The effectiveness of training strategies to improve healthcare provider practices in low-income and middle-income countries

Alexander K Rowe 1, Samantha Y Rowe 2, David H Peters 3, Kathleen A Holloway 4, 5, Dennis Ross-Degnan 6

ABSTRACT

Introduction In low/middle-income countries (LMICs), training is often used to improve healthcare provider (HCP) performance. However, important questions remain about how well training works and the best ways to design training strategies. The objective of this study is to characterise the effectiveness of training strategies to improve HCP practices in LMICs and identify attributes associated with training effectiveness.

Methods We performed a secondary analysis of data from a systematic review on improving HCP performance. The review included controlled trials and interrupted time series, and outcomes measuring HCP practices (eg, percentage of patients correctly treated). Distributions of effect sizes (defined as percentage-point (% point) changes) were described for each training strategy. To identify effective training attributes, we examined studies that directly compared training approaches and performed random-effects linear regression modelling.

Results We analysed data from 199 studies from 51 countries. For outcomes expressed as percentages, educational outreach visits (median effect size when compared with controls: 9.9 %points; IQR: 4.3–20.6) tended to be somewhat more effective than in-service training (median: 7.3 %points; IQR: 3.6–17.4), which seemed more effective than peer-to-peer training (4.0 %points) and self-study (by 2.4–3.2 %points). Mean effectiveness was greater (by 6.0–0.4 %points) for training that incorporated clinical practice and training at HCPs’ work site. Attributes with little or no effect were: training with computers, interactive methods or over multiple sessions; training duration; number of educational methods; distance training; trainers with pedagogical training and topic complexity. For lay HCPs, in-service training had no measurable effect.

Conclusions Although additional research is needed, by characterising the effectiveness of training strategies and identifying attributes of effective training, decision-makers in LMICs can improve how these strategies are selected and implemented.

INTRODUCTION

Healthcare providers (HCPs), including facility-based and community-based health workers, play essential roles in delivering healthcare. However, hundreds of studies have documented inadequate HCP performance in low/middle-income countries (LMICs).1, 2 Estimates of the consequences of poor-quality care range between 4.9–5.2 million1 3 and 5.7–8.4 million1 4 deaths annually.

Strategies to train HCPs are often used to improve performance. For example, from 1988 to 1993, to promote oral rehydration therapy, more than 2000 training courses were held on managing diarrhoea in more than 120 countries.4 From 2011 to 2015, US$95.3 million in country proposals were approved by Gavi for health workforce development and, in particular, training and capacity building for HCPs.

To cite: Rowe AK, Rowe SY, Peters DH, et al. The effectiveness of training strategies to improve healthcare provider practices in low-income and middle-income countries. BMJ Global Health 2021;6:e003229. doi:10.1136/bmjgh-2020-003229

Handling editor Kerry Scott

Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmjgh-2020-003229).

Received 23 June 2020
Revised 20 October 2020
Accepted 10 November 2020
in LMICs (personal communication, Alan Brooks, Gavi, 17 November 2015). From 2017 to 2019, all 72 country-specific annual plans of the President’s Malaria Initiative funded training. For the 2018–2020 funding cycle, grant recipients of The Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM) have budgeted US$510 million for training (personal communication, Benjamin Loevinsohn, GFATM, 15 December 2019).

Despite widespread use, important questions remain about training effectiveness and the best ways to design training strategies in LMICs. A large systematic review of studies from LMICs (the Health Care Provider Performance Review (HCPPR)) found that training was the most often studied strategy, and its effect on HCP practices varied substantially, from −19.9 to 60.8 percentage-points (%-points) (median: 10.3, IQR: 6.1–20.7).1 However, that analysis combined multiple strategies (eg, group training, self-study and peer-to-peer training) and implementation approaches (eg, in terms of educational methods and training duration) into a single heterogeneous group. Other reviews that examined training effectiveness or sought to identify best practices for designing training courses have key limitations, such as providing only non-quantitative, narrative summaries (which obscures the magnitude of effect of recommended design attributes), or including few studies from LMICs.8–12

This report presents a secondary analysis of HCPPR data to characterise the effectiveness of different training strategies (objective 1) and identify attributes associated with group in-service training effectiveness (and thus suggest ways to increase training effectiveness) (objective 2). The results can inform the choice and design of future training strategies and shape the research agenda on improving training effectiveness.

**METHODS**

This report uses the same methods as those used in an HCPPR-based analysis of supervision strategies.13 We analysed data from the HCPPR (PROSPERO registration CRD42016046154). Details of the HCPPR’s inclusion criteria, literature search, data abstraction, risk of bias assessment, effect size estimation and assessment of publication bias have been published elsewhere.14 A summary of those methods and additional details particular to this report are presented below and in online supplemental appendix I (section 1).

**Inclusion criteria**

The HCPPR included published and unpublished studies from LMICs in the public and private sectors that quantitatively evaluated a strategy to improve HCP performance. HCPs were broadly defined as hospital-based, health facility-based or community-based health workers; pharmacists and shopkeepers who sell medicines. Eligible study designs included pre-intervention versus post-intervention studies with a randomised or non-randomised comparison group, post-intervention only studies with a randomised comparison group and interrupted time series (ITS). ITS studies could have a randomised or non-randomised control group, or no separate control group (see online supplemental appendix 1, table B1). The HCPPR included studies on any health condition and there were no language restrictions. For outcomes expressed as a percentage, effect sizes with baseline values of 95% or greater were excluded, as there was little room for improvement.

For this report, we only included studies that tested strategies with a component to train HCPs, although many strategies also had other intervention components. Additionally, we only analysed HCP practice outcomes (eg, patient assessment, diagnosis, treatment, counseling, referral, documentation and consultation time). This outcome category had the largest number of studies in the HCPPR.

**Literature search and data abstraction**

The HCPPR searched 52 electronic databases for published studies and 58 document inventories for unpublished studies from the 1960s to 2016. Personal libraries were screened, experts were queried for unpublished studies and hand searches were performed of bibliographies from previous reviews.

To identify eligible reports, titles and abstracts were screened, and when necessary, a report’s full text was reviewed. Data were abstracted independently by two investigators or research assistants using a standardised form. Discrepancies were resolved through discussion. Data were entered into a database (Microsoft Access, Microsoft, Redmond, Washington, USA). Data elements included HCP type, improvement strategies, outcomes, effect sizes and risk of bias domains. Study investigators were queried about details not available in study reports. The data and an associated data dictionary are publicly available at http://www.HCPperformanceview.org.

**Strategy definitions**

In the HCPPR database, the presence of 207 strategy components was coded for each study arm exposed to an improvement strategy, and the components were grouped into categories. For HCP training strategies, five specific categories were created (eg, group in-service training) (box 1, top part). Eleven other categories were more general (eg, supervision, group problem-solving). Placebo strategy components were analysed together with control groups that received no new intervention.

**Risk of bias assessment**

Risk of bias at the study level was categorised as low, moderate, high or very high. Randomised studies, ITS, and non-randomised studies were initially categorised as low, moderate and high risk of bias, respectively. Risk of bias domains (eg, dataset completeness, balance in baseline outcome measurements and so on) were then assessed. A study’s risk of bias category was dropped by
Training strategies for healthcare providers (HCPs) (all categories are mutually exclusive)

1. Group in-service training. On-the-job training primarily led by a facilitator, typically for a group of HCPs in a classroom or clinical setting.*
2. Group preservice training. Training for HCPs before they begin practicing, primarily led by a facilitator, typically for a group of HCPs in a classroom or clinical setting.*
3. Self-study in-service training. On-the-job training with structured sessions in which HCPs study by themselves without direct supervision or attendance in a class. HCPs might occasionally interact with a facilitator, supervisor or peer to discuss the training content.†
4. Educational outreach visits. On-the-job training strategy with face-to-face visits to individual HCPs at their workplace by persons who HCPs regard as an expert or opinion leader to promote best practices. Also known as academic detailing.
5. Peer-to-peer training. On-the-job training led by HCP peers usually at the HCPs’ workplace. For example, an HCP (eg, a facility in-charge) attends a training course and then returns to his/her clinic and shares the training information with other HCPs. Also known as peer education. This strategy is different from educational outreach visits because peer-to-peer training does not involve visits by an external expert or opinion leader.

Other strategy components (all categories are mutually exclusive)‡

1. Community support, for example, community health education or social marketing of health services.
2. Patient support, for example, patient health education via printed materials or home visits.
3. Strengthening infrastructure, for example, provision of medicines or equipment.
4. HCP-directed financial incentives, for example, performance-based payments.
5. Health system financing and other incentives, for example, insurance or reducing a consultation fee.
6. Regulation and governance, for example, accreditation system.
7. Group problem-solving, for example, collaborative improvement or group problem-solving with or without formal teams.
8. Supervision, for example, improving routine supervision, benchmarking or audit with feedback.
9. Other management techniques that do not include group problem-solving and supervision, for example, HCP self-assessment or HCP group process that is not group problem-solving.
10. Printed information or job aid for HCPs that is not an integral part of another component, for example, pamphlet for HCP.§
11. Information and communication technology (includes mHealth and eHealth) for HCPs, for example, computerised decision aid or text message reminders sent to HCPs’ phones.

*Trainees might spend some time working individually to learn the material (eg, reading to prepare for the next day’s class), however such activity was not considered self-study.
†Does not include: (1) educational materials given to HCPs before group in-service training that HCPs are asked to review to prepare for that training (such materials are considered an integral part of the group training, rather than self-study), or (2) educational materials given to HCPs without specific instructions on how HCPs should learn them (eg, distributing a printed treatment manual with the instruction that HCPs should use it, but without further guidance).
‡Does not include: (1) educational materials given to HCPs before group in-service training that HCPs are asked to review to prepare for that training (such materials are considered an integral part of the group training, rather than self-study), or (2) educational materials given to HCPs without specific instructions on how HCPs should learn them (eg, distributing a printed treatment manual with the instruction that HCPs should use it, but without further guidance).
§Other strategy components (especially training) often include printed information for HCPs; and in these cases, the printed information was not considered a separate component.

Estimating effect sizes

The primary outcome measure was the effect size, which was defined as an absolute %-point change and calculated such that positive values indicate improvement. For study outcomes that decreased to indicate improvement (eg, percentage of patients receiving unnecessary treatments), we multiplied effect sizes by –1. For non-ITS studies, effect sizes for outcomes expressed as a percentage (eg, percentage of patients treated correctly) were calculated with equation 1 (if there were baseline values) or equation 2 (if no baseline values). Effect sizes were based on the baseline value closest in time to the beginning of the strategy and the follow-up value furthest in time from the beginning of the strategy.

\[
\text{Effect size} = \left( \frac{\text{follow-up} - \text{baseline}}{\text{baseline}} \right)_{\text{intervention}} - \left( \frac{\text{follow-up} - \text{baseline}}{\text{baseline}} \right)_{\text{control}}
\]

In non-ITS studies, effect sizes for unbounded continuous outcomes (eg, number of medicines prescribed per patient) were calculated with equation 3 (if there were baseline values) or equation 4 (if no baseline values). For unbounded continuous outcomes, if the baseline value for either the intervention or control group equaled zero (when equation 3 was used) or if the follow-up value for the control group equaled zero (when equation 4 was used), the effect size was undefined and thus excluded. Note that such exclusions were rare (<2%).

\[
\text{Effect size} = 100 \times \left( \frac{\text{follow-up} - \text{baseline}}{\text{baseline}} \right)_{\text{intervention}} - \left( \frac{\text{follow-up} - \text{baseline}}{\text{baseline}} \right)_{\text{control}}
\]

For ITS studies, segmented linear regression modelling was performed to estimate a summary effect size that incorporated both level and trend effects. The summary effect size was the outcome level at the midpoint of the follow-up period as predicted by the regression model minus a counterfactual value that equaled the outcome level based on the pre-intervention trend extended to the midpoint of the follow-up period.

