

Combining Rapid Antigen Testing and Syndromic Data Improves Sensitivity and Specificity in Real-World COVID-19 Detection

Fergus J Chadwick^{a,b}, Yacob Haddou^{a,b}, Tasnuva Chowdhury^a, David Pascall^d,
Shayan Chowdhury^e, Jessica Clark^{a,b}, Joanna Andrecka^f, Mikolaj
Kundergorski^{c,b}, Craig Wilkie^{c,b}, Eric Brum^f, Tahmina Shirin^g, A S M
Alamgir^g, Mahbubur Rahman^g, Ahmed Nawsher Alam^g, Farzana Khan^g, Janine
Illian^{c,b}, Ben Swallow^{c,b}, Davina L Hill^{a,b}, Dirk Husmeier^c, Jason
Matthiopoulos^{a,b}, Katie Hampson^{a,b}, Ayesha Sania^h

^a*Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow*

^b*COVID-19 in LMICs Research Group, University of Glasgow*

^c*School of Mathematics and Statistics, University of Glasgow*

^d*MRC Biostatistics Unit, University of Cambridge*

^e*a2i Programme, ICT Ministry/UNDP Bangladesh*

^f*UN FAO in support of the UN Interagency Support Team, Bangladesh*

^g*Institute of Epidemiology, Disease Control and Research, Ministry of Health, Bangladesh*

^h*Division of Developmental Neuroscience, Department of Psychiatry, Columbia University*

1. Introduction

Identification and isolation of COVID-19 cases remains key to the pandemic response across the globe. The faster and more accurately we can identify cases, the more effectively we can provide clinical care, reduce transmission of infection and develop population-level interventions. RT-PCR testing has rapidly become the default, gold-standard test for COVID-19 in applied settings due to its high sensitivity and specificity for COVID-19 [2]. Most of the world's population, however, live in low- and middle-income countries where the laboratory facilities needed to carry out RT-PCR tests are often scarce and hard to reach [4]. COVID-19 diagnosis worldwide, therefore, must be made accessible using inexpensive

*Corresponding Author

Email addresses: f.chadwick.1@research.gla.ac.uk (Fergus J Chadwick), yacob.haddou@glasgow.ac.uk (Yacob Haddou), tasnuvachowdhury2004@gmail.com (Tasnuva Chowdhury), david.pascall@mrc-bsu.cam.ac.uk (David Pascall), shayan.chowdhury@a2i.gov.bd (Shayan Chowdhury), Jessica.Clark@glasgow.ac.uk (Jessica Clark), aandrecka@gmail.com (Joanna Andrecka), mikolaj.kundergorski@gmail.com (Mikolaj Kundergorski), craig.wilkie@glasgow.ac.uk (Craig Wilkie), eric.brum@fao.org (Eric Brum), tahmina.shirin14@gmail.com (Tahmina Shirin), aalamgir@gmail.com (A S M Alamgir), dr_mahbub@yahoo.com (Mahbubur Rahman), anawsher@yahoo.com (Ahmed Nawsher Alam), farzanakhan_25@yahoo.com (Farzana Khan), janine.illian@glasgow.ac.uk (Janine Illian), ben.swallow@glasgow.ac.uk (Ben Swallow), davina.hill@glasgow.ac.uk (Davina L Hill), dirk.husmeier@glasgow.ac.uk (Dirk Husmeier), jason.matthiopoulos@glasgow.ac.uk (Jason Matthiopoulos), katie.hampson@glasgow.ac.uk (Katie Hampson), ays328@mail.harvard.edu (Ayesha Sania)

28 methods that can be carried out locally [6].

29 An increasingly popular alternative to RT-PCR is rapid antigen testing
30 (RAT) [7]. Like RT-PCR, these tests have high specificity for COVID-19 while
31 being less expensive, easier to implement, and faster to produce results [8].
32 RATs also require less commitment and discomfort for patients. For RT-PCR
33 testing, patients must travel to a designated site (such as a hospital or testing
34 booth) or have highly visible PPE-clad officials visit their home. Then, invasive
35 nasopharyngeal swabs must be taken and there is a delay in receiving the result
36 (between one day and a week in Bangladesh). In contrast, RAT can be conducted
37 on nasal or saliva samples, completed in the home and results are available in 30
38 minutes. RATs can be taken by persons with limited training, thus decreasing
39 the time and expense associated with identifying cases. Together, these traits
40 make RATs an appealing alternative to RT-PCR. However, several concerns
41 have been raised about the sensitivity of RAT [9] leading to more false negative
42 diagnoses.

43 Another alternative to RT-PCR, one that has been used since the start of
44 the pandemic, is identifying cases through symptom-thresholding [10]. In this
45 approach, a patient presenting with a fever and one or more viral pneumonia
46 symptoms is treated as a COVID-19 positive patient. The main advantage of
47 this approach is the ease of implementation. For example, in Bangladesh, a lower
48 middle-income country, much of the initial support and reporting of infections
49 locally is provided by community support teams (CSTs) composed of local
50 volunteers with basic training. The CSTs can easily collect symptomatic data
51 in the community and provide care where the criteria are met. However, these
52 symptom-thresholds were developed early in the outbreak, and thus necessarily
53 drawn from clinical intuition, rather than data. Consequently, the relationship
54 between the criteria and the true COVID-19 status is often weak, with low
55 specificity leading to a very large number of false positive diagnoses.

