the Federal Republic of Germany and German Democratic Republic was used for the period 1983–1989, and thereafter, data from the whole country.

Data coded according to ICD-9 as 180 (cervix uteri cancer), 182 (cancer of the corpus uteri) and 179 (uterine cancer, unspecified) were extracted for countries with data series comprising 10 or more recent years. Mortality data coded at the 3-digit level of the 8th revision of the ICD combined deaths due to corpus uteri (ICD-8-182.0) and "uterus, unspecified" (ICD-8-182.9).

Fig. 1 illustrates the ICD codes for cancers of the uterus in the different revisions of ICD. For countries reporting mortality data according to the codes of the "short list", deaths from cancer of the corpus uteri were indistinguishable from "uterine cancer, unspecified" (and in ICD-7, also from "other uterine cancer"). This combined category is marked as (182 + 179) throughout this text.

2.2. Reallocation procedures

The reallocation must consider the redistribution of unspecified cancer deaths to cervix or corpus uteri cancer, as well as the reassignment of deaths in the combined category to their original coding as corpus uteri and unspecified, described in a flow chart (Fig. 2). The formulae used for the reallocation can be found in Appendix A.

ICD	"Short list"	3-digit category	4-digit category
ICD-7	A052 cervix	171 cervix	
	A053 other & unspecified	172 corpus 173 other 174 unspecified	
ICD-8	A055 cervix	180 cervix	
	A056 other uterus	181 chorioepithelioma	
		182 other	182.0 corpus 182.9 unspecified
ICD-9	B120 cervix	180 cervix	
	B122 corpus & unspecified	179 unspecified 182 corpus	
ICD-10	1037 cervix	C53 cervix	
	1038 corpus & unspecified	C54 corpus C55 unspecified	

Fig. 1. ICD-codes for uterine cancer. ICD, International Statistical Classification of Diseases, Injuries and Causes of Death.

2.2.1. Reallocation of unspecified deaths

When the proportion of uterine cancer deaths coded to the unspecified category is low (<25% of the total), the distribution of specified cancer deaths (between corpus and cervix) in the same dataset can be used to reallocate them [6]. However, when a larger proportion of deaths are in the unspecified category, an external "reference population" is used to provide appropriate proportions for reallocation purposes [7]. The reference populations used are shown in Table 1.

Thus, when unspecified deaths comprise <25% uterine cancer deaths, $\hat{p}_{180}$, the estimated proportion of cervical cancer deaths is $d_{180}/(d_{180} + d_{182})$, with d_{180} and d_{182} the number of deaths recorded as cervix and corpus uteri cancer, respectively. The ratio of deaths from cancer of the cervix to corpus is strongly age-dependent (deaths from corpus cancer are rare before the age of 50 years), so that reallocation requires proportions to be age-specific. For each single year, therefore, the data were grouped by age into intervals 0–39, 40–49, 50–59, 60–69 and 70 years and older. If <10 uterine cancer cases were observed within an age group, wider age groups were chosen to ensure robust estimation of the proportions for the respective years.

Since the estimate of $\hat{p}_{180}$ may be inaccurate when a large proportion of deaths are unspecified, we used this method only when <25% of deaths (at ages 30 years and older) were coded to the unspecified

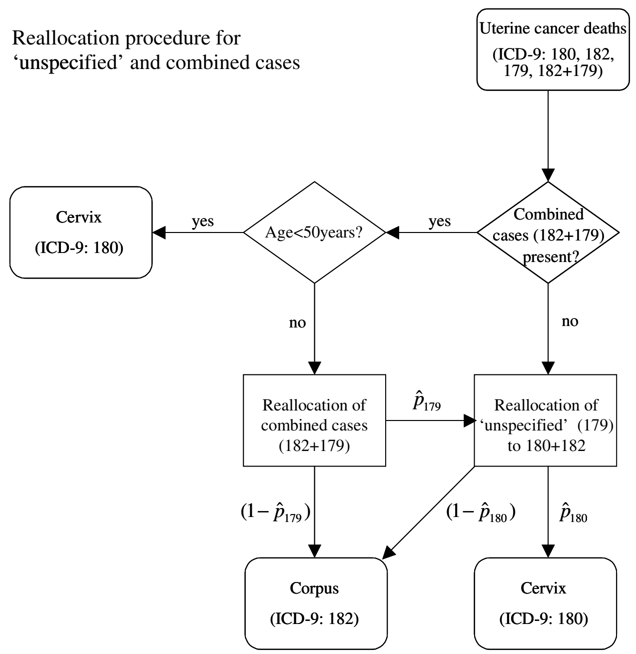

Fig. 2. Flowchart describing the mechanism of reallocation of deaths coded as "uterus, unspecified" and deaths in the combined category (ICD-9: 182 + 179).