Analysis

For objective 1 (characterise the effectiveness of training strategies), we analysed five types of study comparisons (box 2). To estimate strategy effectiveness, the effect size for each study comparison was defined as the median of all effect sizes (MES) within the comparison. That is, results of multiple outcomes from the same study comparison were collapsed into a single
Box 2  Study comparisons used to characterise the effectiveness of different training strategies (objective 1)

| Non-equivalency studies (success is an effect size with a large magnitude) |
|--------------------------------------------------------------------------|
| ► Comparison of a training strategy* alone† versus a (no-intervention) control group. |
| ► Comparison of one training strategy* alone† versus a different training strategy* alone.† |
| ► Comparison of a training strategy* combined with a specific group of other strategy components versus that same specific group of other strategy components‡ (eg, ‘educational outreach visits plus supervision’ vs ‘supervision’). |
| ► Comparison of one training strategy* combined with a specific group of other strategy components‡ versus a different training strategy* combined with that same specific group of other strategy components‡ (eg, ‘educational outreach visits plus supervision’ vs ‘group in-service training plus supervision’). |

| Equivalency studies (success is an effect size close to zero) |
|-------------------------------------------------------------|
| ► Comparison of a training strategy* alone† versus a ‘gold standard’ comparison group (eg, intervention group of nurse-midwives trained to perform tubal ligation surgery vs a gold standard comparison group of physicians). |

*Any of the five training strategies listed in the top part of box 1. †That is, not combined with other strategy components. Printed materials, which are usually used in training courses, were not considered to be a separate component (eg, the strategy ‘in-service training plus printed guideline’ was considered to be ‘in-service training alone’). ‡One or more of the 11 strategy categories in the bottom part of box 1.

Box 3  Variables in the models used to identify in-service training attributes associated with training effectiveness (objective 2)

| In-service training attributes |
|-------------------------------|
| ► Training duration (in days). |
| ► Training methods (type and number of methods used): lectures, interactive discussions, clinical practice, role-play and other methods. |
| ► Printed materials for healthcare providers used. |
| ► Computers used for at least part of the training. |
| ► Group size. |
| ► Topic complexity (single vs multiple topics). |
| ► Use of professional trainers (ie, with any training in pedagogical methods). |
| ► Use of trainers who were content experts. |
| ► Some or all training was on-site where healthcare providers routinely work. |
| ► Training delivered over one continuous period versus multiple sessions (eg, a 4-day curriculum delivered via four separate 1-day sessions (eg, four Mondays in a row)). |

| Interactions |
|--------------|
| ► Topic complexity×natural logarithm of training duration. |
| ► Time since training was conducted×whether training was combined with supervision. |

| Confounders |
|-------------|
| ► Baseline performance level. |
| ► Time since training was conducted. |

RESULTS

Literature search

The HCPR screened 216483 citations and included 2272 reports (online supplemental appendix 1, figure A). Of those, 384 reports were eligible for this analysis. These reports presented 1200 effect sizes from 240 comparisons in 199 studies (see online supplemental appendix 1, tables A1–A4 for sample size details; and online supplemental appendix 2 for study citations and study-specific details). These studies were conducted in 51 LMICs and represented a diversity of methods, geographical settings, HCP types, work environments, health conditions and hierarchical databases: training alone (N=55 studies), training with or without supervision (N=73), and training with or without any other intervention components (N=152). We restricted this analysis to studies of professional HCPs, training duration <20 days (studies with missing duration included) and percentage outcomes. Additional details are presented in online supplemental appendix 1 (section 1, pages 3–4).

To characterise cost, we analysed group in-service training for professional HCPs, as it was the type tested by the largest number of studies. As studies varied in size, in terms of numbers of HCPs trained and training duration (with longer courses being more expensive), we calculated the cost per HCP per day of training.

All analyses were performed with SAS, V.9.4 (SAS Institute, Inc. Cary, North Carolina, USA).
practices (online supplemental appendix 1, tables B1–B4; and online supplemental appendix 2 for study-specific details). More than half of studies (57.8%) had randomised designs, and 40.2% had a low or moderate risk of bias. Median follow-up time was 4.0 months (from 183 studies that reported follow-up time; IQR: 2.0–7.5), median number of health facilities per study was 16 (from 157 studies; IQR: 6–51) and median number of HCPs per study was 98 (from 125 studies; IQR: 49–167). Most studies (80.4%) were published since 2000. We found no evidence of publication bias (online supplemental appendix 1, figure B).

Effectiveness of training strategies (objective 1)

Table 1 presents the effects of training strategies on the practices of professional HCPs. There are five main findings, all supported by low-quality evidence primarily because many studies had a high risk of bias. Since only four study comparisons involved strategies with computers, they were analysed together with studies without computers.

First, for group in-service training (hereafter referred to as ‘in-service training’), when compared with controls, the median performance improvement ranged from 7.3 to 17.4 %-points (table 1, rows 1–2; figure 1; second and fourth histograms in online supplemental appendix 1, figure D). For example, for a percentage outcome with a typical baseline performance level of 40% and training effect of 7.3 %-points, the post-training performance level would be 47.3%. However, training effects were very heterogeneous: in the largest group of studies (60 comparisons from 55 studies), one-quarter of MES values were small (<3.6 %-points) and one-quarter were large (17.4–68.1 %-points). Effect sizes tended to be lower among higher quality studies, with a median improvement of 5.1 %-points for the 28 study comparisons with a low or moderate risk of bias (table 1, footnotes). The one equivalency study of in-service training compared with a gold standard control had a ‘perfect’ effect size of zero (Box 2). The marginal effect of in-service training when added to other strategy components (medians ranging from –7.3 to 3.7 %-points) was smaller than the effect of in-service training when compared with a no-intervention reference group.

Second, educational outreach visits (EOVs) tended to be somewhat more effective than in-service training, although results of individual studies varied widely. The median effect of EOV compared with controls was 9.9 %-points, and the marginal effect when EOV was added to other strategy components was 21.5 %-points (table 1, rows 5–6). Direct comparisons of the two strategies revealed differences that were small (0.8 %-points) to modest (6.4 %-points) (table 1, rows 7–8). The one study (with a very high risk of bias) that combined the two strategies had no effect (~2.5 %-points).

Third, group in-service training might be more effective than self-study by 2.0–9.3 %-points (table 1, rows 10–11). The one study (high risk of bias) that combined both strategies had a large effect (24.0 %-points). Fourth, group preservice training for a small portion (eg, a one-semester course) of a preservice training programme had a moderate effect (16.9 %-points) compared with typical preservice training (table 1, row 13). Fifth, while all studies of peer-to-peer training had a high risk of bias, one study found a small effect of peer-to-peer training (4.0 %-points) (table 1, row 14); and median effects of peer-to-peer training combined with in-service training ranged from 8.4 to 25.0 %-points.

We found only five eligible studies (all high risk of bias) of training strategies to improve lay HCP practices (online supplemental appendix 1, table C). All findings were supported by low-quality evidence. There was essentially no effect for in-service training (median MES=–0.9 %-points) and EOV (0.2 %-points). The effect of group preservice training (one study of a 3-day preservice training programme) was 9.1 %-points.

Attributes associated with in-service training effectiveness (objective 2)

Modelling of the ‘training alone’ database included 58 comparisons from 55 studies. This database was derived from the 60 ‘group in-service training versus control’ comparisons (table 1, row 1), with two comparisons excluded because training durations were over 20 days (durations unfeasible for most programmes). See online supplemental appendix 1, tables A5–A6 for sample sizes and risk-of-bias categories for all three modelling databases. Adjusted R² values of the models ranged from 0.102 to 0.340, indicating that they explain only a small-to-moderate amount of the variation in effect sizes.

The analyses of attributes associated with group in-service training effectiveness were supported by low-quality evidence primarily because many studies had a high risk of bias. Several attributes had statistically significant associations with training effectiveness (table 2, rows 1–4). The mean effect of in-service training when some or all training was done off-site and no training was done at the site where HCPs routinely work was 6.0–10.4 %-points greater than when all training was done on-site. In-service training that incorporated clinical practice tended to be more effective than training without this method by 6.9–7.4 %-points. The mean effect of in-service training alone declined with time since training by 0.8–1.0 %-points per month, with training effectiveness waning to zero after 19.8–22.5 months, on average. When training was combined with supervision, the mean effect did not appear to decrease over time (online supplemental appendix 1, figure C), although this result was sensitive to outliers. Finally, lower baseline performance was associated with greater response to training; for every 10 %-point decrease in baseline performance level, mean in-service training effectiveness was 1.1–1.5 %-points higher.

Several training attributes tested by only one study each had relatively large effects that must be interpreted with caution (table 2, rows 5–7): in-service training tailored to HCPs’ stage of readiness to change (23.3 %-points),...
### Table 1 Effectiveness of training strategies on the practices of professional healthcare providers

| Strategies tested* | No of study comparisons (risk of bias: low, moderate, high, very high) | Median MES†, in %-points (IQR; range) |
|--------------------|-----------------------------------------------------------------------|--------------------------------------|
| **Intervention arm** | **Reference arm** | **Outcome scale** | **Median MES†, in %-points (IQR; range)** |
| **Group in-service training** | **Group in-service training** | Controls | Percentage | 60 (9, 19, 17, 15) | 7.3‡ (3.6–17.4; -21.3 to 68.1) |
| **Group in-service training** | **Group in-service training** | Controls | Continuous | 16 (3, 5, 2, 6) | 17.4§ (-2.3 to 28.5; -25.0 to 81.4) |
| **Group in-service training plus other strategy components** | **Other strategy components** | Controls | Percentage | 13 (6, 3, 4, 0) | 3.7¶ (-0.1 to 5.8; -2.7 to 23.6) |
| **Group in-service training plus other strategy components** | **Other strategy components** | Controls | Continuous | 4 (1, 1, 2, 0) | -7.3 (-20.6 to 3.6; -25.8 to 6.4) |
| **Educational outreach visits** | **Educational outreach visits** | Controls | Percentage | 8 (0, 2, 3, 3) | 9.9 (4.3–20.6; 2.8–30.9) |
| **Educational outreach visits plus other strategy components** | **Other strategy components** | Controls | Percentage | 3 (2, 1, 0, 0) | 21.5 (NA; 5.4–30.7) |
| **Group in-service training versus (or combined with) educational outreach visits** | **Educational outreach visits** | Controls | Percentage | 1 (0, 0, 1, 0) | 0.8 (NA; NA) |
| **Group in-service training plus other strategy components** | **Educational outreach visits plus other strategy components** | Controls | Percentage | 2 (2, 0, 0, 0) | -6.4 (NA; -5.8 to -7.0) |
| **Group in-service training plus educational outreach visits** | **Educational outreach visits** | Controls | Percentage | 1 (0, 0, 1, 0) | 2.5 (NA; NA) |
| **Group in-service training versus (or combined with) self-study in-service training** | **Self-study in-service training** | Controls | Percentage | 2 (1, 1, 0, 0) | 9.3 (NA; 4.6–14.0) |
| **Group in-service training plus other strategy components** | **Self-study in-service training plus other strategy components** | Controls | Percentage | 2 (0, 0, 1, 0) | 2.0 (NA; -1.0 to 5.0) |
| **Group in-service training plus self-study in-service training** | **Self-study in-service training** | Controls | Percentage | 1 (0, 0, 1, 0) | 24.0 (NA; NA) |
| **Group preservice training** | **Controls** | Controls | Percentage | 3 (1, 1, 1, 0) | 16.9 (NA; 15.0–46.7) |
| **Peer-to-peer training** | **Controls** | Controls | Percentage | 1 (0, 0, 0, 1) | 4.0 (NA; NA) |
| **Peer-to-peer training plus group in-service training** | **Controls** | Controls | Percentage | 3 (0, 0, 3, 0) | 8.4 (NA; 1.8–66.2) |
| **Peer-to-peer training plus group in-service training plus other strategy components** | **Other strategy components** | Controls | Percentage | 1 (0, 0, 1, 0) | 25.0 (NA; NA) |

*See boxes 1 and 2 for descriptions of the strategies and the comparisons, respectively.
†Effect sizes calculated as the intervention arm improvement minus reference arm improvement.
‡Among studies with a low or moderate risk of bias, median MES=5.1 %-points (IQR: 2.5–14.0; range: -3.0 to 42.8); median MES for high or very high risk of bias studies=9.7 (IQR: 5.1–19.8; range: -21.3 to 68.1).
§Among studies with a low or moderate risk of bias, median MES=15.1 %-points (IQR: -3.8 to 21.2; range: -25.0 to 81.4); median MES for high or very high risk of bias studies=20.3 (IQR: 1.9–41.4; range: -19.2 to 57.3).
¶Among studies with a low or moderate risk of bias, median MES=4.5 %-points (IQR: 2.1–5.8; range: -2.0 to 23.6); median MES for high or very high risk of bias studies=1.4 (IQR: -1.8 to 5.4; range: -2.7 to 7.0).
MES, median effect size; NA, not applicable; %-points, percentage-points.
Preservice training with group feedback (19.0 %-points) and in-service training on a protocol (8.4 %-points). In contrast, several attributes had little or no effect on training effectiveness (<4.5 %-points): training duration; training with computers, interactive methods (e.g., role-play) or written materials; training over multiple sessions; number of educational methods employed; training via live video interactive sessions; use of trainers with pedagogical training and topic complexity (online supplemental appendix 1, table E, rows 7–10). The effects of training group size, using trainers with content expertise, non-interactive lectures, and an interaction between training duration and topic complexity were unclear because of conflicting results (Table 2, rows 8–11).

