Use of SARS-CoV-2 antigen-detection rapid diagnostic tests for COVID-19 self-testing

INTERIM GUIDANCE

9 March 2022

Web Annex A. GRADE table: Should COVID-19 self-testing, using SARS-CoV-2 Ag-RDTs, be offered as an additional approach?

Lukas E. Brümmer, Stephan Katzenschlager, Hannah Tolle, Seda Yerlikaya, Christian Erdmann, Ioana D. Olaru, Maurizio Grilli, Nira R. Pollock, Claudia M. Denkinger

WHO reference number: WHO/2019-nCoV/Ag-RDTs/Self_testing/Web_annex_A/2022.1

The named authors alone are responsible for the views expressed in this publication.

This publication forms part of the WHO interim guidance: Use of SARS-CoV-2 antigen rapid diagnostic tests for COVID-19 self-testing: Interim guidance, 9 March 2022. It is being made publicly available for transparency purposes and information, in accordance with the *WHO handbook for guideline development*, 2nd edition (2014).

GRADE table: Should COVID-19 self-testing, using Ag-RDTs, be offered as an additional approach?

The following annex summarizes the certainty of evidence according to the GRADE approach. All outcomes in the GRADE table are presented in the order of criticalness determined by the guideline development group (GDG). Figure 1 illustrates the full rankings of each outcome.

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Accuracy – sensitivity (Ag-RDT self-testing vs. rRT-PCR)									
2[a]	observational studies	not serious[b]	not serious[c]	not serious[d]	serious[e]	none	Normalized to a study population with 1,000 participants, 72 true positive and 52 false negative self-testing results were reported. Across the included data sets, sensitivity ranged from 48.9% to 82.5%.	⨁⨁⨁◯ Moderate	CRITICAL
Accuracy – specificity (Ag-RDT self-testing vs. rRT-PCR)									
2[a]	observational studies	not serious[b]	not serious[c]	not serious[d]	not serious[e]	none	Normalized to a study population with 1,000 participants, 874 true negative and 2 false positive self-testing results were reported. Specificity was high across the included data sets with a range of 99.7% to 100%	⨁⨁⨁⨁ High	CRITICAL
Accuracy - concordance (Ag-RDT self-testing vs. Ag-RDT performed by professionals)									
1[f]	observational studies	not serious[b]	serious[c]	not serious[d]	serious[e]	none	Kappa: 0.98 (out of 1.00); PPA: 91.4% (95% CI 77.6 to 97.0) NPA: 99.1% (95% CI 95.0 to 100).	⨁⨁◯◯ Low	CRITICAL

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Individual health outcome - Linkage for positive tests									
12[g]	observational studies	serious[h]	not serious[i]	not serious[d]	not serious[j]	none	In 11 data sets (91.7% of all data sets that reported this outcome), persons were required to quarantine/isolate following a positive test result. In five of these, this was accompanied by the requirement to do an rRT-PCR test for confirmatory testing. In one further data set (8.3%), an invitation to conduct an rRT-PCR test was the only measurement reported.	⊕◯◯◯ Very low	CRITICAL
Individual health outcome - Testing uptake									
9[k]	observational studies	serious[l]	not serious[i]	serious[m]	not serious[j]	none	When only considering data sets where study design offered self-testing as an option and was voluntary (5 data sets), median uptake of self-testing was 58.6% (Q1 = 44.5%; Q3 = 61.8%). In another four data sets, studies' designs required self-testing as part of study participation. Across all nine data sets, the median testing uptake was 83.6% (Q1 = 58.6%; Q3 = 98.5%).	⊕◯◯◯ Very low	CRITICAL
Individual health outcome - Time to diagnosis									
8[n]	observational studies	serious[o]	not serious[i]	not serious[d]	not serious[j]	none	Self-testing usually provides results within 20 to 30 minutes. Self-testing was used to decide on whether people need to isolate in a one-off testing regime (3 data sets) and in regular school testing (5 data sets). In no data set, self-testing was used for clinical diagnosis in symptomatic persons.	⊕◯◯◯ Very low	IMPORTANT