Table 1
Reference populations, and the countries for which they provide estimates of the ratio of deaths by site within the unspecified and combined (182 + 179) categories

Reference	Countries
1. Finland	Estonia, Iceland, Latvia, Lithuania, Sweden
2. Hungary	Belarus, Bulgaria, Poland, Republic of Moldova, Russia, Romania, Slovenia, Ukraine
3. Ireland	UK, Northern Ireland
4. The Netherlands	Austria, Belgium, Croatia, France, Germany, Greece, Italy, Luxembourg, Malta, Portugal, Spain, Switzerland

UK, United Kingdom.

category, although if only a single year exceeded the upper limit of 25% unspecified, the average proportion derived from the aggregate of the preceding and subsequent year was used instead. For populations exceeding the upper level of 25% for two or more years in sequence, the proportions used to reallocate the unspecified were derived from one of four reference populations as shown in Table 1. The choice of the reference countries was made on pragmatic grounds, largely dictated by data quality, period of availability within geographical regions, and considered socio-economic factors in a very broad sense. For example, Hungary, a reference for the countries of Eastern Europe was chosen in preference for to the Czech Republic, because data were available for a longer period (1970–2001 vs. 1986–2001 for the Czech Republic).

2.2.2. Reallocation of deaths in the combined category

As explained above (see also Fig. 1), the coding of deaths to a category comprising both corpus uteri cancer deaths (182) and uterus, unspecified (179) was most common when the 8th revision of ICD was in use.

As cancer of the corpus uteri is very rare in premenopausal women, all deaths in the combined category (182 + 179) were assigned directly to cervical cancer in women under 50 years, whereas for older women a reallocation procedure is required. For each country, the mean of the age-specific ratio of corpus: unspecified from the two years preceding, and the two years following the period with the combined category was used, to reallocate deaths to corpus uteri and uterus unspecified. As for unspecified tumours, the estimated age-specific proportions were calculated for ages 50–59, 60–69, 70+ years and were widened iteratively to ensure that all proportions were calculated from at least 10 uterine cancer cases. This procedure retains the time-specific distribution of corpus and unspecified among deaths coded to the combined category. For some countries, only the combined category of corpus uteri and uterus, unspecified was available for the entire period of observation (see Table 2), and for reallocation of the deaths in these countries, the reference populations in Table 1 were used. The ratio of corpus to unspecified in the reference populations was calculated from age-specific three-year moving averages of the number of unspecified deaths divided by all deaths coded as either corpus or unspecified. A moving average was used in order to ensure a robust estimation of the year and age group specific proportions.

2.2.3. Application of the rules and pragmatic adaptations

Having evaluated the availability and the quality of the mortality data for uterine cancer, the time series were restricted to the period 1969 onwards, which resulted in exclusion of those years for which data quality was poor. As a further consequence of the reallocation process, only the time period corresponding to that for the respective reference population could be reproduced for those countries where a reference was required to estimate the number of deaths from cervix and corpus cancer. Whenever possible a within-country estimate was preferred to that based on the reference population if additional data were available. As before, averages of the ratio of cervix to corpus and corpus to unspecified, were calculated to obtain a within-country estimate of the respective proportion. Table 3 summarises these years for which a within-country proportion was applied together with the corresponding years used for their estimation. In this paper, the results are presented as age-truncated standardised mortality rates [8] applying the world standard population [9] using age groups 25–49 and 50–74 years.

Table 2
Data availability and description of the data quality for the years observed

Population	Years		Unspecified ⩽ 25% in women ⩾ 30 years	Deaths in combined category present $d_{182+179} > 0$
	Min	Max		
Austria*	1955	2002		1955–1968
Belgium**	1954	1997		1954, 1968, 1977, 1978, 1985, 1987–1991
Bulgaria**	1964	2002		1964, 1965, 1968–1979, 1983
Belarus**	1981	2001		1981–2001
Croatia*	1985	2002		
Czech Republic	1986	2001	1986–2001	
Denmark++	1951	1999	1969–1981, 1988–1999	1951–1954, 1982–1987
Estonia**	1981	2002	1994, 1998, 1999	1981–1993
Finland++	1952	2002	1968–1983, 1985–1987, 1989–2002	1952–1954, 1984, 1988
France*	1950	1999		1950–1954
Germany*	1983	2001	1989	
Greece*	1961	1999		1961–1965
Hungary++	1955	2002	1980–2002	1955–1969
Iceland**	1951	1999	1971, 1973, 1974, 1976, 1978–1988, 1990–1992, 1994, 1995, 1999	1951–1970
Ireland**	1950	2000	1969–1973, 1975–1977, 1984, 1985, 1987–2000	1950–1954, 1974
Italy*	1951	2000		1951–1954
Latvia**	1980	2002	1996	1980–1995
Lithuania**	1981	2002	1993–2002	1981–1992
Luxembourg**	1967	2002		1967–1978, 1979–1997
Malta**	1965	2002	1988, 1997, 1998, 2000, 2002	1965–1977
Republic of Moldova**	1981	2002	1996–2002	1981–1995
The Netherlands++	1950	2000	1972–2000	1950–1954, 1970, 1971
Norway	1951	2001	1969–2001	1951–1954
Poland**	1959	2001	1999–2001	1959–1960, 1969–1996
Portugal**	1955	2000		1955–1983
Romania**	1959	2002	1999–2002	1959–1998
Russia**	1980	2002		1980–2002
Slovenia*	1985	2002	1993, 1995, 1998, 1999	
Spain**	1951	2000	2000	1951–1954, 1968–1974
Sweden**	1951	2001	1969, 1987, 1989–1991	1951–1955, 1970–1986
Switzerland**	1951	2000	1969–1994	1951–1968, 1995–2000
Ukraine**	1981	2000		1981–2000
UK, England and Wales	1950	2000	1968–2000	1950–1954
UK, Scotland++	1950	2000	1969–2000	1950–1954, 1968
UK, Northern Ireland**	1950	2000	1996, 2000	1950–1954, 1968–1978

