Web Annex B. GRADE Table: Should HIV self-testing be offered as an additional HIV testing approach?
Contents

Table B1. GRADE table for HIVST compared to SOC for HTS ................................................................. 4

Table B2. GRADE table for HIVST + linkage intervention compared to HIVST only for linkage to HIV CARE or ART ................................................................. 33
# Table B1. GRADE table for HIVST compared to SOC for HTS

**Author(s):** Muhammad S. Jamil, T. Charles Witzel, Ingrid Wilson  
**Question:** HIVST compared to SOC for HTS

<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>Nº of patients</th>
<th>Effect</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIVST</td>
<td>SOC</td>
<td>Relative (95% CI)</td>
<td>Absolute (95% CI)</td>
</tr>
<tr>
<td>Nº of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td>23 a</td>
<td>randomised trials b</td>
<td>serious c</td>
<td>not serious d</td>
<td>not serious e</td>
</tr>
<tr>
<td>13 g</td>
<td>randomised trials h</td>
<td>very serious i</td>
<td>not serious j</td>
<td>not serious k</td>
</tr>
<tr>
<td>10 m</td>
<td>randomised trials n</td>
<td>serious o</td>
<td>not serious p</td>
<td>not serious q</td>
</tr>
<tr>
<td>7 s</td>
<td>randomised trials t</td>
<td>serious u</td>
<td>not serious v</td>
<td>not serious w</td>
</tr>
<tr>
<td>3 x</td>
<td>randomised trials y</td>
<td>serious z</td>
<td>not serious aa</td>
<td>not serious ab</td>
</tr>
</tbody>
</table>

### Uptake of HIV testing (overall)

- **Number of patients:** 23,000
- **Relative effect:** 1.62
- **Absolute effect:** 250 more per 1,000

### Uptake of HIV testing (general population)

- **Number of patients:** 13,000
- **Relative effect:** 2.09
- **Absolute effect:** 436 more per 1,000

### Uptake of HIV testing (key populations)

- **Number of patients:** 10,000
- **Relative effect:** 1.28
- **Absolute effect:** 116 more per 1,000

### Uptake of HIV testing (men who have sex with men)

- **Number of patients:** 7,000
- **Relative effect:** 1.37
- **Absolute effect:** 156 more per 1,000

### Uptake of HIV testing (female sex workers)

- **Number of patients:** 3,000
- **Relative effect:** 1.16
- **Absolute effect:** 63 more per 1,000
<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>Nº of patients</th>
<th>Effect</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nº of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td>17</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>6</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>5</td>
<td>randomised trials</td>
<td>very serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>17</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>6</td>
<td>randomised trials</td>
<td>very serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>5</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
</tbody>
</table>

Uptake of HIV testing (women)

Uptake of HIV testing (young people 15-24 years)

Uptake of HIV testing (measurement time-point: ≤6 months)

Uptake of HIV testing (measurement time-point: >6 months)

Uptake of HIV testing (online and mail distribution)

Uptake of HIV testing (facility-based distribution)
<table>
<thead>
<tr>
<th>№ of studies</th>
<th>Study design</th>
<th>Risk of bias</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other considerations</th>
<th>Effect</th>
<th>Absolute (95% CI)</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bg</td>
<td>randomised trials bg</td>
<td>serious bg</td>
<td>not serious bj</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>903/1155 (78.2%) bk</td>
<td>520/810 (64.2%) bk</td>
<td>□□□</td>
<td>MODERATE</td>
</tr>
<tr>
<td>4 bd</td>
<td>randomised trials bd</td>
<td>serious bd</td>
<td>not serious bo</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>4477/5737 (78.0%) bp</td>
<td>840/2578 (32.6%) bp</td>
<td>□□□</td>
<td>MODERATE</td>
</tr>
<tr>
<td>2 bi</td>
<td>randomised trials bi</td>
<td>very serious bi</td>
<td>not serious bi</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>507/823 (61.6%) bz</td>
<td>120/369 (32.5%) bz</td>
<td>□□</td>
<td>LOW</td>
</tr>
<tr>
<td>3 cb</td>
<td>randomised trials cb</td>
<td>very serious cb</td>
<td>not serious cd</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>6530/9559 (68.3%) cc</td>
<td>4419/8516 (51.9%) ce</td>
<td>□□</td>
<td>LOW</td>
</tr>
<tr>
<td>4 cf</td>
<td>randomised trials cf</td>
<td>very serious cf</td>
<td>not serious ch</td>
<td>not serious</td>
<td>not serious</td>
<td>none</td>
<td>1344/3747 (35.9%) ci</td>
<td>414/3564 (11.6%) ci</td>
<td>□□</td>
<td>LOW</td>
</tr>
</tbody>
</table>

**Uptake of HIV testing (secondary distribution - women to male partners)**

- **Relative (95% CI):** 1.28 (1.01 to 1.61)
- **Absolute (95% CI):** 180 more per 1,000 (from 6 more to 392 more)

- **Risk of bias:** serious
- **Study design:** randomised trials
- **Inconsistency:** not serious
- **Indirectness:** not serious
- **Imprecision:** none
- **Other considerations:** none

**Uptake of HIV testing services (secondary distribution - HIV-positive clients to partners)**

- **Relative (95% CI):** 2.63 (1.81 to 3.82)
- **Absolute (95% CI):** 531 more per 1,000 (from 264 more to 919 more)

- **Risk of bias:** very serious
- **Study design:** randomised trials
- **Inconsistency:** not serious
- **Indirectness:** not serious
- **Imprecision:** none
- **Other considerations:** none

**Uptake of HIV testing (secondary distribution - peers)**

- **Relative (95% CI):** 2.03 (1.01 to 4.09)
- **Absolute (95% CI):** 335 more per 1,000 (from 3 more to 1,000 more)

- **Risk of bias:** serious
- **Study design:** randomised trials
- **Inconsistency:** not serious
- **Indirectness:** not serious
- **Imprecision:** none
- **Other considerations:** none

**Uptake of HIV testing (community- or home-based distribution)**

- **Relative (95% CI):** 1.43 (0.95 to 2.13)
- **Absolute (95% CI):** 223 more per 1,000 (from 26 fewer to 586 more)

- **Risk of bias:** very serious
- **Study design:** randomised trials
- **Inconsistency:** not serious
- **Indirectness:** not serious
- **Imprecision:** none
- **Other considerations:** none

**Uptake of HIV testing (HIVST at facilities)**

- **Relative (95% CI):** 2.14 (1.22 to 3.74)
- **Absolute (95% CI):** 132 more per 1,000 (from 26 more to 318 more)

