Vulvar Cancer in Germany: Increase in Incidence and Change in Tumour Biological Characteristics from 1974-2013

SUPPLEMENTARY METHODS

Based on a missing at random (MAR) assumption, multiple imputation (MI) was used to derive datasets with each missing value replaced with a set of plausible values that represent the uncertainty about the missing information. MI by chained equations, a method closely related to Gibbs sampling, was implemented according to a recently proposed approach for data from population-based cancer registries [21].

The data used for the imputation model included information on the age at diagnosis (categories: <= 54, 55-74, and >= 75 years), tumour morphology (squamous cell carcinoma, basal cell carcinoma, adenocarcinoma, Paget disease, malignant melanoma, and other/unspecified neoplasms), histopathologic grade (G1, G2, or G3/4), tumour size (T1, T2, T3, or T4 according to the TNM staging scheme based on editions 4 and 6 of the classification from 1989 and 2003 onward, respectively), extent of disease (no metastases, involvement of regional lymph nodes, or involvement of distant sites and organs), duration of follow-up (<= 11, 12–59, and >=60 months), vital status at end of follow-up, and cause of death (alive, death from vulvar cancer [including unknown primary], death from other causes). For patients with a death certificate only notified vulvar cancer, the mean survival time of patients who had died in the same calendar year was used.

For subsequent calendar periods of 5 years (1989-93, 1994-98, …, 2009-2013), multinomial logit models were estimated to obtain the sampling distributions for the missing values conditional on the values of the other variables. The chosen approach thus assumed that the lack of values was at random within each stratum of the included model variables and

therefore adjusted for differences in the completeness of information between subgroups of patients with regard to patient and tumour characteristics (for example, between those with favourable and those with poor prognosis). After 5 initial iterations, 15 completed datasets were sampled and used for the analyses (convergence was assessed visually based on the estimated logits).

For descriptive analyses and incidence estimation, the estimates derived for each dataset with sampled values were combined according to Rubin's method [37]. Chi-squared tests for homogeneity to test for differences in patient and tumour characteristics over time or across age groups (H0: no differences in the respective distribution of tumor characteristics across the calendar periods 1989-93, 1994-98, …, 2009-13 (Table 1, Supplementary Table 2) or H0: no differences in the distribution of tumour characteristics between patients aged <= 54, 55-74, and >= 75 years in 2009-13 [Table 2, Supplementary Table 3]) or Mood's tests (H0: no difference in the median age at diagnosis across the calendar periods [Table 1, Supplementary Table 2]) were used. The test statistics derived for each dataset with imputed values were combined to provide one single *P* value [44].