56 A natural extension to these symptom-threshold approaches is syndromic
57 modelling. Here, a patient presenting with a fever and one or more viral
58 pneumonia symptoms is treated as a potential COVID-19 patient. However,
59 rather than using a set of pre-determined criteria, a range of symptomatic and
60 risk factor data are collected and then a sub-sample of patients is tested using
61 RT-PCR for COVID-19 [11]. These data are used to fit a model that allows
62 more accurate prediction of how likely a patient is to have COVID-19 through
63 the identification of COVID-19 syndromes [13]. It is worth highlighting at this
64 point that in resource-limited settings, there is very limited provision for testing
65 of asymptomatic cases, despite their important role in disease transmission [14].
66 Even while focusing solely on symptomatic patients, syndromic modelling is a
67 complex and nuanced task. The strength of relationships between symptoms
68 and diseases is not stable through time or across sampling strategies since the
69 relative importance of each symptom for disease diagnosis, in part, depends on
70 the prevalence of other diseases causing similar symptoms in the community
71 [15]. For example, if another disease for which loss of smell is a symptom
72 becomes common, that symptom becomes a worse predictor for COVID-19.
73 Similarly, if everyone who presents has a cough and thus is included in the

sample, then coughing will likely have a very low correlation with COVID-19 (even if the two are strongly related in the general population). While these issues can be overcome by properly considering the population sampled and using appropriately sophisticated statistical methods, the many types of common respiratory disease generally means that even then these models tend to have relatively high false positive rates (low specificity) for COVID-19 [15], although much lower than the symptom-threshold approach.

Poor sensitivity and specificity are problematic in diagnostics but higher error rates than gold-standard methods may be tolerable depending on their scale and impact given the local situation. Low specificity means a large number of false positive classifications, where the patient is told they have COVID-19 but they actually do not. This might lead to patients unnecessarily self-isolating and receiving support which can be expensive to the individuals and local public health bodies, as well as reducing available resources for those who need them [16]. Similarly, low sensitivity means more false negative classifications, where the patient is told they do not have COVID-19 but they actually do, which can lead to a health-risk for the individual and to the disease spreading further [17]. The costs of these misclassifications will depend on local context. When the prevalence of the disease is low, false positives may create local scepticism about the value of testing, or when there are strong population-level mitigations already in place (such as a nationwide lockdown), then false positives might be more costly than false negatives [16]. If the disease is abundant or increasing rapidly then false negatives are likely to be more costly. In most situations, a balance will need to be struck [4].

The two dominant “alternative” testing methods available in resource limited settings, therefore, are both flawed. Relying solely on symptomatic diagnosis will likely overestimate the number of individuals with COVID-19 due to its lack of specificity. Conversely, RATs will give a false impression of control due to the number of positive cases that will be missed. In this paper, we demonstrate how to combine these data types to exploit their complementarity and ameliorate their respective weaknesses. We aim to compare the performance of these two testing methods and the combined approach both in terms of general prediction and as diagnostics under three epidemiological scenarios; and demonstrate that the combined data achieve equal to much lower error rates than the next best method. We then discuss the role of statistically integrating data from multiple imperfect testing methods in resource limited settings to improve the diagnosis of diseases, particularly COVID-19.

2. Methods

Participants included in this study were identified for COVID-19 testing after self-reporting symptoms to the Bangladesh government’s national hotlines for COVID-19 support. Recruitment took place across Dhaka (the capital city of Bangladesh) between 2nd April 2021 and 5th May 2021.

Patients were selected for further testing conditional on the presence of a fever ($>38^{\circ}\text{C}$) at the point of testing and one or more of 14 additional symptoms

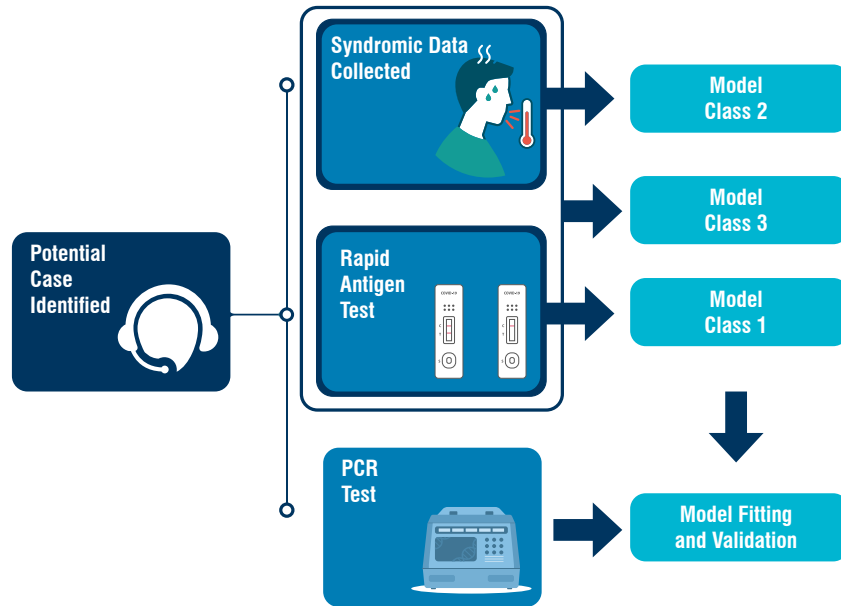


Figure 1: Schematic description of identification of likely COVID-19 patients by community support teams (CSTs), swab collection and model definitions. The teams collected syndromic data (age, gender and presence/absence of 14 predetermined symptoms), and two sets of naso-pharyngeal swabs (one each for Rapid Antigen Testing and RT-PCR). We then used rapid antigen testing (RAT) and syndromic data, two imperfect but inexpensive diagnostics, to generate three model classes: RAT result only in Model Class 1, syndromic data only in Model Class 2, and both RAT result and syndromic data in Model Class 3. The RT-PCR test result is used to train and test each model using temporal cross validation.

