

# Online appendix J

## Public health sector risk of bias assessment report

Table 1. Risk of bias assessment in the public health sector

ITEM	1A. STUDY DESIGN, END OF INTERVENTION (POTENTIAL CONFOUNDERS TAKEN INTO ACCOUNT)	4. INTERVENTION IS CLEARLY DEFINED	5. OUTCOME MEASURES ARE CLEARLY DEFINED AND RELIABLE	6. BASELINE BALANCE (N.A. FOR BEFORE VERSUS AFTER)	7. REPRESENTATIVENESS OF A LARGE SCALE INTERVENTION	8. PRECISION OF ESTIMATE (IN CASE OF REGRESSION)	OVERALL CONFIDENCE IN STUDY (END OF INTERVENTION)	STUDY QUALITY
<b>ABDU (2004)</b> <b>(ID:49952461)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-Medium confidence: Brief description of intervention	-Medium confidence: Outcome named but not described	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>ADITYA (2019)</b> <b>(ID:55552141)</b>	-High confidence: RCT, RDD, instrumental	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with	-High confidence: RCT or baseline balance report and satisfactory	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation

	variable under LATE		reference to validation	(imbalance on 5 or less than 5 percent)				
<b>AKRESH (2012) (ID:49956624)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>ANDRADE (2012) (ID:49947144)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline variables (except RCT)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Low confidence: Standard errors are not clustered	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>ARNOLD (2009) (ID:49924806)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation		-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>AZIZ (2018) (ID:49924807)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>BANERJEE (2010) (ID:49924808)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>BIRAN (2014) (ID:49924810)</b>	-High confidence: RCT, RDD,	-High confidence: Intervention	-High confidence: Outcome measure clearly and fully	-High confidence: RCT or baseline balance report and	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard	-High confidence: High rating across ALL	-High quality

	instrumental variable under LATE	clearly and fully described	described, preferably with reference to validation	satisfactory (imbalance on 5 or less than 5 percent)		errors are clustered by intervention design level	of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	impact evaluation
<b>BOONE (2017) (ID:49957276)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>BOUDREAUX (2014) (ID:49924811)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline variables (except RCT)	-Medium confidence: At least 7b is answered with a "YES" and 7c is not answered as a "NO"	-Medium confidence: Standard errors are clustered but clear at what level	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>BRICENO (2017) (ID:49941491)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline variables (except RCT)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Medium confidence: Standard errors are clustered but clear at what level	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>BROWN (2009) (ID:49956643)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Low confidence: Standard errors are not clustered	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>BRUGH (2018) (ID:49949102)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation

<b>BRUNIE (2014) (ID:49937979)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Low confidence: Standard errors are not clustered	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>BUSZA (2019) (ID:49943710)</b>	-Low confidence: Other matching, DID alone, instrumental variable otherwise	-High confidence: Intervention clearly and fully described	-Medium confidence: Outcome named but not described	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline variables (except RCT)	-Low confidence: None is a "YES" (7a,7b,7c)	-N/A for studies where no clusters are mentioned	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>CAHYADI (2018) (ID:49954192)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>CAMERON (2013) (ID:49924813)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>CARVALHO (2014) (ID:49938716)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline variables (except RCT)	-Low confidence: None is a "YES" (7a,7b,7c)	-High confidence: Standard errors are clustered by intervention design level	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>CHANKOVA (2012) (ID:49937693)</b>	-Low confidence: Other matching, DID	-High confidence: Intervention	-High confidence: Outcome measure clearly and fully described,	-Low confidence: Baseline balance not reported or reported and lack of balance	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are	-Medium confidence: Medium rating across any of the	-Medium quality impact evaluation

	alone, instrumental variable otherwise	clearly and fully described	preferably with reference to validation	on 10 Or more than 10% of baseline variables (except RCT)		clustered by intervention design level	items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	
<b>CHOULAGAI (2017) (ID:49937949)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>CHRISTINE (2012) (ID:55552142)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>CORBETT (2007) (ID:49943298)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>CROCKER (2016) (ID:49938339)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>CROCKER (2016) (ID:49938566)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation

<b>EDMOND (2019) (ID:49947369)</b>	-Low confidence: Other matching, DID alone, instrumental variable otherwise	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>EZEANOLUE (2015) (ID:49944815)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>EZEANOLUE (2017) (ID:49939180)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>FILMER (2018) (ID:49938940)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>FREEMAN (2016) (ID:49957189)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>GALIANI (2012) (ID:49938570)</b>	-High confidence: RCT, RDD, instrumental	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with	-High confidence: RCT or baseline balance report and satisfactory	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by	-High confidence: High rating across ALL of the items 1a,	-High quality impact evaluation