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Individual health outcome - Result reporting									
18[p]	observational studies	serious[q]	not serious[i]	serious[r]	not serious[i]	none	For the majority of data sets (10 data sets; 55.5%), the proportion of results reported was uncertain. In 11.1% (2) of data sets, study participants were contacted by phone if no test results were submitted, leading to 90.7% results reported. Studies where self-testing was required for study participation (6 data sets; 33.3%), the proportion of results reported was assumed to be 100%. It was not possible to differentiate between the proportion of results reported following positive or negative self-testing results.	⊕◯◯◯ Very low	IMPORTANT
Community health outcome – Number/proportion of infectious cases detected									
14[s]	observational studies	serious[t]	not serious[i]	not serious[d]	not serious[i]	none	In ten of these data sets (71.4%), the comparator was 'no testing'. The median test positivity rate was 1.7% (Q1 = 0.3%; Q3 = 1.9%). Due to the comparator of 'no testing', at the time these studies were conducted, it was assumed that none of the cases would have been detected without self-testing.	⊕◯◯◯ Very low	IMPORTANT
Community health outcome – Impact on virus transmission									
2[u]	observational studies	serious[v]	serious[w]	not serious[d]	very serious[x]	none	One data set reported that self-testing could enable indoor care home visits without significantly increasing the proportion of outbreaks in these care homes compared to others where visitors were limited to outdoor visits only. In the other data set, daily self-testing was used as an alternative to self-quarantine for contacts of cases, but no significant change in secondary cases was detected.	⊕◯◯◯ Very low	IMPORTANT

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Broader societal effects – Impact on absenteeism or economic outputs									
4[y]	observational studies	serious[z]	not serious[i]	not serious[d]	not serious[i]	none	In two data sets, where daily self-testing was compared to quarantine for contacts of confirmed SARS-CoV-2 cases, self-testing reduced the quarantine time to zero, as contacts were not required to isolate when providing a negative Ag-RDT self-testing result daily. In one of these data sets, work absenteeism was also reduced, since the persons using self-testing to leave quarantine were police officers, fire fighters, and hospital staff. Self-testing was also reported to support indoor care-home visits and to increase the wellbeing of children in school.	⨁◯◯◯ Very low	IMPORTANT
Accuracy – Proportion of user errors									
1[f]	observational studies	not serious[b]	serious[c]	not serious[d]	not serious[i]	none	15.5% of the sampling steps and 15.0% of testing steps, were found to have deviations by study participants. However, these did not impede the self-test's performance.	⨁⨁◯◯ Low	IMPORTANT
Individual health outcome – Linkage for negatives									
11[aa]	observational studies	serious[ab]	not serious[i]	not serious[d]	not serious[i]	none	Negative self-testing results were followed by the continuation of operation (DS=8), the permission to leave quarantine (DS=2), or indoor visits at a care home (DS=1). In eight of these DS, people had to retest within a given time interval.	⨁◯◯◯ Very low	IMPORTANT

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Individual health outcome – Behaviour change									
3[ac]	observational studies	serious[ad]	not serious[i]	not serious[d]	not serious[i]	none	In one data set, where daily self-testing was used as an alternative to quarantine after contact with a SARS-CoV-2 case, 77.1% of people with a negative result who would have otherwise stayed in quarantine reported to meet other persons. On the contrary, in two other data sets where self-testing was used as a screening tool in school, changes in risky behaviour did not occur.	⊕◯◯◯ Very low	IMPORTANT
Community health outcome – Impact on morbidity and/or mortality									
0								-	IMPORTANT
Individual health outcome - Social harm									
0								-	CRITICAL
Individual health outcome – Testing frequency									
0								-	CRITICAL

Certainty assessment							Impact	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Effects on the health system									
0								-	NOT IMPORTANT
Misuse									
0								-	NOT IMPORTANT
Adverse events									
0								-	NOT IMPORTANT