- **Risk of bias:** very serious
- **Study design:** randomised trials
- **Inconsistency:** not serious
- **Indirectness:** not serious
- **Imprecision:** none
- **Other considerations:** none

**Uptake of HIV testing (no or basic support)**
<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>№ of patients</th>
<th>Effect</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>№ of studies</strong></td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td>4</td>
<td>randomised trials c</td>
<td>very serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>8</td>
<td>randomised trials c</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>3</td>
<td>randomised trials c</td>
<td>serious</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td>4</td>
<td>randomised trials c</td>
<td>very serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>3</td>
<td>randomised trials d</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>1</td>
<td>randomised trials</td>
<td>serious</td>
<td>serious</td>
<td>not serious</td>
</tr>
</tbody>
</table>

Uptake of HIV testing (IFU enhancement, video or study hotline)

Uptake of HIV testing (group demonstration)

Uptake of HIV testing (in-person demonstration or training)

Uptake of HIV testing (in-person observation or supervision)

Uptake of HIV testing (virtual real-time support or supervision)
### Certainty assessment

<table>
<thead>
<tr>
<th>№ of studies</th>
<th>Study design</th>
<th>Risk of bias</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other considerations</th>
<th>HIST</th>
<th>SOC</th>
<th>Relative (95% CI)</th>
<th>Absolute (95% CI)</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>№ of patients</td>
<td>Effect</td>
<td>Certainty</td>
<td>Importance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Frequency of HIV testing (men who have sex with men)**

- **3** randomised trials
  - **serious** not serious
  - **not serious** not serious
  - **none** not serious
  - **Frequency of HIV testing (men who have sex with men)**
    - Among 3 trials 1242 participants in the HIST arm reported 6351 HIV tests over 12 - 15 months compared to 1222 participants reporting 2107 HIV tests in the SOC arm during the same period. This represents 2.6 additional tests per participant during follow-up period (mean difference: 2.6, 95% CI: 1.2 - 4.0)
  - **HIV positivity rate among randomized (overall)**
    - **16** randomised trials
      - **serious** not serious
      - **not serious** not serious
      - **none** not serious
      - **HIV positivity rate among tested (overall)**
        - **17** randomised trials
          - **serious** not serious
          - **not serious** not serious
          - **none** not serious
          - **HIV positivity rate among tested (general population)**
            - **8** randomised trials
              - **serious** not serious
              - **not serious** not serious
              - **none** not serious
              - **HIV positivity rate among tested (key populations)**
                - **9** randomised trials
                  - **serious** not serious
                  - **not serious** not serious
                  - **none** not serious
                  - **HIV positivity rate among tested (men who have sex with men)**
                    - **6** randomised trials
                      - **serious** not serious
                      - **not serious** not serious
                      - **none** not serious
                      - **HIV positivity rate among tested (female sex workers)**

<p>| 3 | randomised trials | serious | not serious | not serious | not serious | none | Among 3 trials 1242 participants in the HIST arm reported 6351 HIV tests over 12 - 15 months compared to 1222 participants reporting 2107 HIV tests in the SOC arm during the same period. This represents 2.6 additional tests per participant during follow-up period (mean difference: 2.6, 95% CI: 1.2 - 4.0) | ⬤⬸⬸ | MODERATE | IMPORTANT |
| 16 | randomised trials | serious | not serious | not serious | not serious | none | 388/16636 (2.3%) | 143/10330 (1.4%) | RR 1.06 (0.76 to 1.48) | 1 more per 1,000 (from 3 fewer to 7 more) | ⬤⬸⬸ | MODERATE | CRITICAL |
| 17 | randomised trials | serious | not serious | not serious | not serious | none | 444/13824 (3.2%) | 357/9689 (3.7%) | RR 0.97 (0.74 to 1.27) | 1 fewer per 1,000 (from 10 fewer to 10 more) | ⬤⬸⬸ | MODERATE | CRITICAL |
| 8 | randomised trials | serious | not serious | not serious | not serious | none | 227/11088 (2.0%) | 243/7932 (3.1%) | RR 0.81 (0.45 to 1.47) | 6 fewer per 1,000 (from 17 fewer to 14 more) | ⬤⬸⬸ | MODERATE | CRITICAL |
| 9 | randomised trials | serious | not serious | not serious | not serious | none | 217/2736 (7.9%) | 114/1757 (6.5%) | RR 0.99 (0.75 to 1.30) | 1 fewer per 1,000 (from 16 fewer to 19 more) | ⬤⬸⬸ | MODERATE | CRITICAL |
| 6 | randomised trials | serious | not serious | not serious | not serious | none | 34/1538 (2.2%) | 14/1152 (1.2%) | RR 1.59 (0.87 to 2.89) | 7 more per 1,000 (from 2 fewer to 23 more) | ⬤⬸⬸ | MODERATE | CRITICAL |</p>
<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>Nº of patients</th>
<th>Effect</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nº of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td>3</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>13</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>4</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>4</td>
<td>randomised trials</td>
<td>very serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>13</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>4</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Certainty assessment</td>
<td>№ of patients</td>
<td>Effect</td>
<td>Certainty</td>
<td>Importance</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------</td>
<td>--------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>HIV positivity rate among tested (facility-based distribution)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV positivity rate among tested (secondary distribution - women to male partners)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV positivity rate among tested (secondary distribution - HIV-positive clients to partners)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV positivity rate among tested (secondary distribution - peers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV positivity rate among tested (community- or home-based distribution)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Web Annex B
<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>№ of patients</th>
<th>Effect</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>№ of patients</td>
<td>Relative (95% CI)</td>
<td>Absolute (95% CI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HIVST</td>
<td>SOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>(95% CI)</td>
<td>(95% CI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>randomised trials</td>
<td>very serious</td>
<td>not serious</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage to additional or confirmatory HIV testing (HIVST arm only)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>randomised trials</td>
<td>serious</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among randomized (Overall)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (Overall)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (general population)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (key populations)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (men who have sex with men)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certainty assessment</td>
<td>№ of patients</td>
<td>Effect</td>
<td>Certainty</td>
<td>Importance</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------</td>
<td>--------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td>№ of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (female sex workers)</td>
<td>3</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (men)</td>
<td>2</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (women)</td>
<td>8</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (young people 15-24 years)</td>
<td>4</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (measurement time-point: ≤6 months)</td>
<td>2</td>
<td>randomised trials</td>
<td>very serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Linkage to ART initiation or HIV care among HIV positive (measurement time-point: &gt;6 months)</td>
<td>8</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
<td>Imprecision</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>---------------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>3</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td></td>
<td></td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>9</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>1</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>1</td>
<td>randomised trials</td>
<td>serious</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td>1</td>
<td>randomised trials</td>
<td>serious</td>
<td>serious</td>
<td>not serious</td>
</tr>
<tr>
<td>2</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
</tbody>
</table>