**Figure 1** Effectiveness of training strategies for professional healthcare providers in low-income and middle-income countries, as assessed with outcomes expressed as percentages. In-service=group in-service training, pre-service=group preservice training, self-study=self-study in-service training, peer-to-peer=peer-to-peer training, N=number of study comparisons. Red indicates results from a single study, which should be interpreted with caution. The numbers next to each spoke are the median of median effect sizes, in percentage points, and (in parentheses) the number of study comparisons. For each comparison, the arrow points toward the study group with greater effectiveness. For example, preservice training was more effective than controls by a median of 16.9 percentage-points, and (paradoxically) controls were more effective than in-service training plus educational outreach visits by a median of 2.5 percentage-points. These are non-training strategy components (e.g., supervision) that could vary among study comparisons, but are the same for any two arms of a given study comparison, for example, educational outreach visits plus supervision versus supervision.

**Cost of in-service training**
Among 84 study arms of professional HCPs exposed to in-service training (from 68 studies), data on cost or from an economic evaluation of any type were available for only 26 arms or 31.0% (even after actively querying...
Table 2: Associations of training attributes on training effectiveness for the practices of professional healthcare providers (HCPs)

| Attribute | Findings |
|-----------|----------|
| **Attributes associated with training effectiveness based on >1 study** | |
| Location of training activities: where HCPs routinely work (on-site) versus all training off-site | ► **Direct evidence**: none. No head-to-head study examined this attribute.  
  ► **Indirect evidence**: having some or all training on-site was more effective than all training off-site by a mean of 6.0–10.4 %-points.  
  ► Details in online supplemental appendix 1, table E, row 1. |
| Use of clinical practice as a training method | ► **Direct evidence**: none. No head-to-head study examined this attribute.  
  ► **Indirect evidence**: training with clinical practice was more effective than training without clinical practice by a mean of 6.9–7.4 %-points.  
  ► Details in online supplemental appendix 1, table E, row 2. |
| Time since training | ► **Direct evidence**: change over time in the marginal effect of supervision given training was 0.3 %-points per month (p=0.58) for 0.5–5.5 months after training.  
  ► **Indirect evidence**: mean effect of training only (without supervision) decreased by 0.8–1.0 %-points per month after training, with the effect predicted to reach zero after 19.8–22.5 months, on average. Mean effect of training plus supervision did not decrease over time (there was a trend of increasing effect of 0.2–0.3 %-points per month, which was not statistically significant). The latter result was sensitive to outliers.  
  ► Details in online supplemental appendix 1, table E, row 3, and figure C. |
| Baseline performance level | ► **Direct evidence**: none. No head-to-head study examined this attribute.  
  ► **Indirect evidence**: mean effect of training decreased by 0.11–0.15 %-points for every 1 % point increase in baseline performance level.  
  ► Details in online supplemental appendix 1, table E, row 4. |
| **Attributes associated with training effectiveness based on only 1 study (ie, interpret with caution)** | |
| Tailoring in-service training to HCPs’ stage of readiness to change | ► **Direct evidence**: training tailored to HCPs’ stage of readiness to change was more effective than non-tailored training by a median of 23.3 %-points.  
  ► **Indirect evidence**: none. The HCPPR database did not include this attribute.  
  ► Details in online supplemental appendix 1, table E, row 5. |
| Preservice training with group feedback about pretraining evaluation results | ► **Direct evidence**: preservice training with group feedback about pretraining evaluation results was more effective than with individual feedback by 19.0 %-points.  
  ► **Indirect evidence**: none. Modelling was not performed for preservice because there were too few studies.  
  ► Details in online supplemental appendix 1, table F. |
| Training on a protocol versus training on clinical acumen | ► **Direct evidence**: training on a protocol-based model (HCPs applied screening results to an algorithm), combined with supervision and integration of services, was more effective than training on clinical acumen (what HCPs did with screening results was left to their discretion), combined with supervision and integration of services, by 8.4 %-points.  
  ► **Indirect evidence**: none. The HCPPR database did not include this attribute.  
  ► Details in online supplemental appendix 1, table E, row 6. |
| **Attributes with an unclear association with training effectiveness because direct and indirect evidence was contradictory** | |
| Trainee group size | ► **Direct evidence**: small group training (ie, 2–14 participants) was somewhat more effective than large group training (ie, >14 participants), by a median of 5.3 %-points.  
  ► **Indirect evidence**: large group training was somewhat more effective than small group training by a mean of 5.8–6.1 %-points.  
  ► Details in online supplemental appendix 1, table E, row 11. |
| Trainers with content expertise | ► **Direct evidence**: training by trainers with content expertise (doctors) was slightly more effective than training by trainers without content expertise (paramedics), by 2.5 %-points.  
  ► **Indirect evidence**: training when ‘all trainers were content experts’ was less effective than when not all trainers were content experts, by a mean of 16.1 %-points.  
  ► Details in online supplemental appendix 1, table E, row 12. |
| Use of non-interactive lectures as a training method | ► **Direct evidence**: training with a non-interactive lecture or session (as a sole training method) was more effective than interactive training (as a sole training method) by a median of 5.0 %-points.  
  ► **Indirect evidence**: no significant association. However, all univariable regression coefficient β values were less than 5.0 %-points (range: −2.5 to 3.8 %-points; all non-significant).  
  ► Details in online supplemental appendix 1, table E, row 13. |
investigators); and 15 arms (from 11 studies) had data that allowed us to calculate cost per HCP per day of training. These 11 studies were from countries in Africa, Asia, Latin America and Europe. The median training duration was 5 days (IQR: 4–9, range: 1–19). The median cost per HCP per day of training was $26 (IQR: 4–72, range: 1–94). Cost was not related to study year (which ranged from 1991 to 2013) or whether the training was done on-site versus off-site. Costs tended to be higher in low-income countries (median = $54/HCP/day, N=5 arms) than in middle-income countries (median = $27/HCP/day, N=10 arms).

**DISCUSSION**

In LMICs, training is often used to improve HCP performance and reduce the enormous burden caused by poor quality healthcare. While experts have voiced the need to go beyond training and implement other strategies and health systems interventions, training will continue to be a core element of most improvement approaches. This analysis of HCPPR data characterises the effectiveness of different training strategies and identifies attributes associated with training effectiveness. Strengths of this analysis are the large number of studies that are all from LMICs, the inclusion of head-to-head studies that directly compare training strategies and quantitative results.

For professional HCPs, certain training strategies appear to be more effective at improving HCP practices: on average, EOV was somewhat more effective than in-service training, while the latter might in turn be more effective than peer-to-peer training or self-study. Certain combinations (eg, in-service training plus self-study) might be more effective than others (eg, in-service training plus EOV). We also identified attributes of in-service training that were associated with larger, sustained improvements: on-site training, incorporating clinical practice and combining training with supervision. While our results suggest that certain approaches are more effective, the variability of results and the overall low-quality of evidence suggest that (as the larger HCPPR emphasised) programmes should monitor performance to understand the effect of a given approach in their specific context. For the small number of studies of lay HCPs, in-service training had little or no measurable effect. Much more work is needed in this crucial area given the reliance on this cadre of HCPs in many settings.

Despite analysing 199 studies, the overall quality of evidence about training interventions is weak and substantial knowledge gaps remain. An evidence-based research agenda derived from this study (box 4) suggests greater attention to replication of key results, additional studies of specific promising strategies, better reporting on context, and use of more rigorous study designs and standardised methods.

Our estimates of strategy effectiveness are generally similar to those from other reviews, although methodological differences limit comparisons. The review by Forsetlund et al (with 11 of 81 studies from LMICs) found the median effect of educational meetings on HCP practices was 6 %-points for dichotomous outcomes and 10 %-points for continuous outcomes. The effect of provider education reported by Holloway et al (all studies from LMICs) was 6 %-points. A review on educating lay HCPs identified only two studies, which had divergent results: no effect and 22 %-points. A review on EOV (with 3 of 69 studies from LMICs) found a median effect of 5.6 %-points for dichotomous outcomes. Notably, an HCPPR-based study of strategy effectiveness over time found a decline in the effect of training plus supervision (0.4–0.5 %-points/month)—the opposite of our result. As that study analysed effectiveness at multiple follow-up time points per study, its results probably have greater validity.

Certain reviews present lists of recommended attributes for effective training or learning. Some of these attributes agree with our results (eg, on-site training (personal communication from Alison Trump (presentation: ‘Using evidence to design effective learning to support health worker performance’), Jhpiego, 22 August 2019) and Bluestone et al 2020 and combining training with supervision, while some do not (eg, training over multiple sessions and interactive methods (personal communication from Alison Trump (presentation: **Table 2** Continued

| Attribute                                      | Findings                                                                                                                                                                                                                                                                                                                                 |
|------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Interaction between in-service training duration and the topic complexity of the training | *Direct evidence*: for training on single topics, training effectiveness might have increased with course duration; and for training on multiple topics, training effectiveness seemed unrelated to course duration.  
*Indirect evidence*: for training on single topics, training effectiveness was unrelated to course duration; but for training on multiple topics, effectiveness increased with longer course duration.  
Details in online supplemental appendix 1, table E, row 14.                                                                                                                                                      |

*Direct evidence comes from head-to-head studies (ie, the direct comparison of two training approaches), and indirect evidence comes from modelling results (ie, essentially the difference between the mean effect of a group of ‘intervention vs control’ comparisons from studies that used one training approach, and the mean effect of a group of ‘intervention vs control’ comparisons with a different training approach).  
HCPPR, Health Care Provider Performance Review; %-point, percentage-point.  

Forsetlund et al 2019) and Bluestone et al 2020 and combining training with supervision, while some do not (eg, training over multiple sessions and interactive methods (personal communication from Alison Trump (presentation:
Regarding topic areas, future research should focus on:

- Replicating studies of promising strategies tested with only one or a small number of studies.
- Head-to-head comparisons of key training strategies (eg, in-service training vs educational outreach visits), strategy combinations (eg, in-service training plus peer-to-peer training vs in-service training alone) and training attributes (eg, different training methods and durations).
- Use of computers as a training adjunct, especially self-study in-service training with computers, given the increased availability of personal computers and internet-based training courses.
- Rigorous studies of training strategies to improve the practices of lay or community health workers.
- Better quantitative and qualitative understanding of how context influences strategy effectiveness.

Regarding methods, future research should:

- Use standardised methods, especially for outcomes, strategy description, implementation (including dose and fidelity) and characterisation of study context.
- Prioritise head-to-head studies, which provide stronger evidence for comparing different training approaches.
- Have rigorous study designs, such as interrupted time series with a randomised comparison group, which reduce bias and show how effectiveness changes over time.
- Prioritise the use of practice outcomes expressed as a percentage (rather than continuous outcomes, which can be more difficult to interpret; see online supplemental appendix 1, figure D).
- Have follow-up periods that match the timeframe that programs require for improvements to be sustained (eg, at least 12 months) and include multiple measures of effect so changes (reductions or further improvements) in effectiveness over time can be quantified.
- Include assessments of strategy cost and cost-effectiveness.
- Be designed to better contribute to filling gaps in the evidence base about strategy choice and combinations of components.*

*Studies directly comparing two training approaches without other components are the easiest to interpret. However, given the generally moderate effect of training as a sole strategy, studies should include other enhancing components in both study arms (eg, training approach A + quarterly supervision vs training approach B + quarterly supervision).

CONCLUSIONS

This analysis has characterised the effectiveness of training strategies and identified attributes associated with training effectiveness in LMICs. Although more higher quality research is needed, for now, decision-makers should consider these results to make HCP training in LMICs more effective and ultimately to strengthen health programmes that serve billions of people.

Author affiliations

1Division of Parasitic Diseases and Malaria, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
2CDC Foundation, Atlanta, Georgia, USA
3Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
4International Institute of Health Management Research, Jaipur, India
5Institute of Development Studies, University of Sussex, Brighton, UK
6Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA

Acknowledgements This article is based on information in the Health Care Provider Performance Review (HCPR), a joint program of CDC, Harvard Medical School, WHO, Management Sciences for Health, Johns Hopkins University and the CDC Foundation. We are grateful for the excellent assistance from the data abstractors, librarians, statistical advisors and data managers who worked on the HCPR; and the responses that many authors provided to our questions about their studies.