* Countries with more than 25% unspecified in at least one year in women older than 30 years.
** Countries with deaths in the combined (182 + 179) category not suitable as a reference.
++ Countries with deaths in the combined (182 + 179) category classified as a suitable reference having ⩽25% unspecified elsewhere.

3. Results

Table 2 summarises the years for which mortality data are available and the countries for which ‘‘uterus, unspecified’’ deaths are included, and indicates the years in which the proportion of unspecified deaths is less than or equal to 25%, and the years in which deaths in the combined (182 + 179) category are present.

The full results of the reallocation procedure are available on http://www-dep.iarc.fr/hmp/reallocation.htm. Here, for each country, year of death and five-year age group, the numbers of deaths are presented, as available in the WHO mortality database [5], and the estimated numbers of deaths from cancers of the cervix uteri and corpus uteri, following reallocation of deaths in the unspecified categories (the format is described in Appendix B).

The datasets of three countries, Finland, Denmark and Poland are summarised in detail in Table 4. Apart from two years (1984, 1988) Finland reported the detailed 4-digit classification list and has good data quality (⩽25% unspecified and no deaths in the combined category) throughout the period of study. Denmark has a longer period of combined data, although <25% unspecified in the other years, while detailed data for corpus uteri and unspecified cancer were available in Poland only in 1968 and 1999–2001.

The proportion of uterine cancer deaths coded to the ‘‘uterus, unspecified’’ category ranged from a median percentage of 84% (range 83–85%) in Italy to

Table 3
Summary of populations (and years after 1969) in which there are deviations from the general rule for estimation of $\hat{p}_{179}$ and $\hat{p}_{180}$

Population	Years for estimation of $\hat{p}_{179}$	Years for estimation of $\hat{p}_{180}$	Years used for mean values
Belgium**	1977, 1978		1975, 1976, 1979, 1980
	1985		1983, 1984, 1986
	1987–1991		1986, 1992–1994
Bulgaria **	1969–1979		1980–1982
	1983		1981, 1982, 1984, 1985
Denmark++	1982–1987	1982–1987	1979, 1980, 1988, 1989
Estonia**	1981–1993		1994–1998
Finland++	1984	1984	1982, 1983, 1985, 1986
	1988	1988	1986, 1987, 1989, 1990
Hungary++		1970–1979	1980–1984
Iceland**	1975		1973, 1974, 1976, 1977
Ireland**	1974		1972, 1973, 1975, 1976
		1978–1983	1976, 1977, 1984, 1985
		1986	1984, 1985, 1987, 1988
The Netherlands++	1970, 1971		Start in 1972
Portugal**	1980–1983		1984–1988
Spain**	1968–1974		1975–1979

** Countries with deaths in the combined (182 + 179) category not suitable as a reference.
++ Countries with deaths in the combined (182 + 179) category classified as a suitable reference having ⩽25% unspecified elsewhere.

approximately 3% (range 1–8%) in Norway and Finland (range 0–13%) in 1969–1978. For most of the countries, the median proportion of unspecified by age group was 10% higher in older (50–74 years) than in younger women (aged 25–49 years). Exceptions included Croatia, France, Germany, Greece, Belgium, Latvia, Malta and United Kingdom, where the reverse was observed. In general, there was a clear trend towards fewer unspecified deaths over time in all populations.

We present results for eight countries with different data quality to illustrate the effects of the reallocation procedure on the mortality rates: Latvia, Lithuania and Sweden with the reference country, Finland; and France, Italy and, Spain with the reference country, the Netherlands.

The Netherlands has low proportions of unspecified deaths (8% with range 4–11% in 1972–1981 and range 6–12% in 1991–2000), whereas the likes of Italy and France (66% with range 62–67% in 1969–1978 and 52% with range 48–55% in 1990–1999) had rather high proportions. Finland required modification of two years of data; the corrected rates seem to fit in the overall trend of the slopes). Sweden had a longer period with deaths in the combined category (182 + 179) at the beginning of the observation period (1970–1986) and approximately 25% (range 18–34%) of unspecified deaths in the most recent period of 1992–2001, whereas in Latvia (with the exception of the most recent seven years) only deaths in the combined category were available. A long period with deaths in the combined category is also noted for Lithuania, but between 1993 and 2002, only 3% (range 1–8%) of the deaths are in the unspecified category (Table 2).

Figs. 3–5 show the age-truncated mortality rates for all uterine cancer by age group, and for cervix and corpus cancers before and after reallocation, by age groups.