**Certainty assessment**

**Importance**

**Linkage to ART initiation or HIV care among HIV positive (no linkage support)**

**Linkage to ART initiation or HIV care among HIV positive (phone reminder or follow up)**

**Linkage to ART initiation or HIV care among HIV positive (home visit or in-person referral)**

**Linkage to ART initiation or HIV care among HIV positive (virtual real-time support)**

**Linkage to ART initiation or HIV care among HIV positive (financial incentive)**

**Misuse of HIVST kits - coercion (HIVST arm only)**
<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>№ of patients</th>
<th>Effect</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td>2</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>6</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>4</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>4</td>
<td>randomised trials</td>
<td>serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
</tbody>
</table>

**Adverse events among those randomized**

**Condomless anal sex with male partners among those randomized**

**Condomless anal sex with male partners among those who completed follow-up**

**CI:** Confidence interval; **RR:** Risk ratio

**Explanations**

a. Meta-analysis: Choko, 2019a and Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined in these trials; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting); Choko, 2019b had two study groups (ANC women and ART clients) and these were presented separately. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

b. 14 individual randomized trials, 9 cluster randomized trials.

c. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 14 trials and attrition bias in 3 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm; Patel, 2018: 36% LTFU overall, 44% in the intervention and 27% in the control arm). Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.
d. There was high statistical heterogeneity (Heterogeneity: \( \tau^2 = 0.091; \ Chi^2 = 409.12, df = 25, p < 0.01; I^2 = 94\%, 92\% - 95\% \)). Sub-group analyses by population, study design, measure time-point, and distribution method did not fully explain heterogeneity. Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

e. We did not downgrade for indirectness.

f. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

g. Meta-analysis: Choko, 2019a had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined. Choko, 2019b had two study groups (ANC women and ART clients) and these were presented separately. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

h. 7 individual randomized trials and 6 cluster randomized trials

i. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 10 trials and attrition bias in one trial (Patel, 2018: 36\% LTFU overall, 44\% in the intervention and 27\% in the control arm). Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts. 9 of 13 trials had more than three high risk or unclear risk of bias domains.

j. There was a high statistical heterogeneity (Heterogeneity: \( \tau^2 = 0.133; \ Chi^2 = 213.31, df = 13, p < 0.01; I^2 = 94\%, 91\% - 96\% \)). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

k. We did not downgrade for indirectness but noted that all but one trial were conducted in Africa (6 in Malawi, 4 in Kenya, one in Zambia, one in South Africa, one in the US). This is expected most countries with generalized epidemics are in Africa.

l. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

m. Meta-analysis: Wray, 2018 had two intervention arms which both included HIVST with similar HIVST kit distribution method, therefore both arms were combined. Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

n. 7 individual randomized trials and 3 cluster randomized trials

o. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 5 trials and attrition bias in 2 trials (MacGowan, 2017: 27.1\% LTFU in the intervention arm and 28.5\% in the control arm; Merchant, 2018: 38.4\% LTFU overall, 26\% in the intervention 50\% in the control arm). Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

p. There was high statistical heterogeneity (Heterogeneity: \( \tau^2 = 0.040; \ Chi^2 = 108.69, df = 11, p < 0.01; I^2 = 90\%, 846\% - 93\% \)). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

q. We did not downgrade but noted that only men who have sex with men, transgender women and female sex workers were represented in included studies and results should be viewed with caution when applying to other key populations.

r. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

s. Meta-analysis: Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore intervention arms were combined. in these trials. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

t. 6 individual randomized trials and one cluster randomized trial.
We downgraded once. due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 3 trials and attrition bias in 2 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

There was high statistical heterogeneity (Heterogeneity: Tau² = 0.028; Chi² = 45.50, df = 6, p < 0.01; I² = 87%, 75% - 93%). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

One individual randomized trial and 2 cluster randomized trials.

We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and detection bias (self-reported or non-validated outcomes) in 2 trials and unclear risk in one trial. One trial had unclear risk associated with random sequence generation.

There was high statistical heterogeneity (Heterogeneity: Tau² = 0.052; Chi² = 39.90, df = 4, p < 0.01; I² = 90%, 79% - 95%). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

7 trials were conducted among MSM, 4 were among male partners of ANC women, 2 were among male truck drivers, and for the remaining sub-analysis among men was included. Meta-analysis: Choko, 2019a and Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined in these trials; Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms and presented ANC data only. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

11 individual randomized trials and 6 cluster randomized trials.

We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 11 trials and attrition bias in 2 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

There was high statistical heterogeneity (Heterogeneity: Tau² = 0.084; Chi² = 240.11, df = 17, p < 0.01; I² = 93%, 91% - 95%). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

Three trials were conducted among female sex workers, one was among young women, and for two trials sub-analysis among women was included. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

2 individual randomized trials and 4 cluster randomized trials.

We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and detection bias (self-reported or non-validated outcomes) in 3 trials or unclear detection bias in 2 trials. One cluster randomized trial was subject to recruitment bias. Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.
ak. There was high statistical heterogeneity (Heterogeneity: \( \tau^2 = 0.110; \chi^2 = 62.95, \text{df} = 7, p < 0.01; I^2 = 89\%, 80\% - 94\%)\). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

al. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

am. Age groups 15–24 years were included. One trial was conducted among young women, one among young MSM, and for three sub-analysis among young people was included. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

an. 2 individual randomized trials and 3 cluster randomized trials.

ao. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 4 trials or unclear risk of detection bias in one trial, and attrition bias in one trial (Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts. Each of the trials had more than three high risk or unclear risk of bias domains.

ap. There was high statistical heterogeneity (Heterogeneity: \( \tau^2 = 0.230; \chi^2 = 82.25, \text{df} = 4, p < 0.01; I^2 = 95\%, 91\% - 97\%)\). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

aq. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ar. Meta-analysis: Choko, 2019a had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting); Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but presented sub-groups separately. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

as. 10 individual randomized trials and 7 cluster randomized trials.

at. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 11 trials or unclear risk of detection bias in 3 trials, and attrition bias in 2 trials (Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm; Patel, 2018: 36% LTFU overall, 44% in the intervention and 27% in the control arm). Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

au. There was high statistical heterogeneity (Heterogeneity: \( \tau^2 = 0.131; \chi^2 = 351.02, \text{df} = 20, p < 0.01; I^2 = 95\%, 93\% - 96\%)\). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

av. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

aw. Meta-analysis: Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore intervention arms were combined. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

ax. 4 individual randomized trials and 2 cluster randomized trials.