118 associated with COVID-19 (breathing problems, coughing, diarrhoea, fever
 119 (ongoing), a headache, loss of taste, loss of smell, muscle pain, red eyes, a runny
 120 nose, a sore throat, tiredness, vomiting or a wet cough). The patient's age
 121 and gender were also recorded, but these data were not included in the patient
 122 selection criteria.

123 Nasal swabs and syndromic data were collected from the patient by medical
 124 technologists. One swab each was used for rapid antigen testing (RAT) and
 125 RT-PCR (gold-standard for COVID-19 status). The full questionnaire and
 126 testing protocols are provided in Appendix XX. Participants provided written
 127 informed consent to sample collection and for their test results to be analyzed in
 128 the study.

129 We examined the ability of the two imperfect identification methods, the
 130 syndromic profile and RAT result, to predict the patient's COVID-19 status
 131 when used separately and together. The different data combinations define three
 132 model classes (Figure 1).

Find more
ways to cite
Figure
refig:data-
flowchart in
this text.

133 Model Class 1 uses only the RAT result and is the simplest of the three.
 134 It simply equates a positive RAT result with the patient being PCR positive,
 135 and a negative RAT result with PCR negativity. Model Class 2 uses only the
 136 syndromic data and Model Class 3 combines the RAT result with the syndromic
 137 data.

138 For Model Class 2, we used a Bayesian multivariate probit model [18]. The
 139 multivariate probit structure allows the model to account for the correlations
 140 between, and binary nature of, the symptoms (e.g. loss of taste is often correlated
 141 with loss of smell). By using a Bayesian formulation, we are able to better quantify
 142 the uncertainty in the parameter estimates. Structurally, the multivariate probit
 143 model allows the symptoms and COVID-19 status to be treated as correlated
 144 binary outcomes with an intrinsic rate (the intercept for each variable) and the
 145 patient’s age and gender, while propagating and quantifying uncertainty.

146 In Model Class 3, we model RAT positive patients as PCR positive and
 147 use the syndromic approach outlined for Model Class 2 for the RAT negative
 148 patients. The models were fitted to the data using Hamiltonian Monte Carlo in
 149 the Stan programming language [19].

150 We conducted backwards model selection (starting with the most complex
 151 model feasible, with all 14 symptoms and both covariates) to identify a subset of
 152 models with the highest predictive power under temporal cross validation (Figure
 153 2). Reducing the number of possible models to a small number of the most
 154 predictive models was necessary to reduce computational demand and reduce the
 155 risk of overfitting models to the test scenarios. The large number of symptoms
 156 means that there is a high number of potential model configurations (>131 000
 157 for 14 symptoms and two covariates) which might, by chance, perform well on
 158 the test sets (even under the challenging conditions of temporal cross validation)
 159 but lack transferability. By first using general predictive power to narrow down
 160 the number of candidate models and then testing those models under more
 161 specific scenarios, we are more likely to choose models which generalise well
 162 to new data. The number of candidate models used was not pre-determined.
 163 In fitting the models it became clear that there were “jumps” in performance
 164 (as defined below) between models containing five and four symptoms, so the
 165 models with zero to four symptoms were used as the candidate models.

166 We scored the models’ predictive power using cross entropy. Cross entropy
 167 measures the accuracy of probabilistic predictions for models that predict binary
 168 outcomes using probabilities [20], similar in concept to a mean squared error.
 169 A cross entropy value close to zero corresponds to high levels of accuracy, with
 170 larger values indicating lower accuracy. As the score only uses the predicted
 171 probability and true values, it is possible to directly compare the predictions
 172 of any model for the same test set. More details on the model structure and
 173 selection process, including code, are available in Appendix XX.

174 We then compared models as classifiers using their false positive and false
 175 negative rates in three epidemiological scenarios. In applied settings, models
 176 must often be evaluated on their performance as classifiers rather than just as
 177 prediction engines (i.e. their ability to say a patient is COVID-19 positive or
 178 negative, not simply the probability the patient might be COVID-19 positive or