In Latvia and Lithuania, the mortality rate from uterine cancer is high and there is an increasing trend at young ages (25–49 years) and a slight decrease at ages older than 50 years. In the other six countries, the trends are decreasing in younger women, an observation more evident in Finland and Sweden than in the other countries (Fig. 3).

Table 4
Distribution of uterine cancer deaths over time by ICD-9 code plus combined category for the examples of Finland, Denmark and Poland

Years	Finland++				Denmark++				Poland**			
	d_{180}	d_{182}	d_{179}	$d_{182+179}$	d_{180}	d_{182}	d_{179}	$d_{182+179}$	d_{180}	d_{182}	d_{179}	$d_{182+179}$
1968	142	72	55	0	369	113	78	0	1569	383	1042	0
1969	162	95	43	0	327	115	76	0	1644	0	0	1456
1970	147	94	38	0	293	129	76	0	1697	0	0	1182
1971	162	98	21	0	348	117	60	0	1859	0	0	1132
1972	130	126	15	0	283	138	67	0	1898	0	0	1099
⋮		⋮		⋮		⋮		⋮		⋮		⋮
1979	113	124	4	0	291	148	47	0	1925	0	0	1027
1980	108	97	7	0	285	141	51	0	1997	0	0	1037
1981	98	133	6	0	321	148	58	0	1925	0	0	1049
1982	98	102	5	0	283	0	0	194	1913	0	0	1113
1983	107	107	9	0	276	0	0	230	1946	0	0	1134
1984	92	0	0	127	261	0	0	193	2069	0	0	1145
1985	95	121	9	0	257	0	0	216	2028	0	0	1162
1986	93	110	3	0	262	0	0	228	1973	0	0	1189
1987	80	126	14	0	209	0	0	194	2026	0	0	1212
1988	107	0	0	142	250	137	41	0	2063	0	0	1180
1989	80	141	18	0	234	164	65	0	2020	0	0	1111

** Countries with deaths in the combined (182 + 179) category not suitable as a reference.
++ Countries with deaths in the combined (182 + 179) category classified as a suitable reference having ⩽25% unspecified elsewhere.

For cervical cancer, the pattern that emerges after reallocation in the 25–49 age group closely resembles that for all uterine cancer (compare Figs. 3 and 4). The pattern of trends in Northern Europe is little changed by reallocation; however, in France, Spain and Italy, reallocation reveals clearly declining mortality rates in this age group, and suggests the low observed mortality rate in Italy is in fact an artefact.

At ages 50–74 years, the observed data suggest that cervix cancer mortality is declining in all eight countries. Reallocation does not change this basic pattern; however, the difference between Finland and Sweden becomes more apparent and that between Latvia and Lithuania less so. The low risk in Italy, and the change in the trend in Spain in the early 1980s, are clearly artefacts.

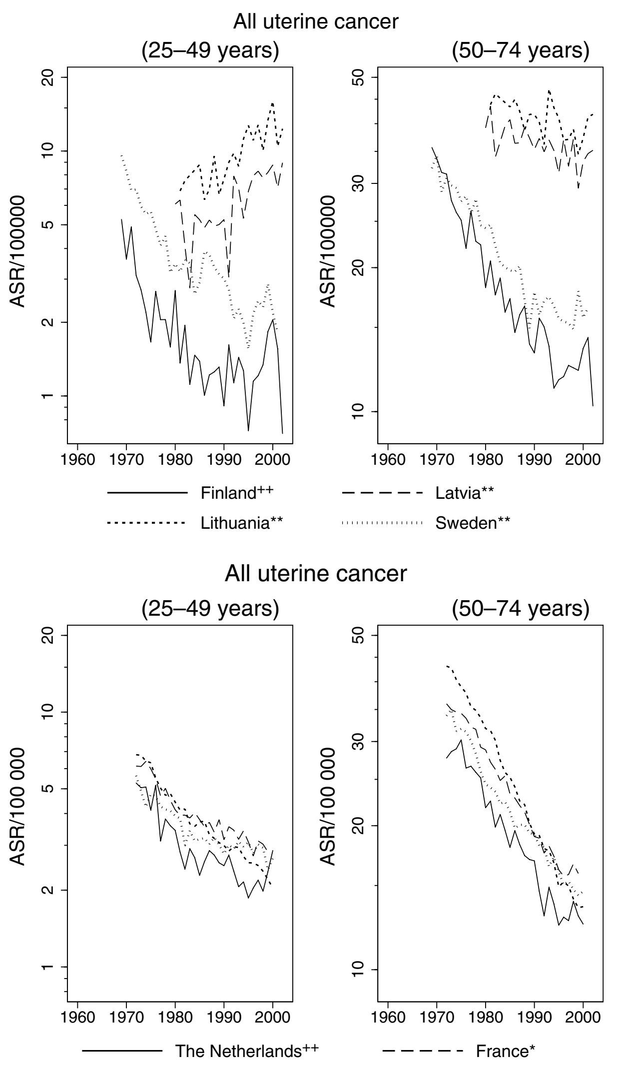

Fig. 3. Truncated age-standardised rates for all uterine cancer by age groups 25–49 and 50–74 years for Latvia, Lithuania and Sweden with the reference country, Finland; and France, Italy and Spain with the reference country, the Netherlands.

