Supplementary Materials: Appendix 1—Mathematical Model

The purpose of this section is to provide the mathematic basis for the high-quality gold standard whereby a series of steps can be used to improve the overall level of agreement between pathologists. Since the true diagnosis is unknown, it is not possible to determine the true rates of correct diagnosis. However, if a few simplifying assumptions are made, probabilities of correct diagnosis can be calculated. These assumptions include: a) The 2 reviewers and adjudicator all have equal probability of misdiagnosis and this probability is independent of the other reviewer and/or adjudicator; b) The probability of misdiagnosis is independent of the individual slide (i.e., each slide has an equal probability of misdiagnosis); c) when the two reviewers disagree on a slide's diagnosis, one of the two reviewers (which one is unknown) is assumed to be correct in diagnosis and the other reviewer is assumed to have an incorrect diagnosis.

Part A: Initial Assumptions

Let Pw = probability of a reviewer or adjudicator misdiagnosing a particular slide

Pc = probability of a reviewer or adjudicator correctly diagnosing a particular slide

As such,

Pw(Reviewer A) = probability "Reviewer A" misdiagnoses a particular slide

Pc(Reviewer A) = probability "Reviewer A" correctly diagnoses a particular slide

Pw(Slide 1) = probability "Slide 1" is misdiagnosed by a particular reviewer or adjudicator

1) ***Assumption 1:*** The two reviewers and the adjudicator all have an equal probability of misdiagnosis; and, each reviewer's (or adjudicator's) probability of misdiagnosing a particular slide is independent of the other reviewer (and/or adjudicator).

Pw(Reviewer A) = Pw(Reviewer B) = Pw(Adjudicator)

Pc(Reviewer A) = Pc(Reviewer B) = Pc(Adjudicator)

2) ***Assumption 2:*** Assume that this probability of misdiagnosis is independent of a particular slide, that is the probability of misdiagnosis is the same for every slide.

Pw(Slide 1) = Pw(Slide 2) = Pw(Slide 3)

Pc(Slide 1) = Pc(Slide 2) = Pc(Slide 3)

3) ***Conclusion 1:*** The diagnosis of a particular slide by a particular reviewer (or adjudicator) has two and only two mutually exclusive outcomes: either i) correct diagnosis or ii) wrong diagnosis. Thus, the probabilities add up to 1 (i.e., 100%).

1 = Pc + Pw

Pc = 1 – Pw

Part B: Estimation of the Probability of Misdiagnosis

4) a) Probability of two reviewers agreeing on diagnosis: 591/846 = 0.699

b) Probability of two reviewers disagreeing on diagnosis: 255/846= 0.301

5) ***Assumption 3:*** When two reviewers disagree on a diagnosis, one of the reviewers has the correct diagnosis and the other reviewer has the wrong diagnosis.

The probability of reviewers disagreeing is 0.301 and equals:

(Pw(Reviewer A) • Pc(Reviewer B)) + (Pc(Reviewer A) • Pw(Reviewer B)) = 0.301

Since: Pw(Reviewer A)=Pw(Reviewer B) = Pw; and

Pc(Reviewer A)=Pc(Reviewer B) = Pc,

(Pw•Pc) + (Pw•Pc) = 0.301

2• Pw•Pc = 0.301

Pw•Pc = 0.301/2 = 0.151

Thus, the probability of Reviewer A being wrong and Reviewer B being correct:

Pw • Pc = Pw•(1 - Pw) = 0.151

$-Pw^2 + Pw = 0.151$

$-Pw^2 + Pw - 0.151 = 0$

Giving the standard form of the quadratic equation:

$Pw^2 - Pw + 0.151 = 0$

Using the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$Pw = = \frac{1 \pm \sqrt{1 - 4 \cdot 0.151}}{2} = (0.185, 0.815)$$

If we select from the two possible solutions: Pw = 0.185 then Pc=(1-Pw)=0.815

can be shown to be correct by substituting back into original equation and calculating the total number of reviewers in agreement :

(Pc• Pc) + (Pw• Pw) = 0.698 which equals 0.699 (given slight rounding error observed in data, Table 4).