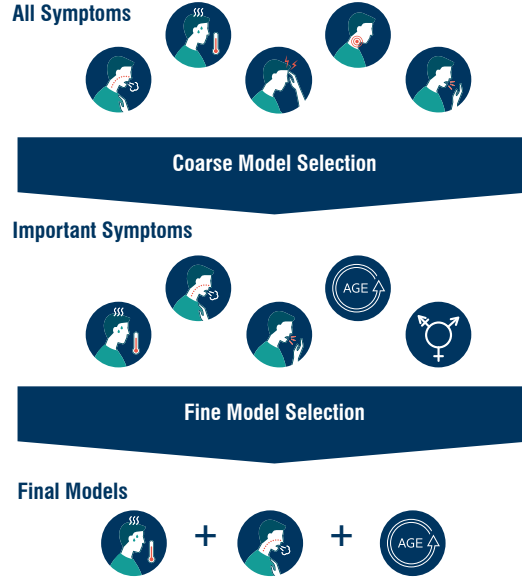


Figure 2: Schematic for rounds of model selection in the multivariate probit component of Model Classes 2 and 3. With 14 symptoms (only 5 shown here for demonstration purposes) and two covariates there are over 131000 possible model combinations. To make exploring these possible models computationally feasible and, to reduce the risk of overfitting, we carried out two rounds of model selection. First, the data are divided into temporal cross validation sets. The multivariate probit connects symptoms to the RT-PCR result through a correlation matrix. In the coarse model selection, the most complex feasible model (all symptoms and covariates) is fit to the training data. The estimated correlations between each symptom and the RT-PCR result are compared for each cross validation set. The symptoms that have non-zero correlations in a systematic direction (i.e. all positively or all negatively correlated with RT-PCR result) are retained. The process is then repeated on each retained set of symptoms until the four symptoms in each model class with the strongest correlation to RT-PCR result. We then conduct a more exhaustive fine model selection on all the possible permutations of the four symptoms and two covariates. In this round, each model is fit to training data and used to predict for the test set, and the quality of those predictions is measured using cross entropy scoring. The cross entropy score is then used to select the best predictive model for each level of model complexity. Only these final models are then used for classification. This reduces the set of models tested as classifiers from >131 000 to just four per model class.

Table 1: For each epidemiological scenario there is a requirement and a performance criterion. The requirement refers to a base level of performance the model must achieve; in general this will be a maximum acceptable error rate of some kind. The requirement determines a threshold for each model which most closely meets that requirement. The performance criterion is then used to determine which model performs the 'best' given that the requirement has been met.

Scenario Name	Requirement	Performance Criterion (Error)
1 Agnostic	Maximise correct classification rates	Sum of error rates
2 Rising Cases	Max. 20% false negative rate	False negative rate
3 Steady, Low-Level Cases	Max. 20% false positive rate	False positive rate

negative). To generate a classification, a probability threshold must be chosen over which patients are classified as COVID-19 positive.

Classifier performance was compared using ROC curves and error rates under three epidemiological scenarios. ROC curves show the true and false positive rates that each model can achieve. To extract the error rate under the epidemiological scenarios (described below), we use the ROC calculations to identify the probability threshold which most closely meets the scenario requirement (see Table 1. Comparing specific scenarios allows classifier performance to be demonstrated in relevant scenarios. Whether measuring classifier performance in specific scenarios or more generally, decisions need to be made about the relative cost and acceptable levels of the two types of misclassification (false positives and negatives). We strongly emphasise that local context should be the guide in applying these methods.

In Scenario 1, we do not consider epidemiological context but simply weight false negative and false positive rates equally by aiming to maximise the overall correct classification rate. Scenario 2 corresponds to the current situation in Bangladesh at time of writing (July 2021), with COVID-19 cases beginning to rapidly increase again. Under these circumstances, false negatives are extremely costly relative to false positives due to the exponential growth of the disease. In Scenario 3, the pandemic is not declining but maintaining a steady rate of cases. In this situation, policy-makers may be keen to keep false positive diagnoses low to prevent lockdown fatigue and to keep the workforce active.

3. Results

A total of 1172 subjects had data available for the current analyses. The mean age of women participants (46.8430034% of the sample) was 37.13 (SD = 14.11), and for men (53% of the sample) was 35.94 (SD = 14.28). Participants were self-selecting and drawn from across Dhaka.

Model selection for Model Class 2 (syndromic data only) and 3 (syndromic and RAT data), each retained age as an explanatory variable and showed a

marked decline in predictive power at more than 4 symptoms. The final four symptoms in order of importance (i.e. the most important symptom was retained in all of the final 4 models, the least important symptom was only retained in the 4 symptom model) were wet cough, runny nose, loss of smell and breathing problems for Model Class 2, and fever, wet cough, tiredness and diarrhoea for Model Class 3. For both Model Class 2 and Model Class 3 model selection retained age as a covariate but not gender.

In the comparison of model predictive performance, Model Class 1 (RAT only) performed worst with a cross entropy of 3.24 (cross entropy values further from zero correspond to worse predictive performance). The median cross entropy values were between 2.52 and 2.59 for models in Class 2 (syndromic data only). Models in Class 3 (combined data model) performed best with cross entropy values between 1.44 and 1.45 (see Figure 3).

General model classification performance is shown by the full ROC curves for each model (Figure 4).

Scenario specific classification performance is shown in Figure 5. In Scenario 1 (“Agnostic,” see Table 1), the median error was 0.34 for models in Class 1 and Class 3 and between 0.98 and 1 for models in Class 2 (Figure 5). In Scenario 2 (“Rising Cases”), Model Class 1 was unable to meet the required false negative rate. The median errors were between 0.82 and 0.94 for models in Class 2, and 0.48 and 0.78 for models in Class 3 (Figure 5). In Scenario 3 (“Steady, Low-Level Cases”), the error in Class 1 was 0.02, 0.02 and the median errors ranged from 0.14 to 0.19 for Class 2, and 0.08 to 0.19 for Class 3 (Figure 5).

4. Discussion

We have demonstrated that combining two imperfect diagnostics yields better prediction of COVID-19 status and greater flexibility than each diagnostic individually. These improvements are non-trivial in real-world settings like Bangladesh, where there are currently thousands of new cases being identified every day and the pandemic growth is accelerating meaning every missed case has a compounding effect. In the most relevant scenario, 2 (“Rising Cases”), where we try to keep the false negative rate low, the combined data model achieves 22 percentage points lower than the false negative rate of the RAT only model, and 34 percentage points lower false positive rate than the syndromic only model. These are large performance gains for any diagnostic, although under the current situation will have a considerable impact on identifying both COVID-19 positive and negative patients. The pattern is similar in Scenario 3 and when comparing the more generic measures of model performance in prediction and classification (Figures 3 and 5, respectively). Across almost all scenarios and metrics, the combined data model has equal or better performance. Furthermore, this boost is achieved with data that are already being collected in Bangladesh and other low- and middle- income countries. Outwith developing and rerunning the models presented in this paper, these improvements are essentially cost free and eminently scalable.

