

Appendix 5 (as supplied by the authors): Multiple imputations of missing data

Patterns of missing values for clinical predictors (n=676)

Complete case analysis would have led to excluding 129/676 patients (19%).

Clinical signs or symptoms	Missing values	Percentage
Sudden onset	7	1.04
Sore throat/pain in throat	32	4.73
Body temperature	3	0.44
Headache	14	2.07
Abdominal pain	9	1.33
Abnormal pharynx		
Erythema/redness	4	0.59
Petechiae (pharynx/palate)/doughnut lesions	42	6.21
Tonsillar swelling	18	2.66
Exudate (pharynx/tonsils)	10	1.48
Cervical lymph nodes		
Swelling/enlargement	4	0.59
Tenderness	28	4.14
Scarlatiniform rash	10	1.48
Conjunctivitis	5	0.74
Rhinitis/coryza/rhinorrhea	3	0.44
Cough	12	1.78
Diarrhea	13	1.92

Characteristics of patients with at least one missing data and complete cases

Clinical signs or symptoms	Patients with at least one missing value (n=129)	Complete cases (n=547)	<i>p</i> ^a
Positive throat culture for GAS	63/129 (48.8)	217/547 (39.7)	0.057
Age, years (SD) ^b	5.7 (2.4)	6.2 (2.5)	0.055
Sudden onset	97/122 (79.5)	440/547 (80.4)	0.815
Sore throat/pain in throat ^c	101/121 (83.5)	441/523 (84.3)	0.817
Body temperature > 38°C	105/129 (81.4)	414/547 (75.7)	0.167
Headache	24/115 (20.9)	162/547 (29.6)	0.058
Abdominal pain	48/120 (40.0)	176/547 (32.2)	0.100
Abnormal pharynx			
Erythema/redness ^d	110/126 (87.3)	489/546 (89.6)	0.463
Petechiae (pharynx/palate)/doughnut lesions	15/87 (17.2)	101/547 (18.5)	0.784
Tonsillar swelling	86/111 (77.5)	391/547 (71.5)	0.197
Exudate (pharynx/tonsils) ^d	28/125 (22.4)	116/541 (21.4)	0.815
Cervical lymph nodes			
Swelling/enlargement	86/125 (68.8)	302/547 (55.2)	0.006
Tenderness	22/101 (21.8)	109/547 (19.9)	0.670
Scarlatiniform rash	17/119 (14.3)	52/547 (9.5)	0.121
Conjunctivitis	3/124 (2.4)	16/547 (2.9)	0.759
Rhinitis/coryza/rhinorrhea	48/126 (38.1)	204/547 (37.3)	0.867
Cough	48/117 (41.0)	214/547 (39.1)	0.702
Diarrhea	10/116 (8.6)	23/547 (4.2)	0.047

Data are no./N of patients (%). ^aChi-square or Fisher's exact test; ^bStudent's *t* test; ^cMissing values in the complete case analysis because this criterion was scored only for the Breese score, which accepts missing values; ^dMissing values in the complete case analysis because these variables are always scored in combination with other variables using "and/or" operators, which is compatible with missing values.

Multiple Imputation of missing data

We used multiple imputations with chained equations (MICE) because of missing values in almost all clinical variables necessary for the validation of the CPRs we identified [1]. MICE was performed in STATA 13/SE ($m=10$). Analyses were repeated in each imputed dataset and estimates of interest were combined by using Rubin's rules [2].

Convergence and imputation diagnostics

Convergence of the MICE algorithm was checked by graphically investigating the trends in the means and standard deviations of the imputed values over 100 iterations. The trace plots did not show apparent trends in the summaries of the imputed values, so the number of burn-in iterations was set to 10. The distributions of imputed and observed values were compared graphically. Distributional plots did not show significant departure between imputed values and observed values. We concluded that the fit of the imputation model was good.

Comparison of complete case analysis versus multiple imputations analysis

We compared the results from analyses restricted to complete cases with results from analyses with multiple imputations. Overall, complete case analysis seemed to produce slightly higher estimates for both sensitivity and specificity.

	Sensitivity, % (95% CI)		Specificity, % (95% CI)	
	CC ^a	MICE ^b	CC	MICE
Breese	89.9 (85.1–93.5)	87.9 (84.1–91.8)	82.4 (77.9–86.4)	82.0 (78.2–85.8)
Edmond	68.7 (62.0–74.8)	66.1 (60.5–71.7)	86.1 (81.8–89.6)	85.9 (82.4–89.3)
McIsaac	94.9 (91.1–97.4)	94.3 (91.6–97.0)	53.9 (48.4–59.4)	53.6 (48.7–58.6)
Wald	94.9 (91.1–97.4)	93.9 (91.1–96.7)	63.9 (58.5–69.1)	63.0 (58.2–67.7)
Attia	87.1 (81.9–91.3)	87.1 (83.1–91.1)	87.9 (83.9–91.2)	87.8 (84.5–91.0)
Joachim	91.8 (87.9–94.7)	87.5 (83.5–91.6)	35.1 (30.4–40.0)	40.1 (35.2–44.9)

^aCC, complete case analysis (n=547); ^bMICE, multiple imputations with chained equations analysis (n=676).

References

1. White IR, Royston P, Wood AM: **Multiple imputation using chained equations: Issues and guidance for practice.** *Stat Med* 2011, **30**:377-399.
2. Rubin D: *Multiple Imputation for Nonresponse in Surveys.* Hoboken, NJ: J Wiley; 1987.