Supplementary Table 1 | **Accuracy of different test strategies.**

The pooled proportions of test positives from the different CIN grading strategies with positivity definitions I to VII are shown separately for each *Reference Standard* category (i.e., *no dysplasia*, *cervical intraepithelial neoplasia (CIN) 1*, *CIN2*, *CIN3* and *cervical squamous cell carcinoma (SCC)*). A proportion of 1.0 represents 100% accuracy for identification of the *Reference Standard*, whereas a proportion of 0.0 represents 0% accuracy for identification of the *Reference Standard*.

| | No dysplasia (n=35) Proportion | CIN1 (n=19) Proportion | CIN2 (n=17) Proportion | CIN3 (n=22) Proportion | SCC (n=22) Proportion |
|---|---|---|---|---|---|
| **1. H&E score CIN2+ vs <CIN2** | 0.019 | 0.193 | 0.765 | 0.970 | 1.000 |
| **2. H&E score CIN3+ vs <CIN3** | 0.010 | 0.053 | 0.275 | 0.818 | 1.000 |
| **3. H&E + IHC score CIN2+ vs <CIN2** | 0.010 | 0.175 | 0.863 | 0.985 | 1.000 |
| **4. H&E + IHC score CIN3+ vs <CIN3** | 0.000 | 0.018 | 0.314 | 0.848 | 1.000 |
| **5. Immunoscore 4+ vs <4** | 0.019 | 0.281 | 0.902 | 1.000 | 1.000 |
| **6. Immunoscore 5+ vs <5** | 0.019 | 0.158 | 0.824 | 0.985 | 1.000 |
| **7. Immunoscore 6  vs <6** | 0.010 | 0.018 | 0.549 | 0.955 | 0.985 |

Supplementary Table 2 | **Differences in proportions of test positives between test strategies.**

This table shows the absolute differences between the pooled proportions of test positives of strategies I to VII (all combinations) within each *Reference Standard* category (i.e. *no dysplasia, cervical intraepithelial neoplasia (CIN) 1* (a), *CIN2* (b), *CIN3* (c) and *cervical squamous cell carcinoma (SCC)*). Within each *Reference Standard* category, we calculated the absolute differences between the pooled proportions of test positives for each combination of CIN grading strategies I to VII with 95% Confidence Intervals (95%CI). If a confidence interval did not include the value 0, the difference was considered statistically significant. Significant differences are shown in bold and partly shown in Figure 4 by ∆.

For example, within *Reference Standard CIN2*, the pooled proportion of test positives for strategy I is 0.765 (see Supplementary Table 1). The pooled proportion of test positives for strategy VII is 0.549. The significant difference in positivity between these strategies is 0.765 – 0.549 = 0.216 (95%CI 0.06 – 0.372). This number corresponds to the ∆ 0.216* shown Figure 4.

Within Reference Standard no dysplasia and SCC, none of the differences were significant and data are not shown.

**a) Reference Standard CIN1**

| | 1. H&E score CIN2+ Difference (95%CI) | 2. H&E score CIN3+ Difference (95%CI) | 3. H&E + IHC score CIN2+ Difference (95%CI) | 4. H&E + IHC score CIN3+ Difference (95%CI) | 5. Immunoscore 4+ Difference (95%CI) | 6. Immunoscore 5+ Difference (95%CI) |
|---|---|---|---|---|---|---|
| 1. H&E score CIN2+ | - | - | - | - | - | - |
| 2. H&E score CIN3+ | **0.140 ( 0.033 - 0.248 )** | - | - | - | - | - |
| 3. H&E + IHC score CIN2+ | 0.018 ( -0.091 - 0.126 ) | **0.123 ( 0.007 - 0.238 )** | - | - | - | - |
| 4. H&E + IHC score CIN3+ | **0.175 ( 0.059 - 0.292 )** | 0.035 ( -0.030 - 0.100 ) | **0.158 ( 0.046 - 0.270 )** | - | - | - |
| 5. Immunoscore 4+ | 0.088 ( -0.100 - 0.276 ) | **0.228 ( 0.091 - 0.365 )** | 0.105 ( -0.239 - 0.028 ) | **0.263 ( 0.131 - 0.395 )** | - | - |
| 6. Immunoscore 5+ | 0.035 ( -0.101 - 0.171 ) | 0.105 ( -0.216 - 0.006 ) | 0.018 ( -0.103 - 0.138 ) | **0.140 ( 0.033 - 0.248 )** | **0.123 ( 0.020 - 0.226 )** | - |
| 7. Immunoscore 6 | **0.175 ( 0.059 - 0.292 )** | 0.035 ( -0.030 - 0.100 ) | **0.158 ( 0.046 - 0.270 )** | 0.000 ( -0.018 - 0.018 ) | **0.263 ( 0.131 - 0.395 )** | **0.140 ( 0.033 - 0.248 )** |