ay. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 4 trials or unclear risk of detection bias in one trial, and attrition bias in one trial (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm). Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts. 4 of 6 trials had more than three high risk or unclear risk of bias domains.

az. There was high statistical heterogeneity (Heterogeneity: \( \tau^2 = 0.029; \chi^2 = 42.71, \text{df} = 5, p < 0.01; I^2 = 88\%, 77\% - 94\%)\). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

ba. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.
bb. Meta-analysis: Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore intervention arms were combined. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

bc. 4 individual randomized trials and one cluster randomized trial.

bd. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 3 trials and attrition bias in 2 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

be. There was high moderate statistical heterogeneity (Heterogeneity: Tau² = 0.012; Chi² = 12.51, df = 4, p = 0.01; I² = 68%, 17% - 88%). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

bf. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

bg. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, in this analysis comparison of one arm (facility-based HIVST distribution) with standard of care was included. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

bh. 4 individual randomized trials and 2 cluster randomized trials.

bi. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 3 trials or unclear risk for detection bias in 2 trials, and attrition bias in one trial (Patel, 2018: 36% LTFU overall, 44% in the intervention and 27% in the control arm). There was unclear risk of bias for random sequence generation for Patel, 2018 and Pettifor, 2018.

bj. There was high statistical heterogeneity (Heterogeneity: Tau² = 0.068; Chi² = 60.88, df = 5, p < 0.01; I² = 92%, 85% - 96%). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

bk. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

bl. Meta-analysis: Choko, 2019a had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore intervention arms were combined. Choko, 2019b had two study groups (ANC women and ART clients), only ANC women group was included in this analysis. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

bm. 2 individual randomized trials and 2 cluster randomized trials.

bn. We downgraded once. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in all trials. One trial had unclear risk of attrition bias and two cluster randomized trials were subject to recruitment bias.

bo. There was high statistical heterogeneity (Heterogeneity: Tau² = 0.133; Chi² = 53.54, df = 3, p < 0.01; I² = 94%, 89% - 97%). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

bp. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

bq. Meta-analysis: Choko, 2019b had two study groups (ANC women and ART clients), but only ART group was included in this analysis. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

br. One individual randomized trial and one cluster randomized trial.

bs. We downgraded twice. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in both trials. Choko, 2019b cluster randomized trial was subject to recruitment bias. Each of the trials had more than three high risk or unclear risk of bias domains.
bt. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.211; \chi^2 = 5.45, \text{df} = 1, p = 0.02; I^2 = 82\%$). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

bu. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

bv. Meta-analysis: Both trials (Chanda, 2017 and Ortblad, 2017) had more than one intervention arm which included HIVST but HIVST kit distribution method was different, only one arm which involved direct distribution of HIVST kits to peers compared to standard of care is included in this analysis. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

bw. 2 cluster randomized trials.

bx. We downgraded once. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in both trials.

by. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.018; \chi^2 = 5.02, \text{df} = 1, p = 0.03; I^2 = 80\%$). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

bz. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

cb. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

cc. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

cb. 3 cluster randomized trials.

dc. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials and detection bias (self-reported or non-validated outcomes) in three trials. Some other risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts. All trials had more than three high risk or unclear risk of bias domains.

dd. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.122; \chi^2 = 52.78, \text{df} = 3, p < 0.01 ; I^2 = 96\%, 92\%-98\%$). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

dce. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

cf. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

cg. 2 cluster randomized trials.

dh. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in one trial or unclear detection bias in five trials. Three of 4 trials had unclear risk of bias for random sequence generation. Three of 4 trials had more than three high risk or unclear risk of bias domains.

daj. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.274; \chi^2 = 29.10, \text{df} = 3, p < 0.01; I^2 = 90\%, 76\%-95\%$). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

dcj. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

dck. Meta-analysis: Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore intervention arms were combined. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

dcl. 4 individual randomized trials.

dcm. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 3 or unclear detection bias in one trial and attrition bias in 2 trials (Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm; Patel, 2018: 36% LTFU overall, 44% in the intervention and 27% in the control arm). Each of the trials had more than three high risk or unclear risk of bias domains.
There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.090; \chi^2 = 18.22, df = 3, p < 0.01; I^2 = 84\%, 58\% - 94\%)$. Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

cr. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

cp. Meta-analysis: Choko, 2019a had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined. Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but presented sub-groups separately. Cluster-adj usted analysis was included in meta-analysis for cluster randomized trials.

cq. 5 individual randomized trials and 3 cluster randomized trials.

cr. We downgraded once. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in all trials, detection bias (self-reported or non-validated outcomes) in 6 trials and attrition bias in one trial (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm). Two cluster randomized trials were subject to recruitment bias. Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

cs. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.083; \chi^2 = 126.61, df = 8, p < 0.01; I^2 = 94\%, 90\% - 96\%)$. Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

cr. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

cu. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adj usted analysis was included in meta-analysis for cluster randomized trials.

cv. 3 cluster randomized trials.

cw. We downgraded once. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in all trials. One cluster randomized trial was subject to recruitment bias. One cluster randomized trial had unclear risk of bias for selection bias (allocation concealment), selective reporting, and baseline cluster imbalance.

cx. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.054; \chi^2 = 39.74, df = 4, p < 0.01; I^2 = 90\%, 79\% - 95\%)$. Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

cy. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

cz. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

da. One individual randomized trial and 3 cluster randomized trials.

db. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 3 trials or unclear detection bias in one trial. Three of 4 trials had more than three high risk or unclear risk of bias domains.

dc. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.121; \chi^2 = 85.06, df = 3, p < 0.01; I^2 = 96\%, 94\% - 98\%)$. Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

dd. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

de. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

df. 3 individual randomized trials.
We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and unclear risk for detection bias (self-reported or non-validated outcomes) in 2 trials. Two trials had unclear risk for random sequence generation (selection bias).

dh. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 0.123; \chi^2 = 10.49, df = 2, p < 0.01; I^2 = 81\%, 40\% - 94\%$). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

di. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

dj. We downgraded once due to potential for performance bias due to lack of blinding.

dk. Single trial, we downgraded once as inconsistency cannot be evaluated.

dl. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

dm. Three trials were included in meta-analysis (Jamil, 2017; Katz, 2018; MacGowan, 2017). All trials were conducted among men who have sex with men in high income countries (Australian and the United States).

dn. 3 individual randomized trials.

do. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in one trial or unclear risk of detection bias in one trial, and attrition bias in one trial (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm).

dp. There was high statistical heterogeneity (Heterogeneity: $\tau^2 = 1.435; \chi^2 = 63.07, df = 2, p < 0.01; I^2 = 97\%, 94\% - 98\%$). Study effects from individual RCTs were consistently beneficial and no difference was observed in other critical outcomes. The GDG determined that downgrading for inconsistency was not necessary. We did not downgrade.