Therefore, based upon the assumptions listed above, the probability of a reviewer or adjudicator misdiagnosing a slide is:

Pw=0.185, or 18.5%

Likewise, the probability of a reviewer or adjudicator correctly diagnosing a slide is:

Pc = 1 – Pw

Or Pc=0.815, that is 81.5%

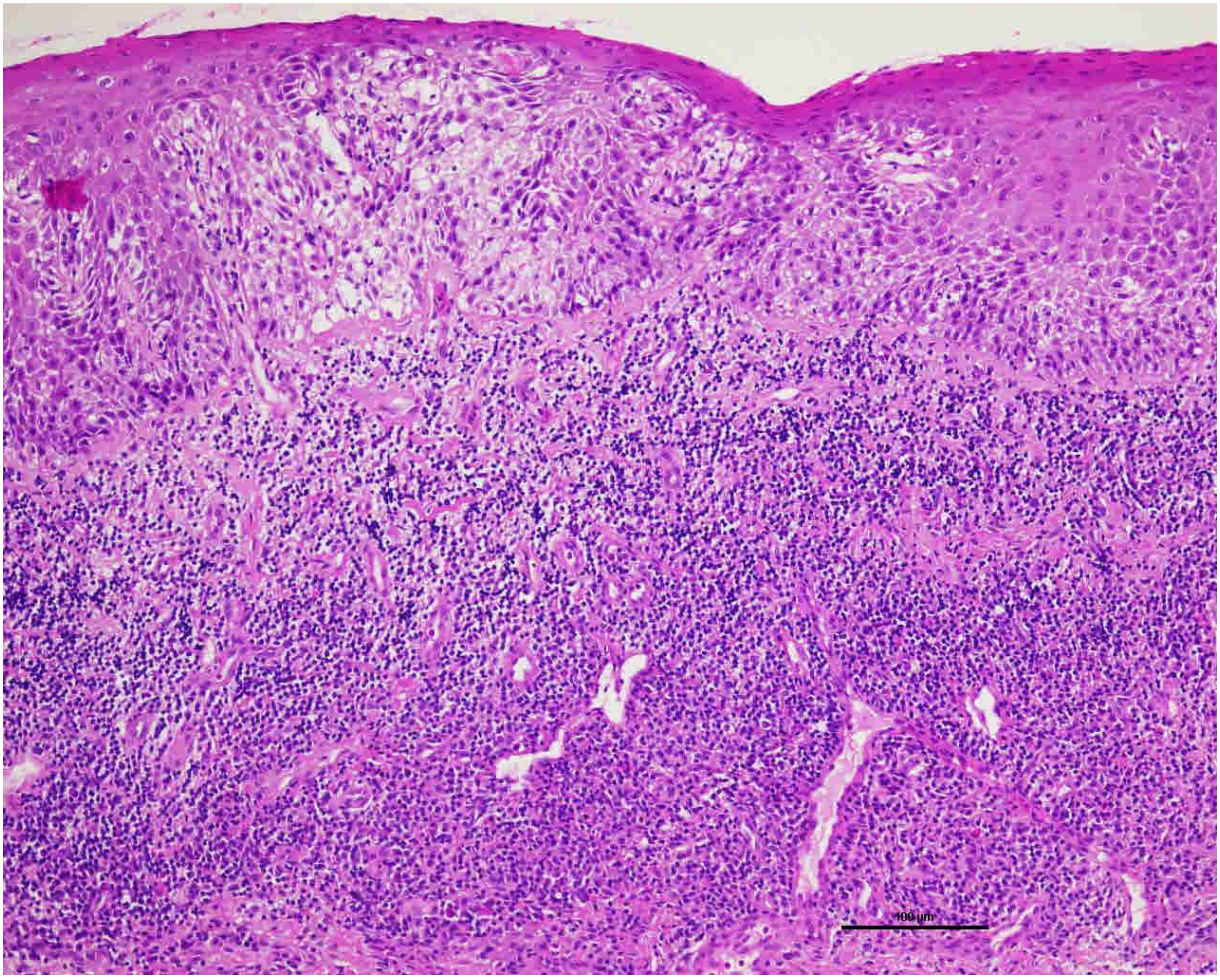
Part C: Overall Probability of Correct Diagnosis Given 2 Reviewers and an Adjudicator to Settle Differences

The 6 probability scenarios in Table 2 are mutually exclusive and thus their probabilities add to 1 (within rounding error). The Pc is the probability of an individual reviewer or adjudicator for a given slide making the correct diagnosis; and, Pw is the probability of a reviewer or adjudicator making a misdiagnosis for a particular slide. Pc was estimated at 0.815 and Pw was estimated at 0.185.