For mortality from cancer of the corpus uteri (ages 50–74 years), the reallocation procedure resulted in the trends for Latvia and Lithuania, and for Finland and Sweden, becoming very similar. This effect is also apparent for France, Italy and Spain, where, following reallocation, the rates are also of similar magnitude. The observed data for these three countries suggested much lower rates, and stable (France) or modestly increasing (Italy, Spain) trends.

Corresponding data for every country can be found at http://www-dep.iarc.fr/hmp/reallocation.htm.

4. Discussion

Problems in analysing uterine cancer mortality data are well recognised in the literature, but to date there have been few attempts to handle the problem introduced by the deaths coded as ''uterus, unspecified'' and by the combined category (corpus uteri cancer and uterus unspecified). Percy and colleagues [10] compared the cause of death on the death certificates with the hospital and pathology diagnosis to check for agreement in coding based on the follow-up of uterine cancer deaths registered by the Surveillance, Epidemiology and End-Results (SEER) in 1977–1979. They described that even during the late 1970s, 25% of the uterine cancer deaths were still not being specified and emphasised the importance of achieving further improvement in accuracy.

Bosch [11] described the difficulties of analysing mortality data from regions in Spain where the apparently increasing trends in mortality from cervical and corpus uteri cancer were largely due to a decreasing proportion of deaths certified as unspecified over the time period studied. Levi and colleagues [12] and, more recently, Peto and colleagues [13] restricted their analysis of cervical cancer mortality in Europe to women aged 20–44 years and assumed that all uterine cancer deaths in this age group were due to cervix uteri cancer. Other strategies are summarised in detail by Arbyn and Geys [7] and include, in addition, analyses restricted to certified cervical cancer deaths, pooled analyses of all uterine cancer, and combining data classified as corpus uteri cancer with unspecified deaths as a surrogate for corpus uteri cancer.

When analysing time trends, it is essential to allow for a changing ratio of cervix:corpus cancer deaths. For example, the establishment of organised screening programmes [14] can have major impacts on the relative number of deaths from cancers of the cervix and corpus uteri. The solution proposed by Jensen and colleagues [6] to the problem of unspecified was to use the ratio of cervix to corpus deaths among the unspecified. This is probably satisfactory if the proportion of unspecified deaths is low and that the cervix to corpus ratio is available for the individual age groups.