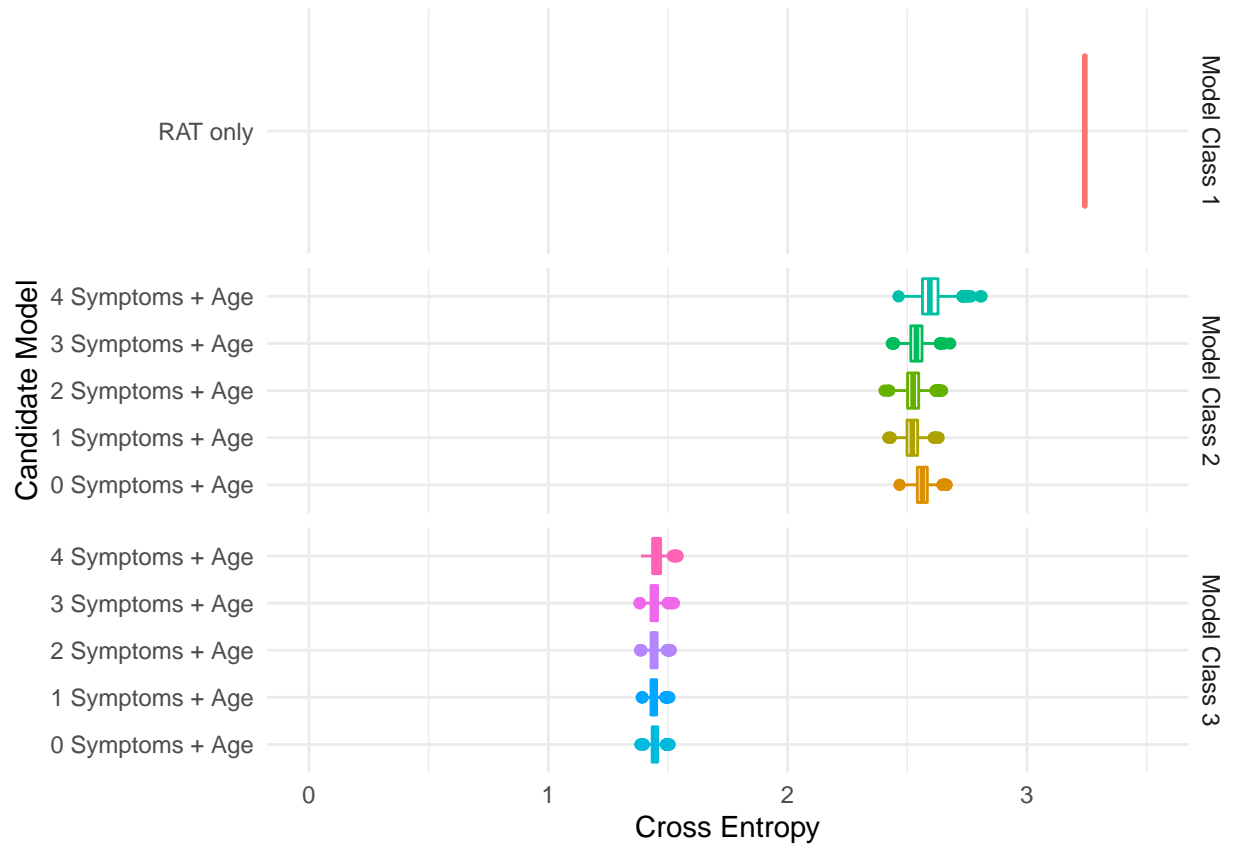


Figure 3: Predictive performance of candidate models. Interquartile ranges for the posterior cross entropy of the best candidate models at each level of model complexity tested under temporal cross-validation. Cross entropy is a measure of distance from the truth, so values closer to zero indicate better models. The intermediate complexity models perform best at prediction, although performance is similar across all the models within each model class (1: rapid antigen testing (RAT) only; 2: syndromic data only; and 3: combined RAT and syndromic data).