Contributors AKR, DR-D and KAH conceived the review. All authors (AKR, SYR, DHP, KAH and DR-D) developed the methods and the review's protocol. AKR and SYR had primary responsibility for the analysis and initial draft of the manuscript. All authors contributed substantially to the analysis, interpretation of the results and completion of the manuscript. All authors approved the final manuscript. In addition, AKR, the corresponding author, had full access to all the reports included in the review and had final responsibility for the decision to submit for publication.

Funding This review was supported by funding from the CDC Foundation through a grant from the Bill & Melinda Gates Foundation (grant OPP52730), and from CDC, and a World Bank—Netherlands Partnership Program Grant (project number P09685).

Disclaimer The findings and conclusions presented in this report are those of the authors and do not necessarily reflect the official position of the Centers for Disease Control and Prevention.
Control and Prevention or the CDC Foundation. The funders had no influence on the design and conduct of the review; collection, management, analysis and interpretation of the data; or preparation of this report.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The data are publicly available at https://www.HCPperformanceview.org.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation or adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs
Alexander K Rowe http://orcid.org/0000-0002-1383-4341
Samantha Y Rowe http://orcid.org/0000-0003-3729-579X
David H Peters http://orcid.org/0000-0001-8377-3444
Kathleen A Holloway http://orcid.org/0000-0001-5032-6357
Dennis Ross-Degnan http://orcid.org/0000-0002-0066-1242

REFERENCES
1. Rowe AK, Rowe SY, Peters DH, et al. Effectiveness of strategies to improve health-care provider practices in low-income and middle-income countries: a systematic review. *Lancet Glob Health* 2018;6:e1163–75. doi:10.1016/S2214-109X(18)30398-X
2. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. *Lancet Glob Health* 2018;6:e1196–252. doi:10.1016/S2214-109X(18)30386-3
3. National Academies of Sciences, Engineering, and Medicine. *Crossing the global quality chasm: improving health care worldwide*. Washington, DC: The National Academies Press, 2018. doi:10.17226/25152
4. World Health Organization. *Ninth programme report 1992–1993*. Geneva: Programme for Control of Diarrhoeal Diseases, World Health Organization, 1994.
5. President’s Malaria Initiative. Fiscal year 2017 malaria operational plans. Available: https://www.pmi.gov/resource-library/mops/fy-2017 [Accessed 11 Feb 2019].
6. President’s Malaria Initiative. Fiscal year 2018 malaria operational plans. Available: https://www.pmi.gov/resource-library/mops/fy-2018 [Accessed 13 Feb 2019].
7. President’s Malaria Initiative. Fiscal year 2019 malaria operational plans. Available: https://www.pmi.gov/resource-library/mops/fy-2019 [Accessed 14 Feb 2019].
8. Dielmann M, Gerretsen B, van der Walt GJ. Human resource management interventions to improve health workers’ performance in low and middle-income countries: a realist review. *Health Res Policy Syst* 2009;7:7. doi:10.1186/1478-4505-7-7
9. Forsetlund L, Bjernad A, Rashidian A, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2009;CD003030. doi:10.1002/14651858.CD003030.pub2
10. Jhpiego. Breaking translation: translating evidence into effective learning. Available: https://www.jhpiego.org/wp-content/uploads/2019/06/Effective-Evidence-Based-Learning-Brief.pdf [Accessed 22 Aug 2019].
11. O’Brien MA, Rogers S, Jamtvedt G, et al. Educational outreach visits: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2007;CD004009. doi:10.1002/14651858.CD004009.pub2
12. Pantoea T, Opiyo N, Lewin S, et al. Implementation strategies for health systems in low-income countries: an overview of systematic reviews. *Cochrane Database Syst Rev* 2017;9:CD011086. doi:10.1002/14651858.CD011086.pub2
13. Rowe SY, Peters DH, Holloway KA, et al. The effectiveness of supervision strategies to improve health care provider practices in low- and middle-income countries. *Unpublished manuscript.*
14. Rowe SY, Peters DH, Holloway KA, et al. A systematic review of the effectiveness of strategies to improve health care provider performance in low- and middle-income countries: methods and descriptive results. *PloS One* 2019;14:e0217617. doi:10.1371/journal.pone.0217617
15. World Health Organization, Organisation for Economic Co-operation and Development, International Bank for Reconstruction and Development. Delivering quality health services: a global imperative for universal health coverage. Geneva: World Health Organization, 2018. http://apps.who.int/iris/bitstream/handle/10665/272465/978924153906-0eng.pdf?ua=1
16. World Health Organization. Handbook for national policy quality and strategy: a practical approach for developing policy and strategy to improve quality of care. Geneva: World Health Organization, 2018. http://apps.who.int/iris/bitstream/handle/10665/272357/9789241565661-eng.pdf?ua=1. (accessed on 2020 June 17).
17. Holloway KA, Ivanovska V, Wagner AK, et al. Have we improved use of medicines in developing and transitional countries and do we know how to? Two decades of evidence. *Trop Med Int Health* 2013;18:656–64. doi:10.1111/tmi.12123
18. Sorsdahl K, Ipser JC, Stein DJ. Interventions for educating traditional healers about STD and HIV medicine. *Cochrane Database Syst Rev* 2009;CD007190. doi:10.1002/14651858.CD007190.pub2
19. Arsenault C, Rowe SY, Ross-Degnan D, et al. Does the effectiveness of strategies to improve health care provider practices in low- and middle-income countries change over time? Secondary analysis of a systematic review. *Unpublished manuscript under review.*
20. Bluestone J, Johnson P, Fullerton J, et al. Effective in-service training design and delivery: evidence from an integrative literature review. *Hum Resour Health* 2013;11:51. doi:10.1186/1478-4491-11-51
Appendix 1 for:

Rowe AK, Rowe SY, Peters DH, Holloway KA, Ross-Degnan D. The effectiveness of training strategies to improve health care provider practices in low- and middle-income countries

Section 1. Methodological details

Section 2. Additional results
Section 1. Methodological details

Objective 1: Characterize the effectiveness of different training strategies

- To define the non-training part of a strategy (i.e. the “other” strategy components), we used strategy definitions at the “component category” level, rather than the “individual component” level—although the non-training parts of the strategy in two arms from the same study were usually identical at the individual component level. For example, if arm 1 was “training + provision of drug supply” and arm 2 was “provision of drug supply”, then the non-training part is the same.

- Several component categories (e.g., printed information for health care providers [HCPs], patient support, and community support) contain numerous individual components. On a few occasions, the individual strategy components between study arms did not exactly match; however, the categories of strategy components matched. For example, in one study comparison: arm 1 had “poster for HCP + poster for community + drug supply” and arm 2 had “training + educational video for community + poster for HCP + poster for community + drug supply.” Even though the individual “community support” components were not exactly the same between the arms (Arm 1: poster for community vs. Arm 2: educational video + poster for community), both arms had components in the “community support” category so that the “other” non-training parts of the strategies were the same in both arms.

- If study arm 1 had printed information for HCPs but no training, and study arm 2 had training with printed information for HCPs handed out (originally coded as “training only”), the strategy in arm 2 was re-defined as “training + printed information for HCPs” so that both arms had the same “other” non-training component of “printed information for HCPs”. This approach better reflects what each study arm was exposed to and improves the interpretability of the effect sizes.

- Indirect analysis (true control comparisons) and direct analysis (head-to-head comparisons) results were presented together in a “network” diagram (Figure 2). However, network meta-analysis was not performed because it would not have added much to the simpler analysis used. Specifically (as can be seen in Figure 2): in the main network (with the no-intervention control reference group), there is only 1 closed loop (which adds a single study comparison, while the other spokes have 59 and 8 comparisons). Additionally, in the smaller network (with non-training strategy components as the reference group), the “other X” nodes represent a diversity of strategies, so they are not combinable (i.e., they only make sense for “within study” comparisons, which quantify the marginal effect of the training strategy conditioned on other components in the strategy).
Objective 2: Identify attributes of in-service training associated with training effectiveness

- Indirect analysis (modeling of study comparisons with a no-intervention control group)

1. For each of the three databases of training studies (i.e., training only, training +/- supervision, training +/- other strategy components), we created a mixed model with a random-intercept (in which the cluster was the study) using a 4-step approach: 1) univariable analyses of individual training attributes; 2) attributes with univariable p-values < 0.10 were identified; 3) step 2 attributes were included in a multivariable model (except the variable for “some or all training on-site”, despite its having a univariable p < 0.10, because this variable was missing for about one-fifth of effect sizes*); and 4) if the step 3 model included the “duration-topic complexity” interaction or the “supervision-time” interaction, and that interaction was not significant (p < 0.05), then the interaction and components (duration & complexity, or time) were removed from the model. To examine the association for the “some or all training on-site” variable (which was often missing), we developed an alternative model with the following 4-step approach: 1) univariable analyses of individual training attributes; 2) attributes with univariable p-values < 0.10 were identified; 3) step 2 attributes were included in a multivariable model that included the “some or all training on-site” variable; and 4) if the step 3 model included the “duration-topic complexity” interaction or the “supervision-time” interaction, and that interaction was not significant (p < 0.05), then the interaction and components (duration & complexity, or time) were removed from the model. For the models containing the “some or all training on-site” variable, we did not consider results for other training attributes (e.g., training duration); these models were only used to evaluate the effect of some/all on-site training (adjusted for other factors, as potential confounders).

2. The following training attributes were excluded because they were highly unbalanced (i.e., one level of the attribute had <5 comparisons): whether training used all four key educational methods (clinical practice, interactive session, non-interactive lecture, and role play), and whether training used computer-based methods.

3. Attempts to add training attributes not included in the step 4 model, or to include non-training effect modifiers, resulted in unstable models. Out of concerns that more complex models might be over-specified, we only tested one set of additional models that included variables for baseline performance and time since training, as they were known predictors of effect size. Thus, for each of the three databases, we had four final models (see Tables D1–D3).
   a) Model 1: no predictors forced into the model
   b) Model 2: baseline performance and time since training forced into the model
   c) Model 3: some/all on-site training forced into the model
   d) Model 4: some/all on-site training, baseline performance, and time since training forced into the model

4. Details on eligibility
   a) Inclusion criteria: 1) professional HCP studies (i.e., no lay HCP predominant studies) with at least one comparison of in-service group training versus a true control, and 2) training duration < 20 days (studies with missing training duration were included).
   b) Exclusion criteria: 1) educational outreach visits, 2) peer-to-peer training, 3) self-study, 4) pre-service training, and 5) equivalency studies.
c) Note. Regarding distance learning, studies were eligible if there was a classroom of trainees with an off-site trainer (i.e., studies of HCPs studying in isolation were considered “self-study” and excluded).

5. Additional details on modeling for the three training databases
   a) Database of studies on training only (see Table D1). See methods above; no additional details.
   b) Database of studies on training +/- supervision (see Table D2). All models included 1 indicator variable for the presence of any supervision.
   c) Database of studies on training +/- other strategy components (see Table D3). All models included nine indicator variables for the presence of non-training components (i.e., community supports, patient supports, strengthen infrastructure, health systems financing and other incentives, governance or regulation, group problem solving, supervision, other management techniques, and information and communication technology for HCPs). Two indicator variables for the presence of other non-training component categories (i.e., HCP-directed financial incentives, and printed information or job aids for HCPs) were excluded from models because they were highly unbalanced (i.e., one level of the variable had <10 comparisons).

* The variable for “some or all training on-site” had missing values for: 93 (25.0%) of 372 observations in the training only dataset, 113 (24.0%) of 470 observations in the training +/- supervision dataset, and 160 (18.7%) of 856 observations in the training +/- other components dataset.

