

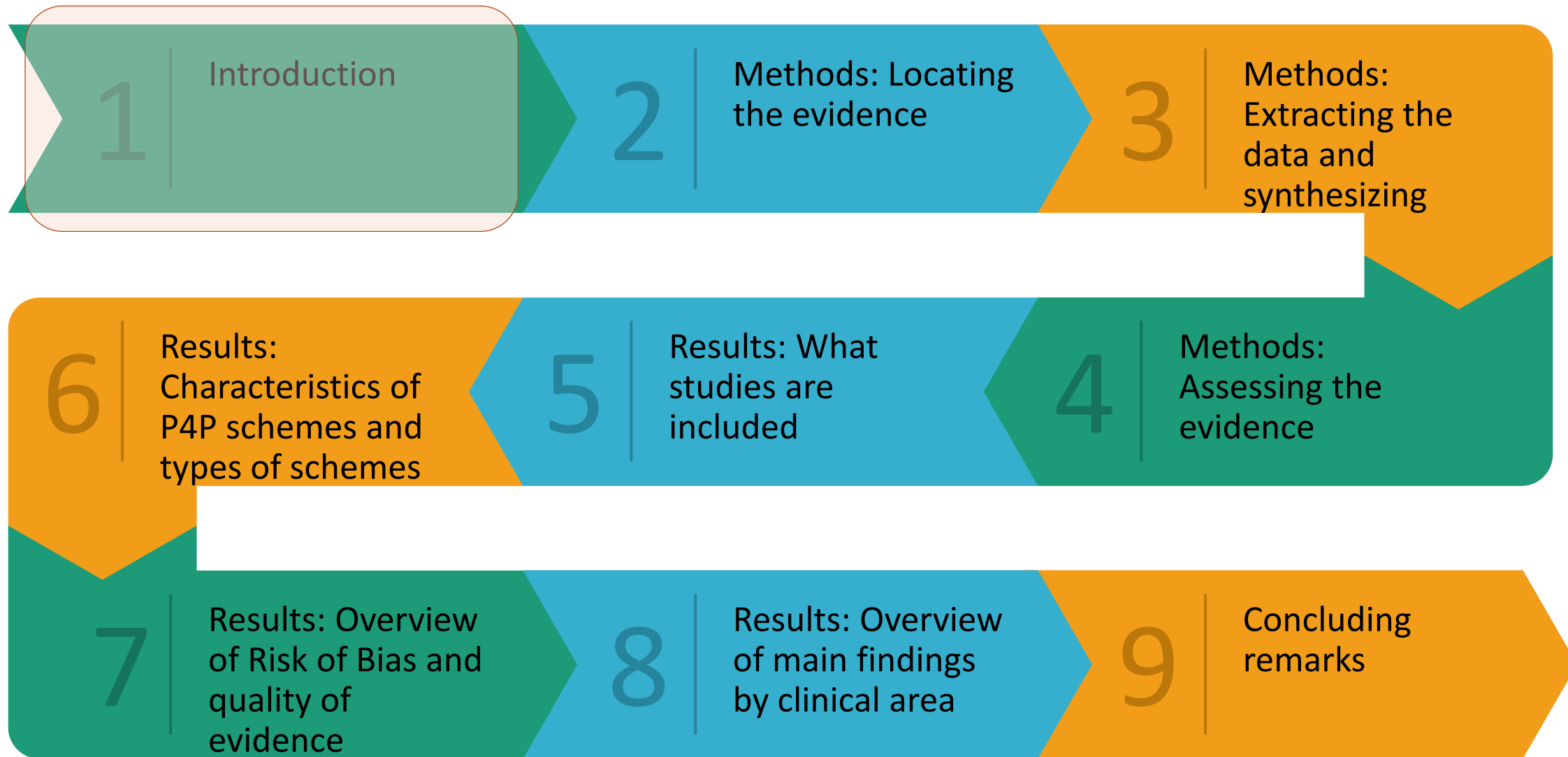
Paying for performance to improve the delivery of health interventions in LMICs: findings from updated Cochrane review