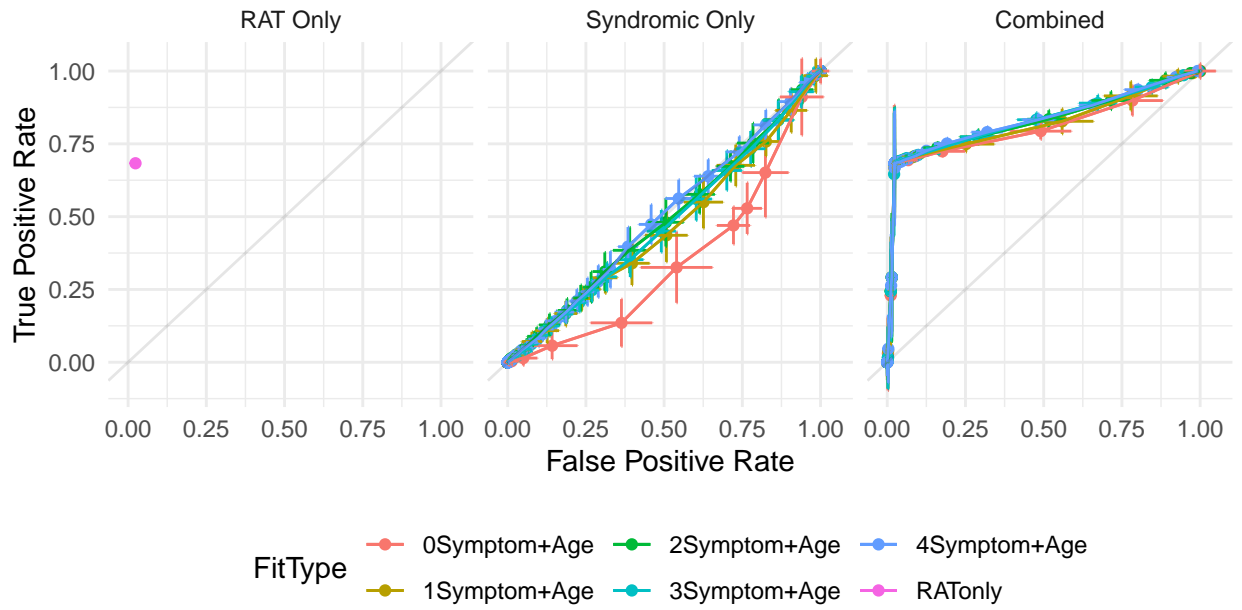


Figure 4: Receiver operating characteristics for rapid antigen testing (RAT) only approach (Model Class 1) and posterior mean (\pm posterior standard deviation) receiver operating characteristics for Class 2 (syndromic data only) and 3 (syndromic and RAT data) models. These curves demonstrate the performance of the model for any hypothetical scenario as defined by the axes (as opposed to Figure 5 which demonstrates model performance in specific epidemiological scenarios which are realisations of a single point in this space).

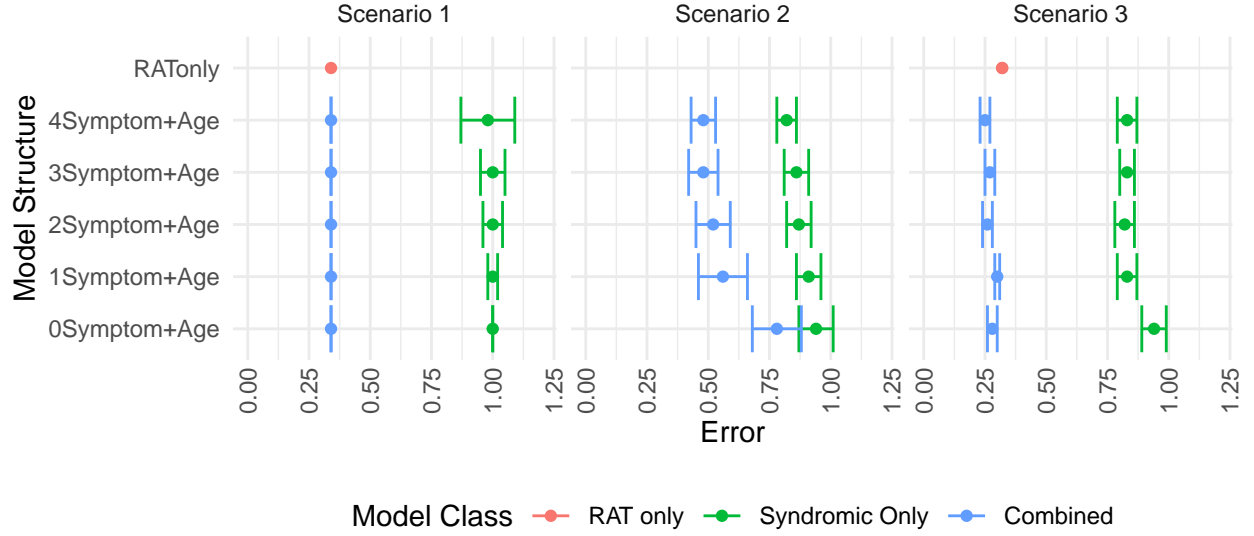


Figure 5: Performance of models under each scenario measured by errors defined in Table 2. Low errors correspond to better model performance. There is no error rate defined for the Model Class 1 (RAT only model) in Scenario 2 as the model failed to meet the requirement for that scenario (making the error functionally infinite).

251 Syndromic identification and RATs are fast, inexpensive and can be performed
 252 at patients' homes by minimally trained personnel. These imperfect detection
 253 methods have been developed as inexpensive alternatives to RT-PCR. While even
 254 the improvements described above will never allow these methods to compete
 255 with RT-PCR in terms of sensitivity and specificity, Rapid Antigen Testing and
 256 syndromic diagnosis also hold several further advantages. Unlike RT-PCR where
 257 patients have to go to designated testing centres, samples need to be taken by
 258 trained technologists and results take a day to a week to come back, both of the
 259 imperfect diagnostics can be delivered in the community with almost instant
 260 result delivery. This has several advantages. Firstly, increased accessibility by
 261 removing the need to travel for testing, thus reduces bias particularly against
 262 poorer, sicker and older people and reduces the risk of transmission in transit.
 263 Secondly, linked to cost, it is much easier to scale up testing when specialist
 264 training and expensive equipment and biosafety procedures are not required.
 265 Thirdly, it allows for assessment of an individual's wellbeing in the home context,
 266 and thus facilitates tailoring interventions to where they are most needed.