• Direct analysis (head-to-head comparisons)
  1. Eligible comparisons were: a) in-service training approach A vs. in-service training approach B (e.g., a 6-day training versus an 11-day training), and b) in-service training approach A + other strategy components vs. in-service training approach B + other strategy components.
Box A. Attributes of successful training according to a specialist in the science of how people learn

1. The training uses analogies as bridges to link new knowledge to prior knowledge
2. Trainers actively attempt to identify and address misconceptions directly
3. Information is categorized (i.e., presented using relevant categories)
4. Information is sequenced (i.e., presented in a logical sequence)
5. Trainees practice both individual skills (e.g., performing a skin pinch to evaluate dehydration in a child with diarrhea) and the entire set of desired practices (e.g., all aspects of evaluating a child with diarrhea) (this agrees with Malcolm Knowles Principle of Andragogy #2: Experience should be at the root of all learning tasks and activities)
6. Trainers provide informational feedback (i.e., rather than only praise or criticism)
7. Complex information is simplified
8. Training uses images
9. Training uses mnemonics
10. Training uses stories, case studies, problem-based learning, or simulations (this agrees with Malcolm Knowles Principle of Andragogy #4: Adult learning should be problem-centered, rather than content-oriented)
11. Procedures are broken down into steps
12. The training has breaks to avoid overwhelming trainees
13. Trainees are asked to discuss, debate, or persuade each other
14. Some aspect of the training involves trainees collaborating with each other
15. Trainees are asked to teach each other
16. Trainers help trainees tie the training’s objective to a self-relevant, self-transcendent purpose (e.g., for training on treating an illness, trainers helped trainees understand that improving treatment practices will both make them a better, more respected health worker and save the lives of people in their community) (this agrees with Malcolm Knowles Principle of Andragogy #3: Adult learning should have immediate relevance to real life)
17. Trainers recognize trainees growing competence
18. Trainers help trainees develop self-efficacy (i.e., confidence that trainees can perform the required tasks)
19. Trainers ask trainees to make a plan on how the new knowledge would be put to use. The plan includes specifying outcomes and how the outcomes will be measured, and setting goals that are short-term, specific, and moderately challenging.
20. Trainers should get feedback from adult learners (based on Malcolm Knowles Principle of Andragogy #1: Adult learners must be involved in the design and development of their learning experience).

Adapted from a presentation, citation: Annie Murphy Paul. "Learning Science." Presentation at the Teach to Reach Summit, Seattle, Washington, November 2, 2015.
Section 2. Additional results

**Section 2a. Detailed flowchart of the literature search of the systematic review on which this study is based**

Figure A. Detailed flowchart of the literature search

Health Care Provider Performance Review Detailed Flowchart of Literature Search

Abbreviation: HCP = health care provider.

- Early in the initial review’s search of on-line document inventories and websites, detailed records were not kept of the number of citations that were screened. Thus, the number of exclusions is unknown; the exact number of records screened is unknown, but was more than 23,265 (which reflects the number once detailed records began to be kept); and the exact number of included articles is unknown, but was more than 205 (which reflects the number once detailed records began to be kept).

- Early in the initial review’s search of the bibliographies of the 510 previous reviews and other papers, detailed records of the search were not kept. Thus, the number of exclusions and full-text assessments are unknown; and the exact number of included articles is unknown, but was more than 247 (which reflects the number once detailed records began to be kept).
Section 2b. Sample size information

Table A1. Sample size information: analysis of percentage and continuous practice outcomes for study objectives 1 and 2 combined

| Health care provider category | True control comparisons | Head-to-head comparisons | Total          |
|-------------------------------|--------------------------|--------------------------|----------------|
| LHW predominant               | 15 effect sizes          | 2 effect sizes           | 17 effect sizes|
|                               | 4 comparisons            | 2 comparisons            | 6 comparisons  |
|                               | 4 studies                | 1 study                  | 5 studies      |
|                                |                          |                          |                |
| Not LHW predominant           | 953 effect sizes         | 230 effect sizes         | 1183 effect sizes |
|                               | 186 comparisons          | 48 comparisons           | 234 comparisons|
|                               | 168 studies              | 38 studies               | 194 studies    |
|                                |                          |                          |                |
| Total                         | 968 effect sizes         | 232 effect sizes         | 1200 effect sizes |
|                               | 190 comparisons          | 50 comparisons           | 240 comparisons|
|                               | 172 studies              | 39 studies               | 199 studies    |

Abbreviation: LHW = lay health worker.

a These studies evaluated the effect of a strategy designed to improve LHW practices, even if other types of health workers were providing services in the study setting.

b One effect size for a percentage practice outcome from a study of non-LHW-predominant health care providers was an equivalency comparison with a gold standard control group (COMP_IDnew 3640000112: intervention group: in-service training for midwives vs. control group: no in-service training for physicians).

Table A2. Sample size information: analysis of percentage practice outcomes for study objective 1

| Health care provider category | True control comparisons | Head-to-head comparisons | Total          |
|-------------------------------|--------------------------|--------------------------|----------------|
| LHW predominant               | 15 effect sizes          | 2 effect sizes           | 17 effect sizes|
|                               | 4 comparisons            | 2 comparisons            | 6 comparisons  |
|                               | 4 studies                | 1 study                  | 5 studies      |
|                                |                          |                          |                |
| Not LHW predominant           | 460 effect sizes         | 82 effect sizes          | 542 effect sizes |
|                               | 78 comparisons           | 24 comparisons           | 102 comparisons|
|                               | 73 studies               | 19 studies               | 85 studies     |
|                                |                          |                          |                |
| Total                         | 475 effect sizes         | 84 effect sizes          | 559 effect sizes |
|                               | 82 comparisons           | 27 comparisons           | 108 comparisons|
|                               | 77 studies               | 21 studies               | 90 studies     |

Abbreviation: LHW = lay health worker.

a These studies evaluated the effect of a strategy designed to improve LHW practices, even if other types of health workers were providing services in the study setting.

b Eleven effect sizes from 3 true control comparisons of “in-service training alone” from 2 studies involved a training duration longer than 20 days: 5 effect sizes with a 30-day training (COMP_IDnew 6920000112), 5 effect sizes with a 40-day training (COMP_IDnew 6920000113), and 1 effect size (ES_ID 3640000107007) with a 60-day training, which was also an equivalency comparison.
# Table A3. Sample size information: analysis of continuous practice outcomes for study objective 1

| Health care provider category | True control comparisons | Head-to-head comparisons | Total |
|-------------------------------|--------------------------|--------------------------|-------|
| LHW predominant<sup>a</sup>  | 0 effect sizes           | 0 effect sizes           | 0 effect sizes |
| Not LHW predominant           | 27 effect sizes          | 10 effect sizes          | 37 effect sizes |
|                               | 4 comparisons            | 16 comparisons           | 20 comparisons |
|                               | 4 studies                | 16 studies               | 18 studies |

Abbreviation: LHW = lay health worker.

<sup>a</sup> These studies evaluated the effect of a strategy designed to improve LHW practices, even if other types of health workers were providing services in the study setting.

# Table A4. Sample size information: analysis of percentage practice outcomes for study objective 2

| Health care provider category | True control comparisons | Head-to-head comparisons | Total |
|-------------------------------|--------------------------|--------------------------|-------|
| LHW predominant<sup>a</sup>  | 0 effect sizes           | 0 effect sizes           | 0 effect sizes |
| Not LHW predominant           | 856 effect sizes         | 138 effect sizes         | 994 effect sizes |
|                               | 168 comparisons          | 24 comparisons           | 192 comparisons |
|                               | 152 studies              | 20 studies               | 169 studies |

Abbreviation: LHW = lay health worker.

<sup>a</sup> These studies evaluated the effect of a strategy designed to improve LHW practices, even if other types of health workers were providing services in the study setting.
Table A5. Sample size and risk-of-bias information for the three databases used in the modeling analysis to identify attributes associated with in-service training effectiveness: effect size level (study objective 2)

| Database of studies that tested training only: 372 effect sizes from 58 comparisons from 55 studies | Frequency | Percent |
|---------------------------------|-----------|---------|
| Very high                       | 126       | 33.9    |
| High                            | 99        | 26.6    |
| Moderate                        | 111       | 29.8    |
| Low                             | 36        | 9.7     |

| Database of studies that tested training +/- supervision: 470 effect sizes from 79 comparisons from 73 studies | Frequency | Percent |
|-------------------------------------------------------------------------------------------------|-----------|---------|
| Very high                                        | 130       | 27.7    |
| High                                             | 147       | 31.3    |
| Moderate                                         | 153       | 32.6    |
| Low                                              | 40        | 8.5     |

| Database of studies that tested training +/- other components: 856 effect sizes from 168 comparisons from 152 studies | Frequency | Percent |
|---------------------------------------------------------------------------------------------------------------|-----------|---------|
| Very high                                                                                                    | 245       | 28.6    |
| High                                                                                                         | 232       | 27.1    |
| Moderate                                                                                                     | 258       | 30.1    |
| Low                                                                                                          | 121       | 14.1    |
Table A6. Sample size and risk-of-bias information for the three databases used in the modeling analysis to identify attributes associated with in-service training effectiveness: study level (study objective 2)

| Database of studies that tested training only: 55 studies | Risk of bias category | Frequency | Percent |
|-----------------------------------------------------------|-----------------------|-----------|---------|
|                                                            | Very high             | 15        | 27.3    |
|                                                            | High                  | 16        | 29.1    |
|                                                            | Moderate              | 15        | 27.3    |
|                                                            | Low                   | 9         | 16.4    |

| Database of studies that tested training +/- supervision: 73 studies | Risk of bias category | Frequency | Percent |
|---------------------------------------------------------------------|-----------------------|-----------|---------|
|                                                                      | Very high             | 19        | 26.0    |
|                                                                      | High                  | 24        | 32.9    |
|                                                                      | Moderate              | 19        | 26.0    |
|                                                                      | Low                   | 11        | 15.1    |

| Database of studies that tested training +/- other components: 152 studies | Risk of bias category | Frequency | Percent |
|---------------------------------------------------------------------------|-----------------------|-----------|---------|
|                                                                            | Very high             | 48        | 31.6    |
|                                                                            | High                  | 43        | 28.3    |
|                                                                            | Moderate              | 34        | 22.4    |
|                                                                            | Low                   | 27        | 17.8    |
**Section 2c. Descriptive results of included studies**

Table B1. General attributes of included studies

| Study attribute                              | All studies (N=199) |
|----------------------------------------------|---------------------|
| Number of study arms                         |                     |
| 1                                            | 15 (7.5%)           |
| 2                                            | 157 (78.9%)         |
| 3                                            | 20 (10.1%)          |
| 4                                            | 7 (3.5%)            |
| Total number of study arms across all studies| 417                 |
| Total number of comparisons across all studies|                     |
| Strategy vs. true (no intervention) control group | 188 (78.3%)       |
| Strategy A vs. Strategy B with no placebo components | 48 (20.0%)        |
| Strategy vs. placebo control group           | 2 (0.8%)            |
| Strategy A vs. Strategy B with ≥1 placebo component | 2 (0.8%)          |
| Number of effect sizes per study and comparison|                     |
| Median number of effect sizes per study (range) | 3 (1–134)          |
| Median number of effect sizes per comparison (range) | 2 (1–67)          |
| Study designs                                |                     |
| Pre-post study with randomized controls       | 75 (37.7%)          |
| Pre-post study with non-randomized controls   | 68 (34.2%)          |
| Post-only study with randomized controls      | 37 (18.6%)          |
| Interrupted time series with no controls      | 15 (7.5%)           |
| Interrupted time series with randomized controls | 3 (1.5%)           |
| Interrupted time series with non-randomized controls | 1 (0.5%)          |
| Economy of country where study was done       |                     |
| Low income                                   | 79 (39.7%)          |
| Lower-middle income                          | 70 (35.2%)          |
| Upper-middle income                          | 48 (24.1%)          |
| Combination of lower-middle and upper-middle income | 1 (0.5%)        |
| Combination of lower and middle income        | 1 (0.5%)            |
| Risk of bias                                 |                     |
| Low                                          | 36 (18.1%)          |
| Moderate                                     | 44 (22.1%)          |
| High                                         | 60 (30.1%)          |
| Very high                                    | 59 (29.7%)          |
| WHO region where study was conducted         |                     |
| Africa                                       | 70 (35.2%)          |
| Southeast Asia                               | 44 (22.1%)          |
| America                                      | 36 (18.1%)          |
| Study attribute | All studies (N=199) |
|-----------------|---------------------|
| Western Pacific | 27 (13.6%)          |
| Eastern Mediterranean | 18 (9.1%)    |
| Europe          | 3 (1.5%)            |
| Africa, America, Southeast Asia, Western Pacific | 1 (0.5%) |

| Year of publication (or date of document for unpublished reports), by decade | All studies |
|---------------------------------------------------------------------------|-------------|
| 2010 or later (latest year was 2017)*                                      | 69 (34.7%)  |
| 2000–2009                                                                  | 91 (45.7%)  |
| 1990–1999                                                                  | 38 (19.1%)  |
| 1980–1989                                                                  | 1 (0.5%)    |

| Data collection methods (multiple responses allowed per study) | All studies |
|----------------------------------------------------------------|-------------|
| Record or chart review                                         | 103 (51.8%) |
| Interview with patient or patient’s caretaker                 | 67 (33.7%)  |
| Observation of HCP-patient interaction                         | 43 (21.6%)  |
| Interview with HCP                                            | 27 (13.6%)  |
| Simulated client                                              | 27 (13.6%)  |
| Questionnaire for HCP (any administration method)              | 22 (11.1%)  |
| Physical exam of patient by study team                         | 13 (6.5%)   |
| Exam for HCP (e.g., written test for HCP)                      | 8 (4.0%)    |
| Questionnaire for patient or patient’s caretaker               | 6 (3.0%)    |
| Observation of HCP practices not involving real patients       | 5 (2.5%)    |
| Case scenario                                                 | 4 (2.0%)    |
| Observation of facility                                       | 3 (1.5%)    |
| HCP self-assessment                                           | 3 (1.5%)    |
| Interview with administrator                                   | 2 (1.0%)    |
| Observation of patient’s or patient caretaker’s behaviors      | 2 (1.0%)    |
| Questionnaire for an administrator                             | 1 (0.5%)    |

| Urban vs. rural study setting | All studies |
|-------------------------------|-------------|
| Urban +/- peri-urban areas    | 72 (36.2%)  |
| Mix of urban and rural areas  | 46 (23.1%)  |
| Rural areas only              | 39 (19.6%)  |
| Town +/- rural areas          | 10 (5.0%)   |
| Peri-urban areas only         | 4 (2.0%)    |
| Mix of peri-urban and town areas | 1 (0.5%) |
| Unclear or not stated         | 27 (13.6%)  |

| Data available on strategy cost or other economic evaluation (from either the study reports or responses from investigators) | All studies |
|-----------------------------------------------------------------------------------------------------------------------------|-------------|
|                                                                                                                              | 72 (36.2%)  |

