Supplementary appendix

Lorentzen HF, Schmidt SAJ, Sandholdt H, Benfield T. Estimation of the diagnostic accuracy of real-time reverse transcription quantitative polymerase chain reaction for SARS-CoV-2 using re-analysis of published data. Dan Med J 2020;67(9):A04200237

Supplementary methods 1. Further description of latent class analysis

The concept of latent class analysis (LCA) can be illustrated by the following hypothetical example: Imagine that a Martian without knowledge of sexual reproduction observes a group of people on earth. The Martian may notice variation in hair length and stature and may observe an inverse relationship between length of hair and the height of standing individuals. The Martian may speculate that two latent groups/classes could explain the observed characteristics, one with a short stature and long hair (representing females, we could call it latent class "F") and one with a high stature and short hair (representing males; latent class "M"). The Martian would probably find that the two latent groups were of equal sizes (because of the approximately equal sex distribution in the population). Within each latent class, there would no longer be association between observed data (hair, stature) because members of each group would have equally long hair and be of equal stature. This corresponds to local independence in LCA, and a chi-squared test would have p values exceeding 0.1. The Martian might thus find that latent class "F" had high probabilities of long hair and a short stature. In the context of LCA, this latent class can be described by its size, $\pi(F)$, which is approximately 50%, and the conditional probabilities $\pi(\text{long hair} \mid F)$ that would be around say 90% and $\pi(\text{short stature} \mid F)$ that would be around say 95%. Meanwhile, latent class "M" would be characterised by its size $\pi(M)$, which is also approximately 50%, and the conditional probabilities of $\pi(\text{short hair} \mid M)$ and $\pi(\text{long stature} \mid M)$, that would both be high, whereas the conditional probabilities $\pi(\text{long hair} \mid M)$ and $\pi(\text{short stature} \mid M)$ would both be low. A latent model with more latent classes (e.g., short hair–short stature and long hair–long stature) might make it easier for the Martian to place individuals in a group, but would at the same time lead to loss of parsimony and overfitting of the model to the data at the observable level, as it does not contribute to the identification of the prototypical groups of females and males. In LCA, information criteria (e.g., Akaike's information criterion or the Bayesian information criterion) are used to "punish" such overfitting, i.e., to counteract the loss of parsimony and allow identification of the model best describing the data.

Supplementary methods 2. Algorithms used for latent class analysis in lEM software

Unrestricted model:

```
*** INPUT ***

 *Calculation of Se and Sp
 *PCR          CT in table
 *unrestricted LCM

 lat 1
 man 2
 dim 2 2 2
 mod X A|X B|X
 dat [580 308
 21 105]
```

Restricted model:

```
*** INPUT ***

 *Calculation of Se and Sp
 *PCR          CT in table

 lat 1
 man 2
 dim 2 2 2
 mod X
 A|X eq2
 B|X eq2
 des [0 0 -1 0
 0 0 0 0]
 dat [580 308
 21 105]
 sta A|X [.5 .5 .0001 .5]
```