Table S1, Quality assessment of single arm studies (according to NIH tool):

Domains	Bussel 2012	Saleh 2012	Katsutani 2013
1. Was the research question or objective in this paper clearly stated?	Yes	Yes	Yes
2. Were eligibility/selection criteria for the study population prespecified and clearly described?	No	Yes	Yes
3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	Yes	Yes	Yes
4. Were all eligible participants that met the prespecified entry criteria enrolled?	Yes	Yes	Yes
5. Was the sample size sufficiently large to provide confidence in the findings?	No	NR	NR
6. Was the test/service/intervention clearly described and delivered consistently across the study population?	Yes	Yes	Yes
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	Yes	Yes	Yes
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?	No	No	No
9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis?	Yes	Yes	Yes
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	Yes	No	NR
11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	Yes	Yes	Yes
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?	NA	NA	NA
Total scores (Yes = 1, No = 0.5, NR & NA & CD = 0)	9.5	9	8.5
Quality rating: good (11-12 point) or fair (10-8 point) or poor (7-0 points)	**Fair quality**	**Fair quality**	**Fair quality**

NA: not applicable, CD: cannot determine, NR: not reported.

Table S2, Quality assessment of observational studies (according to NIH tool):

Domains	Forsythe 2020	Donga 2017	Cekdemir 2018	Lopez 2015	Giordano 2020	Cheng 2019
1. Was the research question or objective in this paper clearly stated?	Yes	Yes	Yes	Yes	Yes	Yes
2. Was the study population clearly specified and defined?	Yes	Yes	Yes	Yes	Yes	Yes
3. Was the participation rate of eligible persons at least 50%?	NR	NA	NA	NA	NA	NA
4. Were all the subjects selected or recruited from the same or similar populations? Were inclusion and exclusion criteria for being in the study pre-specified and applied uniformly to all participants?	Yes	Yes	Yes	Yes	Yes	Yes
5. Was a sample size justification, power description, or variance and effect estimates provided?	NR	No	NR	NR	No	No
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	Yes	Yes	Yes	Yes	Yes	Yes
7. Was the time frame sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	NA	NA	NA	NA	NA	NA
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?	No	No	No	No	No	Yes
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	NA	NA	NA	NA	NA	NA
10. Was the exposure(s) assessed more than once over time?	NA	NA	NA	NA	NA	NA
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes	Yes	Yes	Yes	Yes
12. Were the outcome assessors blinded to the exposure status of participants?	NR	NA	NA	NA	NA	NA
13. Was loss to follow-up after baseline 20% or less?	Yes	NA	NA	NA	NA	NA
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) an outcome(s)?	No	No	No	No	No	No
Total scores (Yes = 1, No = 0.5, NR & NA & CD = 0)	7	6.5	6	6	6.5	7
Quality rating: good (14-13 point) or fair (9-12 point) or poor (8-0 points)	**Poor quality**	**Poor quality**	**Poor quality**	**Poor quality**	**Poor quality**	**Poor quality**

NA: not applicable, CD: cannot determine, NR: not reported.