**Table 1.** REMARK profile of patients, variables and statistical analyses

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **(a) Patients, treatments and variables** | | | | |
| **Study and marker** | **Remarks** | | | |
| Marker | EBV = Epstein Barr virus infection (negative, positive) | | | |
| Further variables | v1 = age, v2 = parity, v3 = performance status by ECOG, v4 = FIGO stage, v5 = leucocytesa, v6= neutrophils, v7 = lymphocytes, v8 = monocytes, v9 = platelets, v10 = RDW-CVb, v11 = RDW-SD, v12 = neutrophil to lymphocyte ratio, v13 = lymphocyte to monocyte ratio, v14 = serum albuminc, v15 = received treatment for cancerd | | | |
|  |  |  |  |  |
| **Patients** | **n** | **Remarks** | | |
| Assessed for eligibility | 189 | **Disease:** Squamous cervical cancer. **Patient source:** Clinical records of Hospital Rebagliati, Lima, Peru. All patients with anatomopathological diagnosis between 2013 and 2014. **Sample source:** Archived specimens available (formalin fixed, paraffin embedded cervical tissue samples, Hospital Rebagliati, Lima, Peru). | | |
| Excluded | 90 | 4 clinical records missing, 6 adenocarcinoma due wrong diagnosis recorded in clinical records, 16 carcinoma in situ or minimally invasive, 3 previous cervical cone biopsy or hysterectomy, 61 paraffin embedded cervical tissue samples missing. | | |
| Included | 99 | Age ≥ 18 years old; anatomopathological diagnosis of CC would be done in Hospital Rebagliati; and clinical stages of CC were IIB, IIIA or IIIB | | |
|  |  |  |  |  |
| With outcome events | 58 | OS: death from any cause | |  |
|  |  |  |  |  |
| **(b) Statistical analyses of survival outcomes** | | | | |
| **Analysis** | **Patients** | **Events** | **Variables considered** | **Results/remarks** |
| A1: Overall survival curves for EBV-positive and EBV- negative estimated by KM method | 99 | 58 | EBV | Figure 2 |
| A2: Crude Risk Difference from bivariable analysis using a pseudo-observations approach | 99 | 58 | EBV | Table 3 |
| A3: Bivariable in a simple Cox PH regression model1 | 99 | 58 | EBV, v1, v2, v3, v4, v6, v7, v8, v9, v11, v12, v13 | Table 4 |
| A4: Multivariable Cox PH regression model | 99 | 58 | EBV, v1, v2, v3, v4, v6, v7, v8, v9, v11, v12, v13 | Table 4 |
| C1: Check of PH assumption in A4 | 99 | 58 | EBV, v1, v2, v3, v4, v6, v7, v8, v9, v11, v12, v13 | Non-significant result of FPT transformation; Supplementary Figure 2 |
| C2: Check of linearity assumption in A4 | 99 | 58 | EBV, v1, v2, v3, v4, v6, v7, v8, v9, v11, v12, v13 | Significant result of FP transformation for age, NLR an PLR; Supplementary Figure 3 |
| C3: Check influential points in A4 | 99 | 58 | EBV, v1, v2, v3, v4, v6, v7, v8, v9, v11, v12, v13 | No absolute values of DFBETAs for EBV higher than 0.2; Supplementary Figure 4 |
| C4: Internal validation of A4 through stability analysis using bootstrapping | 99 | 58 | EBV, v1, v2, v3, v4, v6, v7, v8, v9, v11, v12, v13 | Relatively high BIF for EBV in A4 and in sensitivity analysis |
| A5: Adjusted Risk Difference from multivariable analysis using a pseudo-observations approach and variables from A4 | 99 | 58 | EBV | Table 3 |
| A6: Overall survival curves for EBV-positive and EBV-negative estimated by A4 | 99 | 58 | EBV | Fig 2 |
| ECOG: Eastern Cooperative Oncology Group, FIGO: Fédération Internationale de Gynécologie et d’Obstétrique, RDW-CV: red blood cell distribution width coefficient for variation, RDW-SD: red blood cell distribution width standard deviation, PH: proportional hazard, FPT: fractional polynomial time, FP: fractional polynomial; BIF: bootstrapped inclusion frequency a Not considered for survival analysis due to collinearity with v6, v7 and v8. b Not considered for survival analysis due to collinearity with v11. We selected v11, instead of v10, because the latter is an indirect measure of RDW (based on other hematological indicators), while the former is direct measure of the red blood cell volume distribution. c Not considered for survival analysis due to the number of missing values was relatively large. d Not considered for survival analysis because it is not a variable measured at the origin time of before it.  1 Assuming a linear functional form for v1, v6 to v9, and v11 to v13. 2 Multivariable fractional polynomial modelling of v1, v6 to v9, and v11 to v13. | | | | |