CI: confidence interval; **rRT-PCR**: real-time reverse transcriptase-polymerase chain reaction

Explanations

a. Lindner, A.K., et al., 2021; Stohr, J.J., et al., 2021 (2 data sets)

b. We used QUADAS-2 to assess risk of bias. Most studies enrolled patients consecutively and assessed the self-testing results blinded to the reference standard result (rRT-PCR or prof. Ag-RDT testing). While for one study it was not clear whether all self-tests were performed as per manufacturer's instructions, this was ensured in the other. Furthermore, we could not detect any potential bias resulting from the study flow and timing. Therefore, we did not downgrade the quality of evidence for this criterion.

c. The heterogeneity/inconsistency in findings, as shown by the wide-ranging point estimates (sensitivity from 48.9% to 82.5%) with only marginally overlapping confidence intervals, is likely to originate from differences in the study population. This is strengthened by the fact that the head-to-head comparison on the same study population shows similar performance for self- and professional-testing. However, as there is only one study available for concordance and user errors, we downgrade for these two outcomes by one.

d. Following current guidance from the GRADE guideline, we do not downgrade by one point for all studies but acknowledge that the study populations are not fully representative of the populations of interest. Furthermore, the intervention did not differ from the one of interest and outcomes were reported directly, therefore indirectness was judged 'not serious'.

e. The number of studies and sample size was small, and only one study reported on concordance. Confidence intervals were large for sensitivity (82.5% [95% CI 67.2 to 92.7)] in Lindner, A.K., et al., 2021; 48.9% [95% CI 41.3 to 56.5] and 61.5% [95% CI 54.2 to 68.4] in Stohr, J.J., et al., 2021), because of the small numbers of positive cases. The confidence intervals for specificity were narrow (100% [95% CI 96.5 to 100] in Lindner, A.K., et al., 2021; 99.9% [95% CI 99.5 to 100] and 99.7% [95% CI 99.3 to 100] in Stohr, J.J., et al., 2021). Therefore, we downgraded one point for sensitivity and concordance, but none for specificity.

f. Lindner, A.K., et al., 2021

g. Hirst, J. A., et al., 2021 (2 data sets); Lamb, G., et al., 2021; Love, N., et al., 2021; Tulloch, J. S. P., et al., 2021; University of Liverpool, 2021 (2 data sets); Wachinger, J., et al., 2021 (2 data sets); Willeit, P., et al., 2021 (3 data sets)

h. Included data sets showed a comparably high risk of bias on the Newcastle Ottawa Scale (median = 3.5 stars; Q1 = 3; Q3 = 5)

i. For this outcome only qualitative data, or quantitative data in isolated studies in well-described but not comparable settings were available, therefore the criterion 'inconsistency' is negligible and rated as 'not serious'.

j. For this outcome only qualitative data, or quantitative data in isolated studies in well-described but not comparable settings were available, therefore the criterion 'imprecision' is negligible and rated as 'not serious'.

k. Hirst, J.A., et al., 2021; Kheiroddin, P., et al., 2021; Lamb, G., et al., 2021; Love, N., et al., 2021; Wachinger, J., et al., 2021; Willeit, P., et al., 2021

l. Included data sets show a comparably high risk of bias on the Newcastle Ottawa Scale (median = 3 stars; Q1 = 3; Q3 = 4). In addition, in four of the data sets (Kheiroddin, P., et al., 2021; Willeit, P., et al., 2021 [3 data sets]) self-testing was compulsory for study participants by study design, strongly impacting the testing uptake.

m. Following current guidance from the GRADE guideline, we do not downgrade by one point for all studies but acknowledge that the study populations are not fully representative of the populations of interest. Nonetheless, for two data sets, the outcome of interest had to be estimated (Kheiroddin, P., et al.) or calculated in multiple calculation steps (Love, N., et al.) from surrogate parameter. Thus, indirectness was rated as 'serious'.