dq. Meta-analysis: Choko, 2019a and Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined in these trials; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting); Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but presented sub-groups separately. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

dr. 11 individual randomized and 6 were cluster randomized trials.

ds. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 6 trials or unclear risk of detection bias in 4 trials, and attrition bias in two trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Three cluster randomized trials were subject to potential recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

dt. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0.117; \chi^2 = 24.26, df = 18, p = 0.15; I^2 = 26\%, 0\% - 57\%)$. We did not downgrade but noted that outcome definition was not consistent across trials: 8 trials defined positivity as confirmed HIV diagnosis; the remaining trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear.

du. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

dv. Meta-analysis: Choko, 2019a and Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined in these trials; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting); Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but presented sub-groups separately. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

dw. 11 individual randomized and 6 were cluster randomized trials.
We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 6 trials or unclear risk of detection bias in 4 trials, and attrition bias in 2 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Three cluster randomized trials were subject to potential recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

dy. There was low statistical heterogeneity (Heterogeneity: Tau² = 0.048; Chi² = 22.14, df = 19, p = 0.28; I² = 14%, 0% - 49%). We did not downgrade but noted that outcome definition was not consistent across trials: 8 trials defined positivity as confirmed HIV diagnosis; the remaining trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear. We also noted that one trial (Nichols, 2019) had significantly lower HIV positivity among those tested in the intervention arm. This trial focused on young people in a setting with high testing coverage.

dz. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ea. RR for HIV positivity among randomized: 1.06 (0.76 - 1.48).

eb. Meta-analysis: Choko, 2019a had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined; Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but presented sub-groups separately. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

ec. 4 individual randomized trials and 4 cluster randomized trials.

ed. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 2 trials and attrition bias in 2 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm. One trial had unclear risk of bias for allocation concealment (selection bias).

ee. There was low statistical heterogeneity (Heterogeneity: Tau² = 0.283; Chi² = 13.12, df = 8, p = 0.11; I² = 39%, 0% - 72%). We did not downgrade but noted that the outcome definition was not consistent across trials: 4 of 9 trials defined positivity as confirmed HIV positive diagnosis; the remaining 5 trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear.

ef. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

eg. Meta-analysis: Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined in these trials; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

eh. 7 individual randomized trials and 2 cluster randomized trials.

ei. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 4 trials or one unclear risk for detection bias, and attrition bias in 2 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Two trials had unclear risk of selection bias (one for random sequence generation and one for allocation concealment).

ej. There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 8.93, df = 10, p = 0.54; I² = 0%, 0% - 55%). We did not downgrade but noted that outcome definition was not consistent across trials: 4 of 9 trials defined positivity as confirmed HIV positive diagnosis; the remaining 5 trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear.

ek. We did not downgrade but noted that only men who have sex with men, transgender women and female sex workers were represented in included studies and results should be viewed with caution when applying to other key populations.

el. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

em. Meta-analysis: Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore intervention arms were combined. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.
en. 6 individual randomized trials.

eo. There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 1.62, df = 5, p = 0.90; I² = 0%, 0% - 22%). We did not downgrade but noted that outcome definition was not consistent across trials: 4 of 6 trials defined positivity as confirmed HIV positive diagnosis; the remaining 5 trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear.

ep. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

eq. Three 3 arm RCTs contributed to this analysis. Kelvin 2019 included two SoC arms, one enhanced and one standard. Enhanced SoC was excluded and standard SoC used in the comparison with the intervention arm. Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST: intervention arms were not pooled as differences in the intervention package appeared likely to impact positivity.

er. 2 individual randomized and one cluster randomized trials.

es. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 2 trials or unclear risk of detection bias in one trial. One trial had unclear risk of bias for random sequence generation (selection bias).

et. There was low statistical heterogeneity (Heterogeneity: Tau² = 0.008; Chi² = 4.24, df = 4, p = 0.35; I² = 6%, 0% - 80%). We did not downgrade.

eu. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ev. 6 trials were conducted among MSM, 3 were among male partners of ANC women, 2 were among male truck drivers, and for the remaining sub-analysis among men was included. Meta-analysis: Choko, 2019a and Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined in these trials; Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but included ANC group only in this analysis. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

ew. 10 individual randomized and 3 cluster randomized trials.

ex. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 4 trials or unclear risk of detection bias in 2 trials, and attrition bias in 2 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

ey. There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 8.62, df = 12, p = 0.73; I² = 0%, 0% - 40%). We did not downgrade but noted that the outcome definition was not consistent across trials: 7 of 13 trials defined positivity as confirmed HIV positive diagnosis; the remaining 7 trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear.

ez. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

fa. 3 trials were conducted among female sex workers and for the remaining one sub-analysis among women was included. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

fb. One individual randomized trial and 3 cluster randomized trials.

fc. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 3 trials or unclear risk of detection bias in one trial. One cluster randomized trial was subject to recruitment bias. Two trials had unclear risk of selection bias (one random sequence generation and one allocation concealment).

fd. There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 4.38, df = 5, p = 0.47; I² = 0%, 0% - 71%). We did not downgrade.

fe. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ff. Two trials were conducted among young people and other two presented sub-analysis among young people. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.
fg. 2 individual randomized trials and two cluster randomized trial.

fh. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in three trials or unclear risk of detection bias in one trial, and attrition bias in one trial (Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). One cluster randomized trial was subject to recruitment bias. All trials had unclear risk of selection bias. 3 of 4 trials had more than three high risk or unclear risk of bias domains.

fi. There was low statistical heterogeneity (Heterogeneity: Tau² = 0.074; Chi² = 3.45, df = 3, p = 0.33; I² = 13%, 0% - 87%). We did not downgrade.

fj. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

fk. Meta-analysis: Choko, 2019a had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore all intervention arms were combined; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting); Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but presented sub-groups separately. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

fl. 7 individual randomized trials and 6 cluster randomized trials.

fm. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 5 trials or unclear risk of detection bias in 4 trials, and attrition bias in 2 trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

fn. There was low statistical heterogeneity (Heterogeneity: Tau² = 0.056; Chi² = 18.08, df = 15, p = 0.26; I² = 17%, 0% - 54%). We did not downgrade but noted that outcome definition was not consistent across trials: 5 of 14 trials defined positivity as confirmed HIV positive diagnosis; the remaining 9 trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear.

fo. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

fp. Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore intervention arms were combined. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

fq. 4 individual randomized trials.

fr. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in one trial and attrition bias in one trial (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm).