267 The symptoms retained in the models may hint at the mechanism by which
 268 combining syndromic data and RAT results improves diagnosis. We have de-
 269 liberately not emphasised the final symptoms chosen through model selection
 270 in this paper as we are focusing on prediction and classification for a unique
 271 sub-population: self-referring, symptomatic patients. We do, however, highlight
 272 that the symptoms retained in the final models for the syndromic-only and

273 combined models are largely different (only one symptom appears in both lists).
 274 We propose that this is due to the fact that RAT is most effective during the
 275 first week of symptom onset, with much worse performance before and after this
 276 period [21]. These first-week symptoms are generally typical of viral pneumonias
 277 [22], and, indeed, when the RAT result is excluded from the model, the most
 278 important symptoms are typical of upper respiratory tract infections. However,
 279 when the RAT result is combined with syndromic data, the most important
 280 symptoms become much more eclectic. It is probable that these symptoms are
 281 either typical of later-stage COVID-19 or are less common presentations of the
 282 disease, possibly caused by co-infection or multimorbidities. Further research is
 283 needed to understand the mechanisms by which symptoms predict COVID-19
 284 and by which RAT misses COVID-19. Of particular interest is whether indi-
 285 viduals that are missed by RAT are less infectious, which could be explored
 286 by using Threshold Cycle (Ct) values from the RT-PCR to compare viral load
 287 with respect to prediction by the different methods [23]. We note also that, as
 288 expected, age was retained in model selection. We were, however, surprised that
 289 gender was removed during model selection. Gender is thought to play a major
 290 role in infection risk [25]. As we are looking to predict symptomatic COVID-19
 291 in symptomatic individuals, generalised risk of infection is perhaps less predictive
 292 than expected, potentially due to the balancing of risk and burden [26].

293 We believe that the combined syndromic and rapid testing model represents
 294 the most promising approach to testing for COVID-19 in low- and middle-
 295 income countries at present. By taking a statistical modelling approach to
 296 case identification, we are able to update our diagnostic process in real time,
 297 allowing this method to readily adapt to new variants (or even new diseases) or
 298 new priorities for resource allocation. The modelling frameworks we have used
 299 are also sufficiently flexible to accommodate new data sources. Of particular
 300 interest are extensions to include the “pandemic context” in the model using
 301 space-time data. Furthermore, by using more sophisticated modelling structures
 302 it is possible to tune error rates to better reflect the local relative costs of false
 303 positives and false negatives. Naturally, these strengths have complementary
 304 limitations. Our models require updating in real-time and can only achieve good
 305 performance if the validation data are of high quality. Similarly, targeting error
 306 rates is only sensible if those rates properly reflect local conditions which is
 307 hard to do in practice. These limitations should be seriously considered but the
 308 alternatives for imperfect testing methods are diagnostics that cannot be tailored
 309 to local conditions at all (and, as such may perform worse than a method which
 310 is sub-optimally tailored to local conditions) or diagnostics which make these
 311 decisions implicitly and not explicitly. We believe that in choosing to be explicit
 312 these decisions are more readily challenged, researched and improved upon. We
 313 also emphasise the need for rigorous experimental design to ensure findings from
 314 the sample population are applicable to the target population and the need for
 315 further research into understanding error rate tradeoffs in applied settings.

316 The methodology we have outlined here is applicable to a wide range of
 317 diseases and settings across low- and middle-income countries. One of the
 318 biggest challenges in diagnosing and tracking many diseases in resource-limited

319 settings is the low availability of access to gold-standard testing (such as RT-PCR
 320 in the case of COVID-19) and high error rate of alternative testing methods.
 321 In this paper, we have outlined the process for coupling a small number of
 322 gold-standard tests with formal statistical integration of alternative testing
 323 methods, to generate high quality diagnostic models. This process readily maps
 324 onto many other case identification problems, including the diagnosis of several
 325 neglected tropical diseases. For example, malaria (gold standard (GS) is also
 326 RT-PCR, imperfect methods (IM) include antigen tests, syndromic diagnosis
 327 and blood smears), schistosomiasis (GS: RT-PCR or autopsy; IM: Kato Katz
 328 egg counts, antibody detection) and rabies (GS: fluorescent antibody test; IM:
 329 light microscopy, differential diagnosis).

330 In conclusion, we believe that the combined syndromic and rapid antigen
 331 testing approach represents the most promising approach to large-scale testing
 332 in low- and middle- income countries at present. By using the small amount
 333 of RT-PCR testing possible and formally integrating multiple imperfect, non-
 334 gold-standard methods, we can tune these diagnostics to our local conditions.
 335 We have demonstrated that these improvements can be impressive in real-world
 336 scenarios, and will have a large impact when scaled to the population sizes in
 337 low- and middle-income countries. As such, these low-cost improvements to
 338 existing testing programs have the potential to identify one to two orders of
 339 magnitude more cases than either gold-standard or alternative methods alone.

340 5. Funding

341 The Bill and Melinda Gates Foundation funded work by FAO (INV-022851),
 342 and University of Glasgow reports funding from Wellcome (207569/Z/17/Z).
 343 The authors declare no competing interests.

344 6. Acknowledgements

345 CSTs + BAHCM/MathsAndStats + Matthiopoulos Research Group + Dan
 346 Haydon

Fergus to
look up
names

347 References

- 348 [1] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al.
 349 Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.
 Eurosurveillance 2020;25:2000045.
- 350 [2] Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection:
 351 Issues affecting the results. Expert Review of Molecular Diagnostics
 2020;20:453–4.
- 352 [3] Chowdhury R, Luhar S, Khan N, Choudhury SR, Matin I, Franco OH.
 Long-term strategies to control COVID-19 in low and middle-income
 353 countries: An options overview of community-based, non-pharmacological
 interventions. European Journal of Epidemiology 2020;35:743–8.