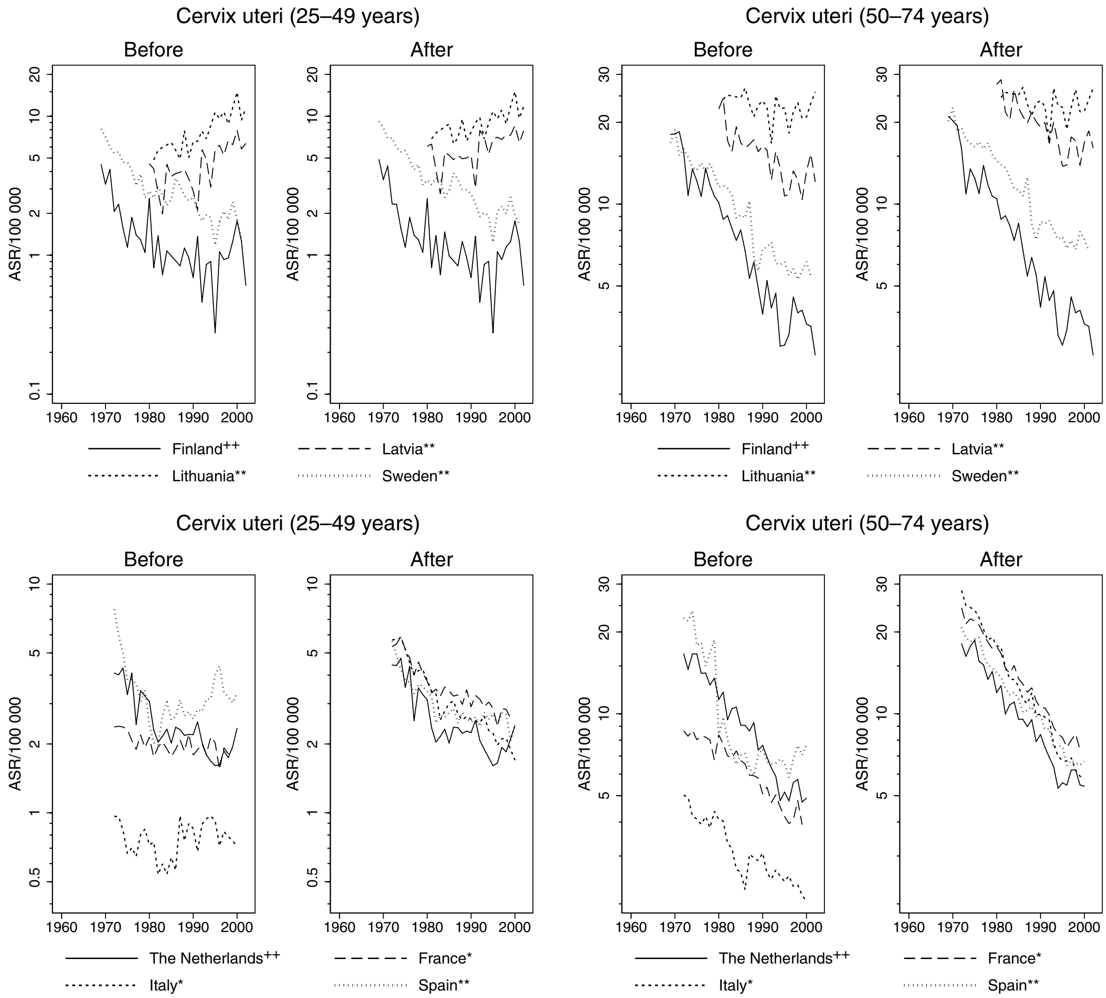

Fig. 4. Truncated age-standardised rates for cervix uteri in women aged 50–74 years before and after reallocation for Latvia, Lithuania and Sweden with the reference country, Finland; and France, Italy and Spain with the reference country, the Netherlands.

Arbyn and Geys [7] extended this approach for application of time trend analysis and derived age-group-specific estimates for European countries in four consecutive periods between 1955 and 1994 using standards derived from certain selected European countries with low proportions of unspecified deaths. Although Arbyn and Geys [7] were aware of the problem induced by the combination of corpus uteri cancer and unspecified deaths, due to the use of ICD-8, general applicable solutions to address the problem of the combined category have not been previously attempted.

For our comprehensive analysis of time trends of uterine cancer mortality with data extracted from the WHO mortality database, we had to deal not only with countries of varying coding precision over time, but also with the problem of the combined category, corpus and unspecified. As in the two approaches described, we estimated proportions of cervical cancer deaths based on populations with higher coding precision, but, unlike Arbyn and Geys [7], we used several sets of year- and age-specific estimates to retain country- or time-specific attributes by choosing well-matching reference populations, whenever possible.

The results for the eight countries chosen to demonstrate the procedure illustrate the substantial effect on the observed trends, and thus the conclusions that might be drawn in interpretation of them. In particular, reallocation showed that many of the large differences were, in fact, a consequence of data incompleteness and could be corrected by applying reallocation methods. The validity of the rates estimated following reallocation cannot be measured directly, but must be evaluated in terms of the plausibility of the revised trends in relation to those observed in populations similar with respect to socio-economic structure and external influences on mortality, especially screening policies. For example, cervical cancer mortality rates were higher in older women and tended to decrease during the 1990s. Only in Latvia and Lithuania did rates remain high (20 per 10^5 person-years), while in Finland, rates decreased to a very