Abbreviations: HCP = Health care provider, WHO = World Health Organization.

* Many reports from 2016 and all from 2017 either were originally identified as unpublished, but were published by the time of the analysis, or were reports that authors or experts provided after the formal literature search had ended.
Table B2. Settings of included studies: places where services were delivered, who owned or operated the service delivery points, and types of health care providers

| Study attribute                                                                 | All studies (N=199) |
|---------------------------------------------------------------------------------|---------------------|
| **Places where services were delivered (multiple responses allowed)**             |                     |
| Outpatient health facility                                                      | 112 (56.3%)         |
| Hospital outpatient department                                                   | 55 (27.6%)          |
| Hospital inpatient wards                                                         | 43 (21.6%)          |
| Household or community setting                                                   | 23 (11.6%)          |
| Pharmacy                                                                         | 17 (8.5%)           |
| Drug shop                                                                        | 14 (7.0%)           |
| Non-hospital health facility inpatient ward                                      | 9 (4.5%)            |
| School                                                                           | 5 (2.5%)            |
| Site in transit to hospital or health facility                                   | 1 (0.5%)            |
| Other outpatient setting                                                        | 3 (1.5%)            |
| **Who owns or operates the place where services were delivered (multiple responses allowed per study)** |                     |
| Public or government                                                             | 142 (71.4%)         |
| Private, for profit                                                              | 35 (17.6%)          |
| Community                                                                       | 25 (12.6%)          |
| Private, not for profit                                                          | 18 (9.1%)           |
| Private, profit status unknown or not reported                                   | 15 (7.5%)           |
| Other                                                                            | 3 (1.5%)            |
| Unclear or not reported                                                          | 15 (7.5%)           |
| **Type of health care providers (multiple responses allowed per study)**         |                     |
| Physician                                                                       | 112 (56.3%)         |
| Nurse                                                                            | 89 (44.7%)          |
| Midwife                                                                          | 36 (18.1%)          |
| Nurse aide                                                                       | 36 (18.1%)          |
| Pharmacist assistant or non-pharmacist drug vendor                              | 27 (13.6%)          |
| Pharmacist                                                                       | 24 (12.1%)          |
| Paramedic or unspecified non-physician                                           | 24 (12.1%)          |
| Lay health worker                                                                | 23 (11.6%)          |
| Clinical officer                                                                 | 15 (7.5%)           |
| Health educator or information officer                                           | 14 (7.0%)           |
| Midwife aide                                                                     | 9 (4.5%)            |
| Student                                                                          | 6 (3.0%)            |
| Laboratorist                                                                     | 4 (2.0%)            |
| Health care provider, type unspecified                                          | 11 (5.5%)           |
| Lay health worker was the predominant type of health care provider               | 5 (2.5%)            |
Table B3. Health conditions addressed by included studies

| Health condition (multiple responses allowed per study) | No. of studies with at least one effect size related to the health condition, among all 199 studies |
|---------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| Multiple (or all) health conditions                     | 51 (25.6%)                                                                                   |
| Acute respiratory infections                            | 31 (15.6%)                                                                                   |
| Pregnancy                                               | 29 (14.6%)                                                                                   |
| Malaria                                                 | 25 (12.6%)                                                                                   |
| Diarrhea                                                 | 25 (12.6%)                                                                                   |
| Reproductive health (not pregnancy related)             | 17 (8.5%)                                                                                    |
| HIV/AIDS +/- other sexually transmitted diseases        | 15 (7.5%)                                                                                    |
| Newborn health conditions                               | 14 (7.0%)                                                                                    |
| Malnutrition                                            | 13 (6.5%)                                                                                    |
| Non-communicable diseases not covered by other categories (e.g., asthma) | 10 (5.0%)                                                                                   |
| Infectious diseases not covered by other categories (e.g., appendicitis) | 8 (4.0%)                                                                                    |
| Mental health                                           | 7 (3.5%)                                                                                     |
| Vaccine-preventable illnesses                           | 7 (3.5%)                                                                                     |
| Sexually transmitted diseases (HIV/AIDS not specifically included) | 6 (3.0%)                                                                                   |
| General medicine use                                    | 5 (2.5%)                                                                                     |
| Tuberculosis                                            | 5 (2.5%)                                                                                     |
| Child health (not covered by other categories, such was well-baby checks) | 3 (1.5%)                                                                                    |
| Heart disease                                           | 3 (1.5%)                                                                                     |
| Infection prevention                                    | 3 (1.5%)                                                                                     |
| Injuries and trauma                                     | 2 (1.0%)                                                                                     |
| Dental health                                            | 1 (0.5%)                                                                                     |
| Hypertension                                            | 1 (0.5%)                                                                                     |
| Non-malaria parasite                                    | 1 (0.5%)                                                                                     |
| Substance abuse                                         | 1 (0.5%)                                                                                     |
Table B4. Practice outcome categories of all 1200 effect sizes from the included studies

| Outcome                        | HCP practice outcome scale | Totals for percentage and continuous outcomes combined |
|--------------------------------|----------------------------|------------------------------------------------------|
|                                | Percentage | Continuous |                                                     |
| Assessment                     | 42 studies  | 1 study    | 42 studies                                           |
|                                | 54 comparisons | 1 comparison | 54 comparisons                                   |
|                                | 215 effect sizes | 1 effect size | 216 effect sizes                                |
| Case management<sup>a</sup>     | 53 studies  | 0 studies   | 53 studies                                           |
|                                | 61 comparisons | 0 comparisons | 61 comparisons                                   |
|                                | 121 effect sizes | 0 effect sizes | 121 effect sizes                                |
| Chemoprophylaxis               | 4 studies    | 0 studies   | 4 studies                                            |
|                                | 4 comparisons | 0 comparisons | 4 comparisons                                     |
|                                | 4 effect sizes | 0 effect sizes | 4 effect sizes                                    |
| Consultation time              | 0 studies    | 2 studies   | 2 studies                                            |
|                                | 0 comparisons | 2 comparisons | 2 comparisons                                     |
|                                | 0 effect sizes | 2 effect sizes | 2 effect sizes                                    |
| Counseling and communication   | 51 studies   | 2 studies   | 51 studies                                           |
|                                | 58 comparisons | 2 comparisons | 59 comparisons                                   |
|                                | 273 effect sizes | 6 effect sizes | 279 effect sizes                                |
| Diagnosis                      | 15 studies   | 0 studies   | 15 studies                                           |
|                                | 19 comparisons | 0 comparisons | 19 comparisons                                   |
|                                | 26 effect sizes | 0 effect sizes | 26 effect sizes                                |
| Documentation by HCP          | 13 studies   | 0 studies   | 13 studies                                           |
|                                | 15 comparisons | 0 comparisons | 15 comparisons                                   |
|                                | 41 effect sizes | 0 effect sizes | 41 effect sizes                                |
| Information accessed by HCP   | 1 study      | 0 studies   | 1 study                                             |
|                                | 1 comparison | 0 comparisons | 1 comparison                                     |
|                                | 5 effect sizes | 0 effect sizes | 5 effect sizes                                    |
| Patient dignity                | 2 studies    | 0 studies   | 2 studies                                            |
|                                | 2 comparisons | 0 comparisons | 2 comparisons                                     |
|                                | 5 effect sizes | 0 effect sizes | 5 effect sizes                                    |
| Referral                       | 14 studies   | 0 studies   | 14 studies                                           |
|                                | 17 comparisons | 0 comparisons | 17 comparisons                                   |
|                                | 30 effect sizes | 0 effect sizes | 30 effect sizes                                |
| Treatment                      | 111 studies  | 14 studies  | 111 studies                                          |
|                                | 142 comparisons | 16 comparisons | 143 comparisons                                   |
|                                | 427 effect sizes | 26 effect sizes | 455 effect sizes                                |
| Universal precautions by HCP  | 6 studies    | 0 studies   | 6 studies                                            |
|                                | 6 comparisons | 0 comparisons | 6 comparisons                                     |
|                                | 11 effect sizes | 0 effect sizes | 11 effect sizes                                    |
| Vaccination                    | 5 studies    | 0 studies   | 5 studies                                            |
|                                | 5 comparisons | 0 comparisons | 5 comparisons                                     |
|                                | 5 effect sizes | 0 effect sizes | 5 effect sizes                                    |
| Total                          | 197 studies  | 18 studies  | 197 studies                                          |
|                                | 238 comparisons | 20 comparisons | 240 comparisons                                   |
|                                | 1163 effect sizes | 37 comparisons | 1200 effect sizes                                |

Abbreviation: HCP = health care provider

<sup>a</sup> Outcomes that include multiple steps of the case-management pathway (e.g., correct diagnosis and treatment).
Section 2d. Assessment of publication bias

First, we performed a visual inspection of a funnel plot of the 78 study comparisons of a training strategy versus a no-intervention comparison group from studies of professional health care providers (Figure B). The effect size for a single study comparison was the median of effect sizes of all practice outcomes expressed as a percentage. Our interpretation was that asymmetry (a sign of potential publication bias) was possible but not clear. Second, we used the statistical test proposed by Egger to identify asymmetry (Egger et al. BMJ 1997; 315: 629–34). We fit the following model using ordinary least squares linear regression: the dependent variable was the standard normal deviate (i.e., the effect size divided by standard error) and the independent variable was the precision (i.e., 1/standard error). Evidence of possible publication bias was defined as a p-value < 0.1 of the model’s intercept. We found no evidence of asymmetry (intercept p-value = 0.65).

Figure B. Funnel plot of 78 study comparisons of a training strategy versus a no-intervention comparison group from studies of professional health care providers (results of percentage outcomes)
Section 2e. Effectiveness of training strategies

Table C. Effectiveness of training strategies on the practices of lay health care providers

| Strategies tested\(a\) | Outcome scale | No. of study comparisons (risk of bias: low, moderate, high, very high) | Median MES\(b\) (range) |
|--------------------------|---------------|-----------------------------------------------------------------------|-------------------------|
| Intervention arm         | Reference arm |                                                                       |                         |
| Group in-service training without computers | Controls | Percentage | 3 (0, 0, 1, 2) | -0.9 (-1.2, 5.6) |
| Educational outreach visits | Other strategy components | Percentage | 1 (0, 0, 1, 0) | 0.2 (NA) |
| Group in-service training without computers versus educational outreach visits | Group in-service training without computers | Percentage | 1 (0, 0, 1, 0) | 0.7 (NA) |
| Group pre-service training without computers | Controls | Percentage | 1 (0, 0, 0, 1) | 9.1 (NA) |

Abbreviations: MES = median effect size, NA = not applicable.

\(a\) See Boxes 1 and 2 in the main article for descriptions of the strategies and the comparisons, respectively.