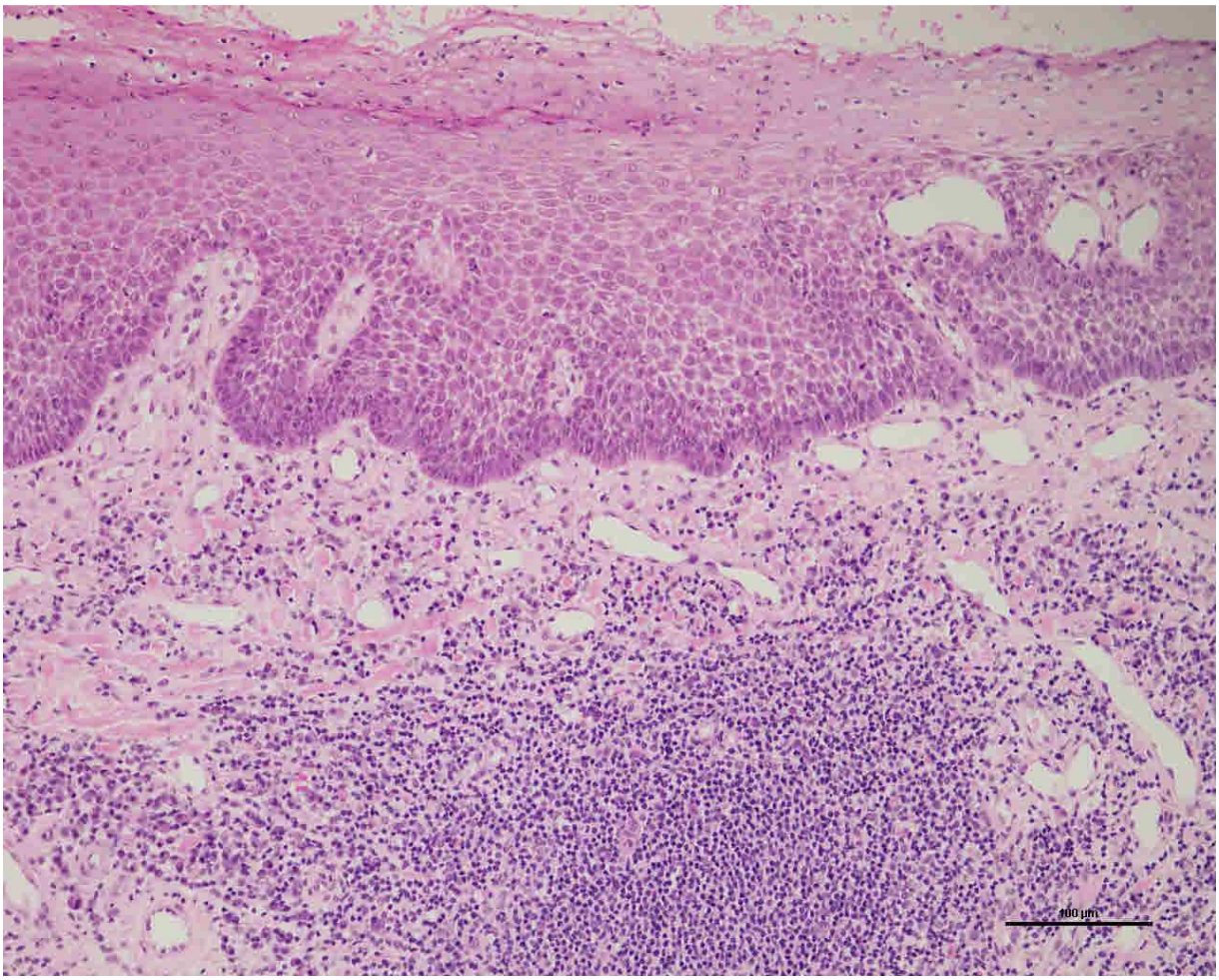
Supplementary Materials: Appendix 2—Representative Pathology Slides

(H&E, 100x)

Benign (consensus review):



Moderate Dysplasia (consensus review):



Malignant (consensus review):