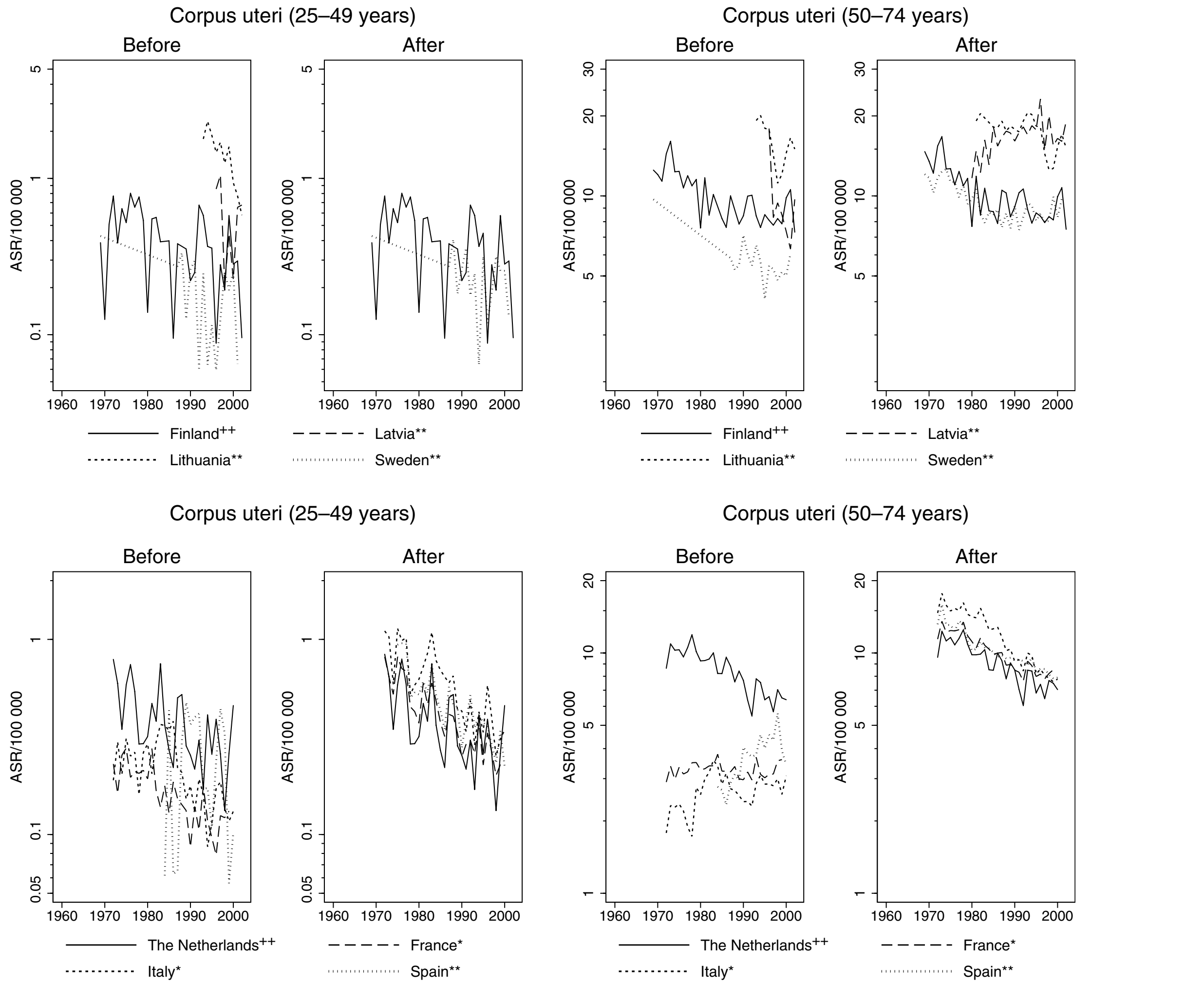

Fig. 5. Truncated age-standardised rates for corpus uteri in women aged 25–49 and 50–74 years before and after reallocation for Latvia, Lithuania and Sweden with the reference country, Finland; and France, Italy and Spain with the reference country, the Netherlands.

low level (less than 5 per 10^5). These trends and differences between countries are consistent with data on screening and its organisation [14–21].

The means of dividing the countries and choosing the reference populations necessarily limited the socio-economic homogeneity within the divisions, and, similarly, homogeneity with respect to screening coverage is also limited. We nevertheless believe that any heterogeneity will have had only a small influence on the validity of the reallocation exercise. While a major impact of the reference country on the reallocated unspecified deaths could be expected if the percentage of unspecified is high (e.g. France, Spain), any effect is likely to have been on the magnitude, but not the direction of the slope of the mortality rate.

We conclude that, for international comparisons of mortality from cancers of the cervix and corpus uteri, and for studies of between- and within-country time trends, the use of data as simply recorded in the routine mortality statistics can be seriously misleading. Reallocation procedures are required to produce estimates of the most probable true rate of mortality. Ideally, country-, year- and age-specific information is required in the reallocation process for time trends analysis, especially when deaths in the combined category are present. It is also preferable to rely on country-specific characteristics of the distribution of corpus uteri cancer and unspecified over time than to use surrogates from a reference population, which might be too heterogeneous and could add extra variability to the trends. The revised estimates of the number deaths from cancer of the corpus uteri and cervix uteri have been made available to others wishing to undertake comparative analyses of European data (see Appendix B for details)

Conflict of interest statement

None declared.

Acknowledgements

This study was part of the Comprehensive Cancer Monitoring Programme in Europe (CaMon) project funded by the European Commission, Agreement No. Sl2.327599 (2001CVG3-512), and of the European Cervical Cancer Screening Network, funded by the European Commission sub-contract agreement No. SPC2002475. P. McCarron is supported by a Department of Health (UK) public health career scientist award. E Weiderpass is supported by a grant from the Swedish Cancer Society.

Appendix A. Reallocation algorithm

Let $\hat{p}_{180} = d_{180}/(d_{180} + d_{182})$ be some estimated proportion of cervical cancer deaths (d_{180}) based on the observed number of uterine cancer deaths and $\hat{p}_{179} = d_{179}/(d_{179} + d_{182})$ in case of deaths in the combined category some estimated proportion of unspecified cases (d_{179}) over unspecified and deaths in the combined category ($d_{182\,+\,179}$). Applying the rule of reallocation, the following numbers of deaths in the combined category are added to uterine cancer deaths unspecified and corpus such that the final number of cervix and corpus uteri cancer deaths after reallocation can be expressed as

$$\hat{d}_{180} = d_{180} + (d_{179} + d_{182+179}\hat{p}_{179})\hat{p}_{180}$$

$$\hat{d}_{182} = d_{182} + d_{179} + d_{182+179} - (d_{179} + d_{182+179}\hat{p}_{179})\hat{p}_{180}, \quad \hat{p}_{179}, \hat{p}_{180} \in (0,1).$$

The rule of direct assignment of all deaths in the combined category in women aged 25–49 years to cervical cancer, and the fact that the presence of deaths in the combined category, excludes the presence of unspecified reduces this algorithm to the simpler formulae

$$\hat{d}_{180} = d_{180} + d_{182+179}$$

$$\hat{d}_{182} = d_{182}, \quad \text{if deaths in the combined category are present}$$

and

$$\hat{d}_{180} = d_{180} + d_{179}\hat{p}_{180}$$

$$\hat{d}_{182} = d_{182} + d_{179} - d_{179}\hat{p}_{180}, \quad \text{else.}$$

Appendix B. Explanations to the revised data

The data file raw_data_adj.csv in the zip-file raw_data_adj.zip downloadable from http://www-dep.iarc.fr/hmp/reallocation.htm contains the mortality data for uterine cancer for 36 European countries or regions extracted from the WHO mortality database published in January 2004. The file comprises the mortality data reported in the coding format ICD-7, ICD-8, ICD-9 and ICD-10 starting with the year 1969.