fs. There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 1.20, df = 3, p = 0.75; I² = 0%, 0% - 62%). We did not downgrade but noted outcome definition was not consistent across trials: confirmed HIV positive diagnosis in three trials and self-reported positivity in one trial.

ft. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

fu. Meta-analysis: Wray, 2018 had more than one intervention arm which included HIVST with similar HIVST kit distribution method, therefore intervention arms were combined.

fv. 4 individual randomized trials.

fw. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 2 trials and attrition bias in two trials (MacGowan, 2017: 27.1% LTFU in the intervention arm and 28.5% in the control arm; Merchant, 2018: 38.4% LTFU overall, 26% in the intervention 50% in the control arm). 3 of 4 trials had more than three high risk or unclear risk of bias domains.

fx. There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 0.88, df = 3, p = 0.83; I² = 0%, 0% - 48%). We did not downgrade but noted that outcome definition was not consistent across trials: 2 of 4 trials defined positivity as confirmed HIV positive diagnosis; the remaining 2 trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear.
fy. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

dz. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting) and only HIVST arm with facility-based HIVST distribution is included in this analysis. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

ga. 2 individual randomized and 2 cluster randomized trials.

gb. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and detection bias (self-reported or non-validated outcomes) in 2 trials.

gc. There was low statistical heterogeneity (Heterogeneity: $\text{Tau}^2 = 0.107; \text{Chi}^2 = 4.25, \text{df} = 3, p = 0.24; I^2 = 29\%, 0\% - 74\%$). We did not downgrade but noted that outcome definition was not consistent across trials: 2 of 4 trials defined positivity as confirmed HIV positive diagnosis; the remaining 2 trials reported reactive HIVST results as positive, relied on self-reported positivity or the definition was unclear.

gd. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

gf. One individual randomized and 2 cluster randomized trials.

gg. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and unclear risk of attrition bias in one trial. Two cluster randomized trials were subject to recruitment bias and one unclear risk for loss of clusters and not reporting cluster-adjusted analysis (we adjusted for clustering in meta-analysis).

gh. There was low statistical heterogeneity (Heterogeneity: $\text{Chi}^2 = 1.60, \text{df} = 2, p = 0.45; I^2 = 0\%, 0\% - 87\%$). We did not downgrade.

gi. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

gj. Meta-analysis: Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but only included ANC group in this analysis. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

gk. One individual randomized and one cluster randomized trial.

gl. We downgraded twice. This was due to potential for performance bias (lack of blinding) in both trials and detection bias (self-reported or non-validated outcomes) in one trial. One trial unclear risk of selection bias (random sequence generation and allocation concealment). One cluster randomized trials were subject to recruitment bias. Each of the trials had more than three high risk or unclear risk of bias domains.

gm. There was low statistical heterogeneity (Heterogeneity: $\text{Chi}^2 = 0.061; \text{df} = 1, p = 0.24$; $I^2 = 26\%$). We did not downgrade.

gn. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

go. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined and only HIVST arm with direct distribution to peers was included in this analysis (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

gp. 2 cluster randomized trials.

gq. We downgraded once. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in both trials.

gr. There was low statistical heterogeneity (Heterogeneity: $\text{Chi}^2 = 0.27, \text{df} = 1, p = 0.61; I^2 = 0\%$). We did not downgrade.
gs. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

gt. 1 cluster randomized trials.

gu. We downgraded once. This was due to potential for performance bias (lack of blinding) and unclear risk of detection bias (self-reported or non-validated outcomes).

gv. Single trial - we downgraded once.

gw. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

gx. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

gy. 3 individual randomized and one cluster randomized trial.

gz. We downgraded twice. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 2 trials or unclear risk of detection bias in 2 trials. Each of the trials had more than three high risk or unclear risk of bias domains.

ha. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0.667$; $\chi^2 = 4.39$, df = 3, $p = 0.22$; $I^2 = 32\%$, 0% - 76%). We did not downgrade.

hb. We downgraded once because confidence intervals around pooled effect estimate are wide, and likely due to small number of events.

hc. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

hd. This outcome is only relevant for intervention (HIVST) arm and reported in 9 trials. Results from 7 trials were pooled and reported as pooled percentage. The remaining two trials did not report data usable for pooled analysis.

he. Three individual randomized trials and 6 cluster randomized trials.

hf. We downgraded once. This was due to potential for performance bias (lack of blinding) in 8 trials, detection bias (self-reported or non-validated outcomes) in 3 trials and attrition bias in one trial. Two cluster randomized trials were subject to recruitment bias and one loss of clusters. Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

hg. There was high statistical heterogeneity (Heterogeneity: $I^2 = 77\%$, $p < 0.01$). The measurement time-point varied (range: 2 weeks - 5 months) and point estimate from individual studies also varied (25% - 76%). We downgraded once.

hh. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with different linkage interventions - we combined arms with similar linkage interventions; Choko, 2019b had two intervention arms which included HIVST and different linkage intervention - we did not combine arms; Shahmanesh, 2019 had two different interventions - we did not combine arms; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST with different HIVST kit distribution method but no linkage intervention, we combined arms (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

hi. 5 individual randomized and 7 cluster randomized trials.

hj. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 5 trials or unclear risk of detection bias in one trial, and unclear attrition bias in one trial. Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

hk. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0$; $\chi^2 = 13.74$, df = 15, $p = 0.56$; $I^2 = 0\%$, 0% - 47%). Sub-group analysis showed heterogeneity was explained by population type. We did not downgrade but noted that outcome definition was not consistent across trials: 7 trials defined linkage as ART initiation whereas other 5 defined it as linkage to any HIV care.

hl. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.
hm. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with different linkage interventions - we combined arms with similar linkage interventions; Choko, 2019b had two intervention arms which included HIVST and different linkage intervention - we did not combine arms; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST with different HIVST kit distribution method but no linkage intervention, we combined arms (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

hn. 5 individual randomized and 6 cluster randomized trials.

ho. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in 5 trials or unclear risk of detection bias in one trial, and unclear attrition bias in one trial. Three cluster randomized trials were subject to recruitment bias. Several risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

hp. There was low statistical heterogeneity (Heterogeneity: Tau² = 0.011; Chi² = 15.71, df = 13, p = 0.29; I² = 15%, 0% - 53%). Sub-group analysis showed heterogeneity was explained by population type. We did not downgrade but noted that outcome definition was not consistent across trials: 7 trials defined linkage as ART initiation whereas other 5 defined it as linkage to any HIV care.