**b) Reference Standard CIN2**

| | 1. H&E score CIN2+ Difference (95%CI) | 2. H&E score CIN3+ Difference (95%CI) | 3. H&E + IHC score CIN2+ Difference (95%CI) | 4. H&E + IHC score CIN3+ Difference (95%CI) | 5. Immunoscore 4+ Difference (95%CI) | 6. Immunoscore 5+ Difference (95%CI) |
|---|---|---|---|---|---|---|
| 1. H&E score CIN2+ | - | - | - | - | - | - |
| 2. H&E score CIN3+ | **0.490 ( 0.333 - 0.647 )** | - | - | - | - | - |
| 3. H&E + IHC score CIN2+ | 0.098 ( -0.199 - 0.003 ) | **0.588 ( 0.434 - 0.743 )** | - | - | - | - |
| 4. H&E + IHC score CIN3+ | **0.451 ( 0.295 - 0.607 )** | 0.039 ( -0.135 - 0.056 ) | **0.549 ( 0.393 - 0.705 )** | - | - | - |
| 5. Immunoscore 4+ | 0.137 ( -0.279 - 0.004 ) | **0.627 ( 0.475 - 0.780 )** | 0.039 ( -0.180 - 0.101 ) | **0.590 ( 0.434 - 0.743 )** | - | - |
| 6. Immunoscore 5+ | 0.059 ( -0.216 - 0.098 ) | **0.549 ( 0.393 - 0.705 )** | 0.039 ( -0.113 - 0.192 ) | **0.510 ( 0.667 - 0.353 )** | 0.078 ( -0.015 - 0.172 ) | - |
| 7. Immunoscore 6 | **0.216 ( 0.060 - 0.372 )** | **0.275 ( 0.121 - 0.428 )** | **0.314 ( 0.145 - 0.482 )** | **0.235 ( 0.383 - 0.087 )** | **0.353 ( 0.202 - 0.504 )** | **0.275 ( 0.132 - 0.417 )** |

**c) Reference Standard CIN3**

| | 1. H&E score CIN2+ Difference (95%CI) | 2. H&E score CIN3+ Difference (95%CI) | 3. H&E + IHC score CIN2+ Difference (95%CI) | 4. H&E + IHC score CIN3+ Difference (95%CI) | 5. Immunoscore 4+ Difference (95%CI) | 6. Immunoscore 5+ Difference (95%CI) |
|---|---|---|---|---|---|---|
| 1. H&E score CIN2+ | - | - | - | - | - | - |
| 2. H&E score CIN3+ | **0.152 ( 0.050 - 0.253 )** | - | - | - | - | - |
| 3. H&E + IHC score CIN2+ | 0.015 ( -0.060 - 0.029 ) | **0.170 ( 0.062 - 0.272 )** | - | - | - | - |
| 4. H&E + IHC score CIN3+ | **0.121 ( 0.027 - 0.215 )** | 0.030 ( -0.087 - 0.026 ) | **0.136 ( 0.038 - 0.234 )** | - | - | - |
| 5. Immunoscore 4+ | 0.030 ( -0.087 - 0.026 ) | **0.182 ( 0.074 - 0.290 )** | 0.015 ( -0.060 - 0.029 ) | **0.150 ( 0.050 - 0.253 )** | - | - |
| 6. Immunoscore 5+ | 0.015 ( -0.082 - 0.051 ) | **0.167 ( 0.062 - 0.272 )** | 0.000 ( -0.057 - 0.057 ) | **0.136 ( 0.234 - 0.038 )** | 0.015 ( -0.029 - 0.060 ) | - |
| 7. Immunoscore 6 | 0.015 ( -0.066 - 0.097 ) | **0.136 ( 0.028 - 0.244 )** | 0.030 ( -0.044 - 0.104 ) | **0.106 ( 0.006 - 0.207 )** | 0.045 ( -0.020 - 0.111 ) | 0.030 ( -0.026 - 0.087 ) |

Supplementary Table 3 | **Reproducibility of test strategies.**

The average of absolute agreements amongst pathologists for the different CIN grading strategies I to VII are shown separately for each *Reference Standard* category *no dysplasia, cervical intraepithelial neoplasia (CIN) 1*, *CIN2*, *CIN3* and *cervical squamous cell carcinoma (SCC)*. An absolute agreement of 1.0 indicates that all pathologists gave the same score (for example, immunoscore 4+ or <4) for each specimen (100% agreement), whereas an absolute agreement of 0 indicates that all pathologists gave a different score for each specimen (0% agreement).

| | No dysplasia (n=35) Absolute Agreement | CIN1 (n=19) Absolute Agreement | CIN2 (n=17) Absolute Agreement | CIN3 (n=22) Absolute Agreement | SCC (n=22) Absolute Agreement |
|---|---|---|---|---|---|
| **1. H&E score CIN2+ vs <CIN2** | 0.962 | 0.684 | 0.804 | 0.939 | 1.000 |
| **2. H&E score CIN3+ vs <CIN3** | 0.981 | 0.895 | 0.451 | 0.697 | 1.000 |
| **3. H&E + IHC score CIN2+ vs <CIN2** | 0.981 | 0.649 | 0.765 | 0.970 | 1.000 |
| **4. H&E + IHC score CIN3+ vs <CIN3** | 1.000 | 0.965 | 0.373 | 0.727 | 1.000 |
| **5. Immunoscore 4+ vs <4** | 0.962 | 0.684 | 0.843 | 1.000 | 1.000 |
| **6. Immunoscore 5+ vs <5** | 0.962 | 0.789 | 0.725 | 0.970 | 1.000 |
| **7. Immunoscore 6 vs <6** | 0.981 | 0.965 | 0.529 | 0.909 | 0.970 |

Supplementary Figure 1 | **Accuracy of different CIN grading strategies including single p16$^{ink4a}$ scores.**

The proportion of test positives for the different CIN grading strategies I to VII are shown separately for each *Reference Standard* category; *no dysplasia, cervical intraepithelial neoplasia (CIN) 1, CIN2, CIN3* and *cervical squamous cell carcinoma (SCC)*. The haematoxylin and eosin (H&E) score (I and II) is based solely on morphologic characteristics, the H&E + immunohistochemistry (IHC) score (III and IV) is rendered after interpretation of Ki-67 and p16$^{ink4a}$ immunostains, and the single p16$^{ink4a}$ score (V, VI and VII) is the value of the scoring for p16$^{ink4a}$ immunostains independently scored from H&E staining by the pathologist. The number of sole p16$^{ink4a}$ positive lesions (for either strategy V, VI or VII) is relatively high compared to the combined Ki-67 and p16$^{ink4a}$ immunoscore.