\(b\) Effect sizes calculated as the intervention arm improvement minus reference arm improvement.
**Section 2f. Training attributes associated with training effectiveness for professional health care providers**

Table D1. Group in-service training attributes associated with training effectiveness: modeling results from studies of training only

| Training attribute or other predictor of effectiveness | Model 1: no predictors forced into the model | Model 2: baseline performance and time since training forced into the model | Model 3: on-site training forced into the model | Model 4: on-site training, baseline performance, and time since training forced into the model |
|-------------------------------------------------------|---------------------------------------------|-------------------------------------------------|-----------------------------------------------|-----------------------------------------------------------------|
|                                                       | β p-value                                   | β p-value                                       | β p-value                                     | β p-value                                                      |
| Intercept                                             | 31.0 0.0003                                 | 33.1 <0.0001                                   | 26.8 0.008                                    | 32.2 0.0001                                                    |
| Small training group size (2–14 trainees)             | –6.4 0.058                                  | –6.1 0.041                                     | –9.5 0.038                                    | –8.5 0.034                                                    |
| Trainers had content expertise                         | –16.1 0.041                                 | NA                                             | –15.1 0.099                                   | NA                                                            |
| Natural logarithm of training duration, in days        | NA                                          | –5.3 0.060                                     | NA                                           | –5.2 0.068                                                    |
| Training on multiple topics                           | NA                                          | –13.4 0.016                                   | NA                                           | –13.9 0.023                                                   |
| Interaction between “logarithm of training duration” and “multiple topics” | NA                                          | 11.4 0.006                                    | NA                                           | 10.9 0.012                                                    |
| Baseline performance                                   | NA                                          | –0.13 0.027                                   | NA                                           | –0.18b 0.013                                                 |
| Time since training, in months                        | NA                                          | –1.0 <0.0001                                   | NA                                           | –1.0 0.0001                                                  |
| On-site training                                      | NA                                          | NA                                             | 10.0 0.031                                    | 9.9 0.018                                                     |
| Adjusted R²                                           | 0.110                                       | 0.231                                          | 0.203                                         | 0.340                                                         |
| No. of observations missing                           | 35/372 (9.4%)                               | 23/372 (6.2%)                                 | 107/372 (28.8%)                              | 111/372 (29.8%)                                               |

Abbreviation: NA = Not applicable, which indicates that a predictor was not included in the model, either because the univariable p-value for a training attribute was ≥ 0.10 or because the predictor was a potential confounder that was not forced into the model.

a For example, a training on managing infectious diseases was taught by a physician with infectious diseases specialist.

b Although this result is statistically significant, it was not used because this model was only used to evaluate the effect of on-site training. Results for other factors (e.g., baseline performance) were not considered as valid as those from other models (without the on-site variable) because this model was based on a database with many missing values.

c At least some of the training was conducted where the health care provider routinely worked.
Table D2. Group-in-service training attributes associated with training effectiveness: modeling results from studies of training with or without supervision

| Training attribute or other predictor of effectiveness | Model 1: no predictors forced into the model | Model 2: baseline performance and time since training forced into the model | Model 3: on-site training forced into the model | Model 4: on-site training, baseline performance, and time since training forced into the model |
|-------------------------------------------------------|---------------------------------------------|-------------------------------------------------------------|-----------------------------------------------|------------------------------------------------------------------|
|                                                       | β               | p-value | β               | p-value | β               | p-value | β               | p-value |
| Intercept                                             | 20.8            | <0.0001 | 28.2            | <0.0001 | 17.0            | 0.0002 | 25.2            | <0.0001 |
| Supervision included in strategy                     | -5.7            | 0.053   | -5.1            | 0.078   | -4.5            | 0.15   | -2.8            | 0.40    |
| Small training group size (2–14 trainees)            | NA              |         | -5.8            | 0.035   | NA              |         | -8.8            | 0.013   |
| Training delivered over multiple sessions a           | -3.8            | 0.39    | NA              |         | -4.0            | 0.31   | NA              |         |
| Training duration, in days                            | -0.6            | 0.11    | -0.8            | 0.10    | -0.4            | 0.31   | -0.5            | 0.30    |
| Interaction between "supervision" and "time since training" | 1.1              | 0.011   | 1.0             | 0.019   | 0.9             | 0.036  | 0.7             | 0.14    |
| Baseline performance                                  | NA              |         | -0.11           | 0.025   | NA              |         | -0.14b          | 0.017   |
| Time since training, in months                       | -0.8            | 0.0003  | -0.8            | <0.0001 | -0.8            | 0.003  | -0.8            | 0.001   |
| On-site training c                                    | NA              |         | NA              |         | 7.7             | 0.059  | 10.4            | 0.007   |
| Adjusted R²                                           | 0.102           |         | 0.177           |         | 0.170           |         | 0.280           |         |
| No. of observations missing                           | 34/470 (7.2%)   |         | 45/470 (9.6%)   |         | 125/470 (26.6%) |         | 135/470 (28.7%) |         |

Abbreviation: NA = Not applicable, which indicates that a predictor was not included in the model, either because the univariable p-value for a training attribute was $\geq 0.10$ or because the predictor was a potential confounder that was not forced into the model.

a For example, a 4-day curriculum delivered via four separate 1-day sessions (e.g., four Mondays in a row).

b Although this result is statistically significant, it was not used because this model was only used to evaluate the effect of on-site training. Results for other factors (e.g., baseline performance) were not considered as valid as those from other models (without the on-site variable) because this model was based on a database with many missing values.

c At least some of the training was conducted where the health care provider routinely worked.
Table D3. Group in-service training attributes associated with training effectiveness: modeling results from studies of training with or without other strategy components

| Training attribute or other predictor of effectiveness | Model 1: no predictors forced into the model | Model 2: baseline performance and time since training forced into the model | Model 3: on-site training forced into the model | Model 4: on-site training, baseline performance, and time since training forced into the model |
|-------------------------------------------------------|---------------------------------------------|-------------------------------------------------|---------------------------------------------|------------------------------------------------------------------------------------------|
|                                                       | β   | p-value | β   | p-value | β   | p-value | β   | p-value |
| Intercept                                             | 13.7 | <0.0001 | 19.6 | <0.0001 | 11.5 | <0.0001 | 17.9 | <0.0001 |
| Community support included in strategy                | -0.3 | 0.94    | 5.6  | 0.12    | 2.2  | 0.67    | 7.2  | 0.12    |
| Patient support included in strategy                  | -5.4 | 0.15    | -3.2 | 0.37    | -7.6 | 0.059   | -4.5 | 0.23    |
| Strengthening infrastructure included in strategy     | 2.6  | 0.55    | 3.8  | 0.38    | 4.2  | 0.36    | 4.5  | 0.32    |
| Health system financing or other incentives included in strategy | 2.7  | 0.61    | 2.0  | 0.73    | 3.5  | 0.52    | 1.4  | 0.82    |
| Regulation or governance included in strategy         | 2.8  | 0.59    | -2.7 | 0.59    | 1.9  | 0.71    | -1.5 | 0.80    |
| Group problem solving included in strategy            | 12.0 | 0.12    | 14.8 | 0.055   | 10.1 | 0.19    | 13.0 | 0.083   |
| Supervision included in strategy                      | 0.8  | 0.68    | -0.1 | 0.95    | -0.1 | 0.97    | -0.6 | 0.73    |
| Other management techniques included in strategy      | 6.6  | 0.12    | 7.0  | 0.078   | 8.8  | 0.032   | 7.5  | 0.066   |
| Information and communication technology included in strategy | -2.1 | 0.74    | -1.9 | 0.75    | -4.0 | 0.52    | -3.8 | 0.51    |
| Training included clinical practice for health care providers | 6.9  | 0.013   | 7.4  | 0.0068  | 5.6  | 0.056   | 6.4  | 0.029   |
| Baseline performance                                  | NA   | -0.15   | 0.0001 | NA | -0.16a | 0.0001 |
| Time since training, in months                        | NA   | -0.2    | 0.14  | NA | -0.2   | 0.27   |
| On-site training                                      | NA   | 5.0     | 0.092 | 6.0 | 0.045  |
| Adjusted R²                                           | 0.106 | 0.199   |          | 0.153 | 0.270 |
| No. of observations missing                           | 0/856 (0%) | 59/856 (6.9%) | 160/856 (18.7%) | 203/856 (23.7%) |

Abbreviation: NA = Not applicable, which indicates that a predictor was not included in the model, either because the univariable p-value for a training attribute was ≥ 0.10 or because the predictor was a potential confounder that was not forced into the model.

a Although this result is statistically significant, it was not used because this model was only used to evaluate the effect of on-site training. Results for other factors (e.g., baseline performance) were not considered as valid as those from other models (without the on-site variable) because this model was based on a database with many missing values.

b At least some of the training was conducted where the health care provider routinely worked.
Table E. Associations of group in-service training attributes on training effectiveness for the practices of professional health care providers: detailed results

| Finding                                                                 | Supporting evidence                                                                 |
|------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Attributes associated with training effectiveness based on > 1 study    | **Direct evidence (results of head-to-head studies)**                                |
| Effect of training when some or all training was done on-site (where HCPs routinely work) was greater than when all training was done off-site by 6.0 to 10.4 %-points | • None. No head-to-head study examined this attribute.                                |
| Effect of training that used clinical practice as a training method was greater than training that did not use this method by 6.9 to 7.4 %-points | **Indirect evidence (model results)**                                                |
|                                                                        | • Some or all training on-site was more effective than all training off-site by a mean of: |
|                                                                        |   ➢  9.9 %-points (Table D1, Model 4, row 9: $\beta = 9.9$, $p = 0.018$)               |
|                                                                        |   ➢  10.0 %-points (Table D1, Model 3, row 9: $\beta = 10.0$, $p = 0.031$)             |
|                                                                        |   ➢  10.4 %-points (Table D2, Model 4, row 9: $\beta = 10.4$, $p = 0.007$)             |
|                                                                        |   ➢  6.0 %-points (Table D3, Model 4, row 14: $\beta = 6.0$, $p = 0.045$)              |
|                                                                        | **Direct evidence (results of head-to-head studies)**                                |
|                                                                        | • None. No head-to-head study examined this attribute.                                |
|                                                                        | **Indirect evidence (model results)**                                                |
|                                                                        | • Training with clinical practice was more effective than training without clinical practice by a mean of: |
|                                                                        |   ➢  6.9 %-points (Table D3, Model 1, row 11: $\beta = 6.9$, $p = 0.013$)              |
|                                                                        |   ➢  7.4 %-points (Table D3, Model 2, row 11: $\beta = 7.4$, $p = 0.0068$)             |
### Direct evidence (results of head-to-head studies)
- Studies that compared training +/- other components versus training + supervision +/- other components (to estimate the marginal effect of supervision given training +/- other components) with at least two post-intervention measures (to allow an assessment of marginal effect over time) found (based on ordinary least squares linear regression modeling) that the change in the marginal effect of supervision over time was 0.3 %-points per month, although this change was not statistically significant ($p = 0.58$) ($N = 3$ MES from 2 studies with a total of 56 post-intervention measures; no. of comparisons with ROB L/M/H/VH: 0/3/0/0, and no. of studies with ROB L/M/H/VH: 0/2/0/0; study follow-up time ranged from 0.5 to 5.5 months). While this analysis does not provide strong support for the interaction (ideally, one would want a marginal effect that significantly increases over time up to at least 20 months after training, which corresponds to the increasing distance between the two lines in Figure C), at least the results match what is predicted by the models of true-control studies (see below) for 0.5 to 5.5 months after training.