The data file includes the following variables:

reg_nr	internal country/region identity number
Reg	country/region name
Period	year
Age	age
Sex	coded as 1 for male (not applicable) and 2 for female
P	population figure
*d*_new180	number of cervix uteri cancer deaths after reallocation
*d*_new182	number of corpus uteri cancer deaths after reallocation
*d*179	original number of uterine cancer deaths coded unspecified as defined here
*d*180	original number of deaths coded as cervix uteri cancer
*d*182	original number of deaths coded as corpus uteri cancer
*d*307	number of deaths in the combined category corpus uteri cancer and unspecified
*p*180_182	proportion used to reallocate the unspecified to cervix uteri (remaining deaths were then defined to be corpus uteri)
*p*182_307	proportion used to reallocate the deaths in the combined category to corpus uteri (remaining deaths were then defined to be unspecified)

References

1. *International Statistical Classification of Diseases, Injuries, and Causes of Death*. 7th Revision. Geneva, World Health Organization, 1957.
2. *International Statistical Classification of Diseases, Injuries, and Causes of Death*. 8th Revision. Geneva, World Health Organization, 1967.
3. *International Statistical Classification of Diseases, Injuries, and Causes of Death*. 9th Revision. Geneva, World Health Organization, 1977.
4. *International Statistical Classification of Diseases and Related Health Problems*. 10th Revision. Geneva, World Health Organization, 1992, 1–3.
5. *WHO Statistical Information System (WHOSIS), Mortality Database*. January 2004. Available from http://www3.who.int/whosis/menu.cfm.
6. Jensen OM, Esteve J, Moller H, Renard H. Cancer in the European Community and its member states. *Eur J Cancer* 1990, **26**, 1167–1256.
7. Arbyn M, Geys H. Trend of cervical cancer mortality in Belgium (1954–1994): tentative solution for the certification problem of unspecified uterine cancer. *Int J Cancer* 2002, **102**, 649–654.
8. Esteve J, Benhamou E, Raymond L. Statistical methods in cancer research. Volume IV. Descriptive epidemiology. *IARC Sci Publ*, 1–302.

9. Doll R, Payne P, Waterhouse J, eds. *Cancer Incidence in Five Continents: A Technical Report*. Berlin, Springer, 1966., Ref Type: Serial (Book, Monograph).
10. Percy CL, Horm JW, Young Jr JL, Asire AJ. Uterine cancers of unspecified origin – a reassessment. *Public Health Rep* 1983, **98**, 176–180.
11. Bosch FX. Trends in cervical cancer mortality. *J Epidemiol Commun Health* 1999, **53**, 392.
12. Levi F, Lucchini F, Negri E, Franceschi S, la Vecchia C. Cervical cancer mortality in young women in Europe: patterns and trends. *Eur J Cancer* 2000, **36**, 2266–2271.
13. Peto J, Gilham C, Fletcher O, Matthews FE. The cervical cancer epidemic that screening has prevented in the UK. *Lancet* 2004, **364**, 249–256.
14. Linos A, Riza E. Comparisons of cervical cancer screening programmes in the European Union. *Eur J Cancer* 2000, **36**, 2260–2265.
15. Anttila A, Nieminen P. Cervical cancer screening programme in Finland. *Eur J Cancer* 2000, **36**, 2209–2214.
16. Schaffer P, Sancho-Garnier H, Fender M, Dellenbach P, Carbillet E, Monnet E, *et al*. Cervical cancer screening in France. *Eur J Cancer* 2000, **36**, 2215–2220.
17. Segnan N, Ronco G, Ciatto S. Cervical cancer screening in Italy. *Eur J Cancer* 2000, **36**, 2235–2239.
18. Fernandez Calvo MT, Hernandez RA, Rosell AI. Cervical cancer screening in Spain. *Eur J Cancer* 2000, **36**, 2250–2254.
19. Dillner J. Cervical cancer screening in Sweden. *Eur J Cancer* 2000, **36**, 2255–2259.
20. Van Ballegooijen M, Hermens R. Cervical cancer screening in the Netherlands. *Eur J Cancer* 2000, **36**, 2244–2246.
21. Aleknaviciene B, Smailyte G, Elaawar B, Kurtinaitis J. Cervical cancer: recent trends of incidence and mortality in Lithuania (Kaunas). *Medicina* 2002, **38**, 223–230.