hq. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

hr. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with different linkage interventions - we combined arms with similar linkage interventions; Choko, 2019b had two intervention arms which included HIVST and different linkage intervention - we did not combine arms (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

hs. 2 individual randomized trials and 4 cluster randomized trials

ht. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and detection bias (self-reported or non-validated outcomes) in one trial or unclear risk of detection bias in one trial. Three cluster randomized trials were subject to recruitment bias. Three trials had unclear risk of selection bias (one random sequence generation and 3 allocation concealment) and some other risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

hu. There was low statistical heterogeneity (Heterogeneity: Tau² = 0.018; Chi² = 11.29, df = 8, p = 0.19; I² = 29%, 0% - 67%). We did not downgrade but noted that the outcome definition was not consistent across trials: some trials defined linkage as ART initiation whereas others defined it as linkage to any HIV care.

hv. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

hw. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

hx. 3 individual randomized trials and two cluster randomized trials.

hy. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and detection bias (self-reported or non-validated outcomes) in 2 trials.

hz. There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 3.24, df = 5, p = 0.52; I² = 0%, 0% - 74%). We did not downgrade but noted that the outcome definition was not consistent across trials: some trials defined linkage as ART initiation whereas others defined it as linkage to any HIV care.

ia. We did not downgrade but noted that trials were conducted among men who have sex with men, transgender women and female sex workers, so results should be viewed with caution when applying to other key populations.

ib. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ic. 3 individual randomized trials.

id. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials.
ie. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0; \chi^2 = 0.27, df = 2, p = 0.87; I^2 = 0\% - 24\%$). We did not downgrade.

if. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ig. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

ih. 2 cluster randomized trials.

ii. We downgraded once. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in both trials.

ij. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0; \chi^2 = 0.63, df = 1, p = 0.43; I^2 = 0\%$). We did not downgrade.

ik. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

il. 3 studies were conducted among MSM, 3 among male partners of women, and for 2 sub-group analysis for men was included. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with different linkage interventions - we combined all intervention arms for this analysis; Choko, 2019b had two intervention arms which included HIVST and had two sub-groups (ANC women and ART clients) - we combined HIVST arms but only included ANC group in this analysis. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trial.

im. 5 individual randomized trials and 3 cluster randomized trials.

in. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, and detection bias (self-reported or non-validated outcomes) in one trial or unclear risk of detection bias in one trial. Three cluster randomized trials were subject to recruitment bias. Some other risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts. Each of the trials had more than three high risk or unclear risk of bias domains.

io. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0.045; \chi^2 = 9.30, df = 7, p = 0.23; I^2 = 25\% - 66\%$). We did not downgrade.

ip. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

iq. 2 trials were conducted among female sex workers and for 2 trials sub-group analysis among women was included. Meta-analysis: Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

ir. One individual randomized and two cluster randomized trials.

is. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and detection bias (self-reported or non-validated outcomes) in 2 trials or unclear risk of detection bias in one trial. Two trials had unclear risk of selection bias (one random sequence generation, two allocation concealment) and some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

it. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0; \chi^2 = 2.48, df = 3, p = 0.48; I^2 = 0\% - 81\%$). We did not downgrade.

iu. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

iv. For both trials sub-analysis among young people was included in the analysis. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

iw. One individual randomized and once cluster randomized trial.

ix. We downgraded twice. This was due to potential for performance bias (lack of blinding) in both trials and unclear risk of selection bias (one random sequence generation and two allocation concealment). One cluster randomized trial was subject to recruitment bias. Each of the trials had more than three high risk or unclear risk of bias domains.
iy. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0$; $\chi^2 = 0.26$, $df = 1$, $p = 0.61$; $I^2 = 0\%$). We did not downgrad.

iz. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ja. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with different linkage interventions - we combined arms with similar linkage interventions; Choko, 2019b had two intervention arms which included HIVST and different linkage intervention - we did not combine arms; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

jb. 3 individual randomized and 5 cluster randomized trials.

cj. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, and detection bias (self-reported or non-validated outcomes) in 4 trials or unclear risk of detection bias in one trial. Three cluster randomized trials were subject to recruitment bias. Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

jd. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0.034$; $\chi^2 = 17.74$, $df = 10$, $p = 0.14$; $I^2 = 32\%$, 0\% - 67\%). We did not downgrad.

ej. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

jf. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

jg. 2 individual randomized trials and one cluster randomized trial.

jh. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and unclear risk of allocation concealment (selection bias) in one trial.

ji. There was low statistical heterogeneity (Heterogeneity: $\tau^2 = 0$; $\chi^2 = 0.46$, $df = 2$, $p = 0.79$; $I^2 = 0\%$, 0\% - 55\%). We did not downgrad.

jj. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

jk. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with different linkage interventions - we only included the arm with no linkage intervention in this analysis; Choko, 2019b had two intervention arms which included HIVST and different linkage intervention - we only included the arm with no linkage intervention in this analysis; Chanda, 2017 and Ortblad, 2017 had more than one intervention arm which included HIVST but HIVST kit distribution method was different, therefore intervention arms were not combined (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

jl. 4 individual randomized trials and 5 cluster randomized trials.

jm. We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials and detection bias (self-reported or non-validated outcomes) in 3 trials or unclear risk of detection bias in one trial. Three cluster randomized trials were subject to recruitment bias. Two trials were unclear risk of selection bias (one random sequence generation and two allocation concealment) and some other risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

jn. There was moderate statistical heterogeneity (Heterogeneity: $\tau^2 = 0.058$; $\chi^2 = 13.03$, $df = 8$, $p = 0.11$; $I^2 = 39\%$, 0\% - 72\%). Heterogeneity can be explained by difference in study populations. We did not downgrad.

jo. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

jp. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with different linkage interventions - we only included the arm with phone reminder for linkage. Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

jq. One cluster randomized trial.

jr. We downgraded once. This was due to potential for performance bias (lack of blinding) and high risk of cluster recruitment bias.
js. Single trial, we downgraded once as inconsistency cannot be evaluated.

jt. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ju. Meta-analysis: Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

jv. 1 cluster randomized trials.

jw. We downgraded once. This was due to potential for performance bias (lack of blinding) in both trials and unclear risk of allocation concealment (selection bias) in one trial.

jx. Single trial - we downgraded once as inconsistency cannot be evaluated.

jy. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

jz. One individual randomized trial.

ka. We downgraded once. This was due to potential for performance bias (lack of blinding).

kb. Single trial, we downgraded once as inconsistency cannot be evaluated.

kc. Confidence intervals around point estimate are wide. We downgraded once.

kd. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

ke. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with different linkage interventions - we combined arms with financial linkage interventions and included in the analysis; Choko, 2019b had two intervention arms which included HIVST and different linkage intervention - we did not combine arms and included the arm with financial linkage intervention in this analysis (control arm sample was adjusted to prevent double counting). Cluster-adjusted analysis was included in meta-analysis for cluster randomized trials.