### Indirect evidence (model results)
- Effect of training alone decreased over time since training; but the effect of training combined with supervision did not decrease over time.
  - Table D2, Model 1 (row 6, $\beta_{\text{interaction}} = 1.1$, $p = 0.011$)
  - Table D2, Model 2 (row 6, $\beta_{\text{interaction}} = 1.0$, $p = 0.019$)
- Effect of training alone (without supervision) decreased over time
  - Table D1, Model 2 (row 8, $\beta = -1.0$ %-points per month, $p < 0.0001$). If one assigns mean values for other variables in the model (baseline performance = 42.7%, $\ln$[training duration] = 1.187, proportion of effect sizes from studies of training with multiple topics = 0.45, and proportion of effect sizes from studies of training with small group size = 0.25), then the training effect size is predicted to reach zero after 19.8 months, on average.
  - Table D2, Model 1 (row 8, $\beta = -0.8$ %-points per month, $p = 0.0003$). If one assigns mean values for other variables in the model (training duration = 4.19 days, proportion of effect sizes from studies of training over multiple sessions = 0.14), then the training effect size is predicted to reach zero after 22.5 months, on average.
  - Table D2, Model 2 (row 8, $\beta = -0.8$ %-points per month, $p < 0.0001$). This regression model is represented by the green line in Figure C. If one assigns mean values for other variables in the model (baseline performance = 40.2%, training duration = 4.0 days, proportion of effect sizes from studies of training with small group size = 0.33), then the training effect size is predicted to reach zero after 22.0 months, on average.
- Effect of training combined with supervision did not decrease over time
  - Table D2, Model 1. If one assigns mean values for other variables in the model (training duration = 4.19 days, proportion of effect sizes from studies of training over multiple sessions = 0.14), then: effect size = 12.2 + (0.3 x time). $P$-value for time = 0.37.
  - Table D2, Model 2. If one assigns mean values for other variables in the model (baseline performance = 40.2%, training duration = 4.0 days, proportion of effect sizes from studies of training with small group size = 0.33), then: effect size = 13.6 + (0.2 x time). $P$-value for time = 0.64. This regression model is represented by the blue line in Figure C.

### Mean effect of training increased by 1.1 to 1.5 %-points for every 10 %-point decrease in baseline performance level
- None. No head-to-head study examined this attribute.

### Direct evidence (results of head-to-head studies)
- None. No head-to-head study examined this attribute.

### Indirect evidence (model results)
- Mean effect of training decreased as baseline performance level increased
  - From Table D1, Model 2, row 7: $\beta = -0.13$ %-points per 1 %-point increase in baseline performance level, $p = 0.027$.
  - From Table D2, Model 2, row 7: $\beta = -0.11$ %-points per 1 %-point increase in baseline performance level, $p = 0.025$.
  - From Table D3, Model 2, row 12: $\beta = -0.15$ %-points per 1 %-point increase in baseline performance level, $p = 0.0001$. 

---

**Effect of training alone decreased over time since training, but the effect of training combined with supervision did not decrease over time**

**Mean effect of training increased by 1.1 to 1.5 %-points for every 10 %-point decrease in baseline performance level**

---

Rowe AK, et al. BMJ Global Health 2021; 6:e003229. doi: 10.1136/bmjgh-2020-003229
| Attributes associated with training effectiveness based on only 1 study (i.e., interpret with caution) | Direct evidence (results of head-to-head studies)                                                                 |
|-------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| Training tailored to HCPs’ stage of readiness to change was more effective than non-tailored training by 23.3 % points | • Training tailored to HCPs’ stage of readiness to change was more effective that non-tailored training by a median of 23.3 % points (N = 2 MES from 1 study; range: 11.0, 36.5; no. of comparisons with ROB L/M/H/VH: 0/0/2/0, and no. of studies with ROB L/M/H/VH: 0/0/1/0; average study follow-up time: 2 months). |
|                                                                                                  | **Indirect evidence (model results)**                                                                                   |
|                                                                                                  | • None. No modeling results because the HCPRR database did not include this attribute.                                 |

| Attributes with a small or no association with training effectiveness (i.e., magnitude < 5 % points) | Direct evidence (results of head-to-head studies)                                                                 |
|-------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| Effects of training with computers and without computers were similar                           | • Training without computers was slightly more effective than training with computers by a median of 0.7 % points (N = 2 MES from 2 studies; range: -1.2, 2.5; no. of comparisons with ROB L/M/H/VH: 2/0/0/0, and no. of studies with ROB L/M/H/VH: 2/0/0/0; average study follow-up time: 2.5 months). |
|                                                                                                  | **Indirect evidence (model results)**                                                                                   |
|                                                                                                  | • None. Attribute was not assessed by modeling because of highly unbalanced data.                                      |

| Training delivered by an in-person trainer was slightly more effective than distance training via live video interactive sessions by 3.6 % points | Direct evidence (results of head-to-head studies)                                                                 |
|-------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
|                                                                                                  | • Training delivered by an in-person trainer was slightly more effective than distance training via live video interactive sessions by 3.6 % points (N = 1 MES from 1 study; no. of comparisons and studies with ROB L/M/H/VH: 0/1/0/0; study follow-up time: 0.03 months).  
• Note that "in-person training + in-person supervision" was more effective than "distance learning training + distance supervision", by 5.85 % points (N = 1 MES). However, this was not a clean comparison of distance versus in-person training because of different supervision approaches. |
|                                                                                                  | **Indirect evidence (model results)**                                                                                   |
|                                                                                                  | • None. No modeling results because the HCPRR database did not include this attribute.                                 |

| Effects of training delivered over one session versus multiple sessions were similar             | Direct evidence (results of head-to-head studies)                                                                 |
|-------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
|                                                                                                  | • Training delivered in one (8 day) session was slightly more effective than 5 days of training over multiple sessions by 0.6 % points (N = 1 MES from 1 study; no. of comparisons and studies with ROB L/M/H/VH: 0/0/0/1; study follow-up time: 34.5 months [note that, given the likely decay of training effect over time, the long follow-up time of this study greatly limits its utility]). |
|                                                                                                  | **Indirect evidence (model results)**                                                                                   |
|                                                                                                  | • No significant association. The one univariable \( \beta \) value was -3.8 % points (p = 0.39) (Table D2, Model 1, row 4), which suggests that training over multiple sessions was slightly less effective than training delivered in one session. |
Seven other attributes were not significantly associated with training effectiveness

Direct evidence (results of head-to-head studies)
- None. No head-to-head study examined these attributes, except for duration, which had no clear pattern (see last row of this table).

Indirect evidence (model results)
- All the following attributes had univariable p-values > 0.1.
  - Training duration in days (as continuous variable, and coded as > 10 days versus ≤ 10 days, and > 13 days versus ≤ 13 days)
  - Training used role play
  - No. of educational methods used (continuous, including clinical practice, interactive sessions, non-interactive sessions, role play, and other method)
  - Training used both interactive sessions and non-interactive lectures
  - Training used written materials
  - Trainers with pedagogical training
  - Topic complexity

Attributes with an unclear association with training effectiveness because direct and indirect evidence was contradictory

Effect of trainee group size is unclear because direct and indirect evidence had contradictory results

Direct evidence (results of head-to-head studies)
- Small group training (i.e., 2–14 participants) was somewhat more effective than large group training (i.e., > 14 participants) by a median of 5.3 %-points (N = 4 MES from 3 studies; range: –6.5, 18.0; no. of comparisons with ROB L/M/H/VH: 0/0/3/1, and no. of studies with ROB L/M/H/VH: 0/0/2/1; average study follow-up time: 2.7 months)

Indirect evidence (results of model results)
- Large group training was somewhat more effective than small group training by a mean of 6.1 %-points (Table D1, Model 2, row 2: β = –6.1, p = 0.041 [note that the reference group was large group size, so the negative β value of –6.1 means that small group was less effective than large group])
- Large group training was somewhat more effective than small group training by a mean of 5.8 %-points (Table D2, Model 2, row 3: β = –5.8, p = 0.035 [note that the reference group was large group size, so the negative β of –5.8 means that small group was less effective than large group])

Effect of having trainers with content expertise is unclear because direct and indirect evidence had contradictory results

Direct evidence (results of head-to-head studies)
- Training by trainers with content expertise (doctors) was slightly better than training by trainers without content expertise (paramedics) by 2.5 %-points (N = 1 MES from 1 study; no. of comparisons with ROB L/M/H/VH: 0/0/1/0, and no. of studies with ROB L/M/H/VH: 0/0/1/0; average study follow-up time: 0.03 months)

Indirect evidence (model results)
- Training when “all trainers were content experts” was lower than when not all trainers were content experts by a mean of 16.1 %-points (Table D1, Model 1, row 3: β = –16.1, p = 0.041 [note that the reference group was not having all trainers who were content experts, so the negative β value of –16.1 means that having all trainers who were content experts was less effective than not having all trainers who were content experts])
Effect of training with non-interactive lectures is unclear because direct and indirect evidence had contradictory results

| Direct evidence (results of head-to-head studies) |
|--------------------------------------------------|
| - Among “training A only versus training B only” study comparisons, training with a non-interactive lecture or session was better than interactive training by 4.3 %-points (N = 1 MES from 1 study; no. of comparisons and studies with ROB L/M/H/VH: 0/0/1/0). |
| - Among “training A only versus training B only” study comparisons and “training A + other strategy components X versus training B + other strategy components X” studies, training with a non-interactive lecture or session was better than interactive training, by median of 5.0 %-points (N = 2 MES from 2 studies; non-interactive minus interactive differences: 4.3 and 5.7 %-points; no. of comparisons and studies with ROB L/M/H/VH: 0/0/2/0). |

Indirect evidence (model results)
- No significant association. Variable in modeling analysis included any interactive training method (i.e., interactive session, clinical practice, or role play). All univariable β values were less than 5.0 %-points (range: -2.5 to 3.8 %-points; all non-significant, with p-value ranging from 0.15 to 0.95).

Effect of an interaction between the natural logarithm of training duration and topic complexity of the training (single topic versus multiple topics) is unclear because direct and indirect evidence had contradictory results

| Direct evidence (results of head-to-head studies) |
|--------------------------------------------------|
| - For training on single topics, training effectiveness might have increased with course duration: a 5-day course was more effective than a 3-day course by 8.7 %-points, a 3-day course (plus peer education, supplies, and incentives) was more effective than a 1-day course (plus peer education, supplies, and incentives) by 13.0 %-points, and a 2-day course (plus peer education) was as effective as a 1.5-day course (plus peer education) with a difference of 0.8 %-points. |
| - For training on multiple topics, training effectiveness seemed unrelated to course duration: an 11-day course was as effective as a 6-day course (difference of 0.3 %-points), and an 8-day course was as effective as a 5-day course (difference of 0.6 %-points). |

Indirect evidence (model results)
- For training on single topics, training effectiveness was unrelated to course duration; but for training on multiple topics, effectiveness increased with longer course duration (Table D1, Model 2, rows 4–6, β_{interaction} = 11.4, p = 0.006). If one assigns mean values for other variables in the model (average baseline performance [42.7%], average follow-up time [5.15 months], and proportion of effect sizes from small-group-size trainings [0.252]), then:
  - For training on single health topics: effect size = 20.9 – (5.3 x ln[course duration in days]). P-value for ln[training duration] = 0.06 (Table D1, Model 2, row 4).
  - For training on multiple health topics: effect size = 7.5 + (6.1 x ln[course duration in days]). P-value for ln[training duration] = 0.0475.

Abbreviations: %-points = percentage points, HCP = health care provider, HCPPR = The Health Care Provider Performance Review, MES = median effect size(s), ROB L/M/H/VH = “risk of bias categories: low/moderate/high/very high”.

25
Figure C. The effect of in-service training over time since training, stratified by the presence of supervision, among studies of professional health care providers with training less than 20 days (analysis of the database of training with or without supervision)

Note 1. Predicted effect sizes assumed average baseline (40.2%), proportion of effect sizes from small-group trainings (0.33), and training duration (4 days).

Note 2. These modeling results were sensitive to the two studies with 27-month follow-up times (one study with supervision and another study without supervision, both with a high risk of bias). When these two studies were removed: a) the time trend for the “no supervision” group was $-0.93$ %-points per month ($p = 0.018$), b) the time trend for the “supervision present” group was $-0.87$ %-points per month ($p = 0.055$), and c) the interaction term had a value of 0.06 ($p = 0.92$).
Table F. Associations of pre-service training attributes on training effectiveness for the practices of professional health care providers: detailed results

| Finding                                                                 | Supporting evidence                                                                 |
|------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Attributes associated with training effectiveness based on only 1 study (i.e., interpret with caution) | Direct evidence (results of head-to-head studies)  
  • Pre-service training with group feedback about pre-training evaluation results was more effective than with individual feedback by 19.0 %-points (N = 1 MES from 1 study; no. of comparisons and studies with ROB L/M/H/VH: 1/0/0/0; study follow-up time: 0.03 months)  

  Indirect evidence (model results)  
  • None. Modeling was not performed for pre-service because there were too few studies. |

Abbreviations: %-points = percentage points, MES = median effect size(s), ROB L/M/H/VH = “risk of bias categories: low/moderate/high/very high”. 

---

Rowe AK, et al. BMJ Global Health 2021; 6:e003229. doi: 10.1136/bmjgh-2020-003229
Figure D. Distributions of effect sizes and median effect sizes for percentage and continuous outcomes for comparisons of group in-service training alone versus controls