- 354 [4] Vandenberg O, Martiny D, Rochas O, Belkum A van, Kozlakidis Z. Con-
355 siderations for diagnostic COVID-19 tests. *Nature Reviews Microbiology*
2021;19:171–83.
- 356 [5] Cash R, Patel V. Has COVID-19 subverted global health? *The Lancet*
357 2020;395:1687–8.
- 358 [6] Olalekan A, Iwalokun B, Akinloye OM, Popoola O, Samuel TA, Akinloye
O. COVID-19 rapid diagnostic test could contain transmission in low-
and middle-income countries. *African Journal of Laboratory Medicine*
359 2020;9:1–8.
- 360 [7] Linares M, Pérez-Tanoira R, Carrero A, Romanyk J, Pérez-García F,
Gómez-Herruz P, et al. Panbio antigen rapid test is reliable to diagnose
SARS-CoV-2 infection in the first 7 days after the onset of symptoms.
361 *Journal of Clinical Virology* 2020;133:104659.
- 362 [8] Boum Y, Fai KN, Nikolay B, Mboringong AB, Bebell LM, Ndifon M,
et al. Performance and operational feasibility of antigen and antibody
rapid diagnostic tests for COVID-19 in symptomatic and asymptomatic
patients in cameroon: A clinical, prospective, diagnostic accuracy study.
363 *The Lancet Infectious Diseases* 2021.
- 364 [9] Mak GC, Cheng PK, Lau SS, Wong KK, Lau C, Lam ET, et al. Evaluation
of rapid antigen test for detection of SARS-CoV-2 virus. *Journal of*
365 *Clinical Virology* 2020;129:104500.
- 366 [10] Jin Y-H, Cai L, Cheng Z-S, Cheng H, Deng T, Fan Y-P, et al. A rapid
advice guideline for the diagnosis and treatment of 2019 novel coronavirus
(2019-nCoV) infected pneumonia (standard version). *Military Medical*
367 *Research* 2020;7:1–23.
- 368 [11] Sim JXY, Conceicao EP, Wee LE, Aung MK, Seow SYW, Teo RCY, et
al. Utilizing the electronic health records to create a syndromic staff
surveillance system during the COVID-19 outbreak. *American Journal of*
369 *Infection Control* 2021;49:685–9.
- 370 [12] Undurraga EA, Chowell G, Mizumoto K. COVID-19 case fatality risk
by age and gender in a high testing setting in latin america: Chile,
371 march–august 2020. *Infectious Diseases of Poverty* 2021;10:1–1.
- 372 [13] Wenham C, Smith J, Morgan R. COVID-19: The gendered impacts of
the outbreak. *The Lancet* 2020;395:846–8.
- 373 [14] Mayorga L, Samartino CG, Flores G, Masuelli S, Sánchez MV, Mayorga
LS, et al. A modelling study highlights the power of detecting and
isolating asymptomatic or very mildly affected individuals for COVID-19
375 epidemic management. *BMC Public Health* 2020;20:1–1.
- 376 [15] Garry S, Abdelmagid N, Baxter L, Roberts N, Waroux O le P de, Ismail
S, et al. Considerations for planning COVID-19 treatment services in
humanitarian responses. *Conflict and Health* 2020;14:1–1.

- 378 [16] Surkova E, Nikolayevskyy V, Drobniewski F. False-positive COVID-19
379 results: Hidden problems and costs. *The Lancet Respiratory Medicine*
2020;8:1167–8.
- 380 [17] West CP, Montori VM, Sampathkumar P. COVID-19 testing: The threat
381 of false-negative results. *Mayo clinic proceedings*, vol. 95, Elsevier; 2020,
382 p. 1127–9.
- [18] Albert JH, Chib S. Bayesian analysis of binary and polychotomous re-
383 sponse data. *Journal of the American Statistical Association* 1993;88:669–
384 79.
- [19] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M,
385 et al. Stan: A probabilistic programming language. *Journal of Statistical*
386 *Software* 2017;76:1–32.
- [20] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and
387 estimation. *Journal of the American Statistical Association* 2007;102:359–
388 78.
- [21] Muhi S, Tayler N, Hoang T, Ballard SA, Graham M, Rojek A, et al. Multi-
site assessment of rapid, point-of-care antigen testing for the diagnosis
389 of SARS-CoV-2 infection in a low-prevalence setting: A validation and
390 implementation study. *The Lancet Regional Health-Western Pacific*
2021;9:100115.
- [22] Xu T, Chen C, Zhu Z, Cui M, Chen C, Dai H, et al. Clinical features
391 and dynamics of viral load in imported and non-imported patients with
392 COVID-19. *International Journal of Infectious Diseases* 2020;94:68–71.
- [23] Albert E, Torres I, Bueno F, Huntley D, Molla E, Fernández-Fuentes MÁ,
393 et al. Field evaluation of a rapid antigen test (panbio™ COVID-19 ag
394 rapid test device) for COVID-19 diagnosis in primary healthcare centres.
395 *Clinical Microbiology and Infection* 2021;27:472–e7.
- [24] Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA,
396 et al. Real-time tracking of self-reported symptoms to predict potential
397 COVID-19. *Nature Medicine* 2020;26:1037–40.
- [25] Wolff D, Nee S, Hickey NS, Marschollek M. Risk factors for covid-19
398 severity and fatality: A structured literature review. *Infection* 2020:1–4.
- [26] Joe W, Kumar A, Rajpal S, Mishra U, Subramanian S. Equal risk, unequal
399 burden? Gender differentials in COVID-19 mortality in india. *Journal of*
Global Health Science 2020;2.