n. Hirst, J.A., et al., 2021 (2 data sets); Lamb, G., et al., 2021; Wachinger, J., et al., 2021 (2 data sets); Willeit, P., et al., 2021 (3 data sets)

o. Included data sets showed a comparably high risk of bias on the Newcastle Ottawa Scale (median = 4 stars; Q1 = 3; Q3 = 5)

p. Downs, L.O., et al., 2021; Hirst, J.A., et al., 2021 (2 data sets); Hoehl, S., et al., 2021; Kheiroddin, P., et al., 2021; Lamb, G., et al., 2021; Love, N., et al., 2021; Stohr, J.J.J.M., et al., 2021 (2 data sets); Tulloch, J.S.P., et al., 2021 (2 data sets); University of Liverpool, 2021 (2 data sets); Wachinger, J., et al., 2021 (2 data sets); Willeit, P., et al., 2021 (3 data sets)

q. Included data sets showed a comparably high risk of bias on the Newcastle Ottawa Scale (median = 3 stars; Q1 = 3; Q3 = 4)

r. Following current guidance from the GRADE guideline, we do not downgrade by one point for all studies but acknowledge that the study populations are not fully representative of the populations of interest. Nonetheless, for six data sets, the proportion of results reported had to be estimated based on the fact that self-testing was required for study participants by study design (Kheiroddin, P., et al., 2021; Tulloch, J.S.P., et al., 2021; University of Liverpool, 2021; Willeit, P., et al., 2021 [3 data sets]), therefore indirectness was rated as 'serious'.

s. Downs, L.O., et al., 2021; Hirst, J.A., et al., 2021 (2 data sets); Hoehl, S., et al., 2021; Lamb, G., et al., 2021; Stohr, J.J.J.M., et al., 2021 (2 data sets); Tulloch, J.S.P., et al., 2021 (2 data sets); University of Liverpool, 2021 (2 data sets); Willeit, P., et al., 2021 (3 data sets)

t. Included data sets showed a comparably high risk of bias on the Newcastle Ottawa Scale (median = 3 stars; Q1 = 3; Q3 = 4.8)

u. Tulloch, J.S.P., et al., 2021 (2 data sets); Love, N., et al., 2021

v. Included data sets showed a comparably high risk of bias on Newcastle Ottawa Scale (median = 4 stars; Q1 = 3; Q3 = 4.5)

w. Data was limited and heterogeneous, with the two studies reporting different sub-outcomes (proportion of outbreaks: Tulloch, J.S.P., et al., 2021; secondary attack rates: Love, N., et al., 2021). Therefore, inconsistency was ranked as 'serious'.

x. Confidence intervals were too wide (proportion of outbreaks: 54.5% (95% CI 23.4% - 83.3%), Tulloch, J.S.P., et al., 2021; secondary attack rates: 6.3% (95% CI: 3.4% - 11.1%), Love, N., et al., 2021) to precisely judge the effect of the intervention on virus transmission.

y. Kheiroddin, P., et al., 2021; Love, N., et al., 2021; Tulloch, J.S.P., et al., 2021; University of Liverpool, 2021

z. Included data sets showed a comparably high risk of bias on the Newcastle Ottawa Scale (median = 4.5 star; Q1 = 4; Q3 = 5.5)

aa. Downs, L.O., et al., 2021; Hirst, J.A., et al., 2021 (2 data sets); Love, N., et al., 2021; Tulloch, J.S.P., 2021; University of Liverpool, 2021; Wachinger, J., et al., 2021 (2 data sets); Willeit, P., et al., 2021 (3 data sets)

ab. Included data sets showed a comparably high risk of bias on the Newcastle Ottawa Scale (median = 3 stars; Q1 = 3; Q3 = 5)

ac. Wachinger, J., et al., 2021 (2 data sets); Love, N., et al., 2021

ad. Included data sets showed a comparably high risk of bias on the Newcastle Ottawa Scale (median = 3 stars; Q1 = 2; Q3 = 4)