kf. 2 cluster randomized trials.

kg. We downgraded once. This was due to potential for performance bias (lack of blinding) in both trials and unclear risk of detection bias (self-reported or non-validated outcomes) in one trial. Both cluster randomized trials were subject to recruitment bias. One trial had unclear risk of bias for attrition or loss of clusters and unadjusted cluster analysis (we included cluster adjusted estimates).

kh. There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 0.08, df = 1, p = 0.77; I² = 0%). We did not downgrade.

ki. Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

kj. Two trials reported this outcome which is only relevant to the intervention (HIVST) arm.

kk. Two cluster randomized trials.

kl. We downgraded once. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in both trials. One cluster randomized trials was subject to recruitment bias. Some risk of bias domains were unclear risk due to lack of information from unpublished reports or conference abstracts.

km. The outcomes were not pooled therefore inconsistency cannot be evaluated. Similar outcomes were reported from both included trials.

kn. We did not downgrade but noted that one trial involved facility-based HIVST and one trial involved home-based HIVST distribution by community workers. Both trials were among general population. The results should be viewed with caution when applying to other populations.

ko. The results are based on very few events so caution is needed when interpreting. We downgraded once.
Two trials were conducted involving HIVST distribution by women to their male partners (Choko, 2019a; Masters, 2016), one involved HIVST distribution by HIV-positive clients to their partners (Dovel, 2019), one involved HIVST distribution by both women to their male partners and HIV-positive clients to their partners (Choko, 2019b), and two involved distribution of HIVST kits or coupons by female sex workers to their peers (Chanda, 2017, Ortblad, 2017).

2 individual randomized and 4 cluster randomized trials.

We downgraded once. This was due to potential for performance bias (lack of blinding) and detection bias (self-reported or non-validated outcomes) in all trials. One cluster randomized trials was subject to recruitment bias and one trial had unclear risk of selection bias (random sequence generation and allocation concealment).

There was low statistical heterogeneity (Heterogeneity: Tau² = 0; Chi² = 2.96, df = 5, p = 0.71; I² = 0%, 0% - 57%). We did not downgrade but noted that type of adverse events varied across trials (intimate partner violence: Chanda, 2017, Choko, 2019a, Masters, 2016, Ortblad, 2017; temporary relationship breakdown: Choko, 2019b; verbal abuse: Dovel, 2019).

All trials involved secondary distribution to partners (in Mulubwba, 2019 a subset of all participants distributed to their partners) or peers so right population of interest for this outcome.

Confidence intervals around pooled effect size were large, likely due to very few events. We downgraded once.

Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

All trials were among men who have sex with men.

We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in all trials, and unclear risk of selection bias (allocation concealment) in one trial.

There was low statistical heterogeneity (Heterogeneity: Tau² = 0.002; Chi² = 3.39, df = 3, p = 0.34; I² = 11%, 0% - 86%). We did not downgrade but noted that outcome definitions varied across trials (Jamil, 2017: condomless anal intercourse with casual male partner(s) during 12 months follow-up; Katz, 2018: non-concordant condomless anal intercourse with male partner(s) in the past 3 months measured at 15 months; Tang, 2018: condomless anal sex with male partner(s) in the past 3 months measured at 6 months (mid-point during follow-up); Wang, 2017: condomless anal intercourse with male partner(s) in the past 3 months measured at 6 months).

All trials were among men who have sex with men so results should be viewed with caution when applying to other populations.

Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

All trials were among men who have sex with men.

We downgraded once. This was due to potential for performance bias (lack of blinding) in all trials, detection bias (self-reported or non-validated outcomes) in all trials, and unclear risk of selection bias (allocation concealment) in one trial.

There was low statistical heterogeneity (Heterogeneity: Tau² = 0.0003; Chi² = 3.06, df = 3, p = 0.38; I² = 2%, 0% - 85%). We did not downgrade but noted that outcome definitions varied across trials (Jamil, 2017: condomless anal intercourse with casual male partner(s) during 12 months follow-up; Katz, 2018: non-concordant condomless anal intercourse with male partner(s) in the past 3 months measured at 15 months; Tang, 2018: condomless anal sex with male partner(s) in the past 3 months measured at 6 months (mid-point during follow-up); Wang, 2017: condomless anal intercourse with male partner(s) in the past 3 months measured at 6 months).

All trials were among men who have sex with men so results should be viewed with caution when applying to other populations.

Original sample size, does not represent effective cluster adjusted sample size in the meta-analysis.

Web Annex B
### Table B2. GRADE table for HIVST + linkage intervention compared to HIVST only for linkage to HIV CARE or ART

**Author(s):** Muhammad S. Jamil  
**Question:** HIVST + linkage intervention compared to HIVST only for linkage to HIV care or ART

<table>
<thead>
<tr>
<th>Certainty assessment</th>
<th>Nº of patients</th>
<th>Effect</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nº of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td><strong>Linkage to HIV care or ART initiation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>randomised trials</td>
<td>serious</td>
<td>serious</td>
<td>not serious</td>
</tr>
</tbody>
</table>

**CI:** Confidence interval; **RR:** Risk ratio

**Explanations**

a. Meta-analysis: Choko, 2019a had multiple intervention arms which included HIVST with financial incentive for linkage - we combined these arms (HIVST only arm sample was adjusted to prevent double counting); Choko, 2019b: HIVST + financial incentive for linkage vs. HIVST only; MacPherson, 2014: home-based HIVST + optional home ART initiation vs. HIVST + standard linkage; Nichols, 2019: HIVST arm had two linkage interventions (HIVST + self-referral and HIVST + in-person referral/escort); we compared HIVST + in-person referral/escort vs. HIVST + self-referral; Sibanda, 2019 had two intervention arms (HIVST + fixed provider financial incentive and HIVST + fixed provider incentive and conditional incentive per linkage); we compared HIVST + fixed and conditional incentive vs. HIVST + fixed incentive. Cluster-adjusted analysis was included in meta-analysis cluster randomized trials.

b. 5 cluster randomized trials.

c. We downgraded once. This was due to potential for performance bias (lack of blinding) in 4 trials or unclear risk of performance bias in one trial, high risk of detection bias (self-reported or non-validated outcomes) in one trial. Four cluster randomized trials were subject to recruitment bias (high or unclear risk). Some risk of bias domains were unclear due to limited information from conference abstracts.

d. There was high statistical heterogeneity (Heterogeneity: Tau² = 0.097; Chi² = 21.75, df = 5, p <0.01; I² = 77%, 49% - 90%). We downgraded once and noted that study design, type of linkage intervention and comparison group varied across studies.

**For ePub**