	variable under LATE		reference to validation	(imbalance on 5 or less than 5 percent)		intervention design level	4, 5, 6, 7 and 8 (or N/A in 8)	
<b>GELDSETZER (2019) (ID:49946415)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>GUIERAS (2015) (ID:49924835)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>HARTER (2019) (ID:49948030)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Low confidence: Standard errors are not clustered	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>HEMMINKI (2013) (ID:49943978)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>HODDINOTT (2004) (ID:49952190)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported on 10 Or more than 10% of baseline variables (except RCT)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>HUDA (2012) (ID:49956996)</b>	-Low confidence: Other	-High confidence: Intervention	-High confidence: Outcome measure clearly and fully	-High confidence: RCT or baseline balance report and	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no	-Low confidence: Low rating across any	-Low quality

	matching, DID alone, instrumental variable otherwise	clearly and fully described	described, preferably with reference to validation	satisfactory (imbalance on 5 or less than 5 percent)		clusters are mentioned	of the items 1a, 4, 5, 6, 7 and 8	impact evaluation
<b>JAFAR (2009)</b> <b>(ID:49924843)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Low confidence: Standard errors are not clustered	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>KIRKWOOD (2013)</b> <b>(ID:49945873)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>KUSUMA (2017)</b> <b>(ID:49944134)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>LAMICHHANE (2017)</b> <b>(ID:49947300)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline variables (except RCT)	-Low confidence: None is a "YES" (7a,7b,7c)	-High confidence: Standard errors are clustered by intervention design level	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>LEE (2019)</b> <b>(ID:49956513)</b>	-Low confidence: Other matching, DID alone, instrumental	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation

	variable otherwise			variables (except RCT)				
<b>LUBY (2009)</b> <b>(ID:49954979)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>LV (2014)</b> <b>(ID:49942798)</b>	-Low confidence: Other matching, DID alone, instrumental variable otherwise CBA	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Medium confidence: Imbalance in 5-10 percent baseline variables	-Low confidence: None is a "YES" (7a,7b,7c)	-N/A for studies where no clusters are mentioned	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>MARQUIS (2018)</b> <b>(ID:49952370)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Medium confidence: Standard errors are clustered but clear at what level	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>MEMON (2015)</b> <b>(ID:49952356)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>MIDHET (2010)</b> <b>(ID:49946648)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Medium confidence: Standard errors are clustered but clear at what level	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation

<b>MILLER (2011) (ID:49947999)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-Low confidence: None is a "YES" (7a,7b,7c)	-N/A for studies where no clusters are mentioned	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>MORRIS (2004) (ID:49943481)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>NJUGUNA (2019) (ID:49952247)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>ODENY (2019) (ID:49943982)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>PARVEZ (2018) (ID:49945881)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>PATIL (2013) (ID:49944874)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation

			reference to validation					
<b>POWELL-JACKSON (2012) (ID:49949023)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Medium confidence: Standard errors are clustered but clear at what level	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>QUAYYUM (2013) (ID:49952311)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>SOLOMON (2019) (ID:49944782)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>STOLLER (2011) (ID:49945153)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>TIAN (2019) (ID:49945485)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>TIRUNEH (2020) (ID:49943435)</b>	-High confidence: RCT, RDD, instrumental	-High confidence: Intervention	-High confidence: Outcome measure clearly and fully described,	-High confidence: RCT or baseline balance report and satisfactory	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no	-High confidence: High rating across ALL of the items 1a,	-High quality impact evaluation

	variable under LATE	clearly and fully described	preferably with reference to validation	(imbalance on 5 or less than 5 percent)		clusters are mentioned	4, 5, 6, 7 and 8 (or N/A in 8)	
<b>TRIYANA (2017) (ID:49946317)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>TUTOR (2014) (ID:49943252)</b>	-Medium confidence: DiD with matching, PSM	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline variables (except RCT)	-Low confidence: None is a "YES" (7a,7b,7c)	-Medium confidence: Standard errors are clustered but clear at what level	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation
<b>WANG (2015) (ID:49946378)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-N/A for studies where no clusters are mentioned	-High confidence: High rating across ALL of the items 1a, 4, 5, 6, 7 and 8 (or N/A in 8)	-High quality impact evaluation
<b>WICH AidIT (2019) (ID:49949514)</b>	-High confidence: RCT, RDD, instrumental variable under LATE	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-High confidence: RCT or baseline balance report and satisfactory (imbalance on 5 or less than 5 percent)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-Medium confidence: Standard errors are clustered but clear at what level	-Medium confidence: Medium rating across any of the items 1a, 4, 5, 6, 7 and 8 (and NO low rating)	-Medium quality impact evaluation
<b>ZHOU (2020) (ID:49946450)</b>	-Low confidence: Other matching, DID alone, instrumental variable otherwise	-High confidence: Intervention clearly and fully described	-High confidence: Outcome measure clearly and fully described, preferably with reference to validation	-Low confidence: Baseline balance not reported or reported and lack of balance on 10 Or more than 10% of baseline variables (except RCT)	-High confidence: All three (7a,7b & 7c) answered as a "YES"	-High confidence: Standard errors are clustered by intervention design level	-Low confidence: Low rating across any of the items 1a, 4, 5, 6, 7 and 8	-Low quality impact evaluation