Professor Sophie Witter and Dr Karin Diaconu



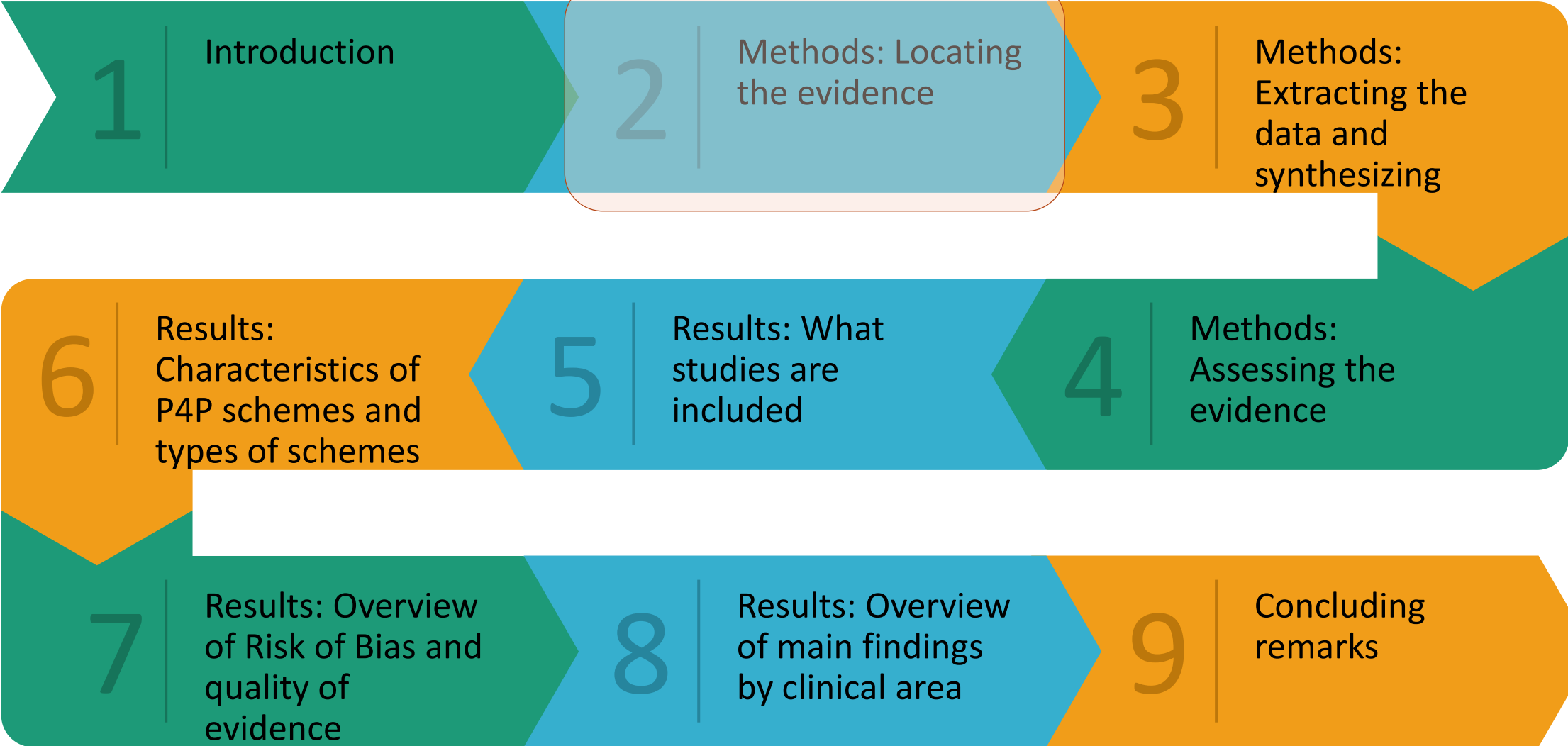
Queen Margaret University

INSTITUTE FOR GLOBAL HEALTH
AND DEVELOPMENT



Introduction

- Continued growth in field – programmes and evaluations
- 2012 review found limited high-quality evidence on effectiveness
- Broadly same methods adopted as per 2012 review
- Challenge of complexity of interventions, contexts and study designs, but we try to reflect this, given greater evidence available in this round, in:
 - Sub-group analysis by P4P design type
 - Comparisons against standard control and comparator interventions
 - Results for targeted versus untargeted indicators
 - Comparing RCT with non-RCT results
- Not all covered today, so do read the review!



Overview of process

- Searches: QMU Team and Cochrane EPOC group
- Study selection: in duplicate, with third reviewer arbitrating
- Data extraction and risk of bias: in duplicate, with quality check
- Data synthesis: two reviewers, with team overseeing
- Summary of finding and overall assessment of quality of evidence: two reviewers, with team overseeing

Searches (2018; updated 2020)

Bibliographic databases

- Cochrane Central Register of Controlled Trials
- Medline, Embase, PsychINFO, Econlit, LILACS, WHOLIS, CINAHL
- 3IE, BLDS, Global Health
- Trial registries: ICTRP, ClinicalTrials.gov

Grey literature databases

- Open Grey
- Grey Literature Report

Websites and other grey literature

- the World Bank, RBFHEALTH, the African Development Bank, the Inter-American Development Bank, U.S. Agency for International Development (USAID), Cordaid, Management Sciences for Health (MSH), Centre for Global Development, World Health Organization (WHO), Swiss Tropical and Public Health Institute (Swiss TPH), Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), KfW Entwicklungsbank, Department for International Development (DFID), The Global Alliance for Vaccines and Immunization (GAVI), The Global Fund to Fight AIDS, Tuberculosis and Malaria, Asian Development Bank and Pan American Health Organization (PAHO).
- Websites: London School of Hygiene and Tropical Medicine, the Harvard School of Public Health, University of Cape Town, Institute of Policy Studies of Sri Lanka (IPS), the Kenya Institute of Policy Analysis and Research (IPAR) and Institute of Tropical Medicine, Belgium, University of Heidelberg, University of Bergen and University of Rotterdam

Inclusion principles

Criteria	What we included
Type of study	<ul style="list-style-type: none">• Randomised and non-randomised trials• Controlled before-after (CBA) studies where: at least two clusters are included in each comparison group; pre and post intervention periods for study and control groups are the same; choice of the control site is appropriate.• Interrupted time series (ITS) studies with at least three measurements before and after introducing the intervention
Type of participant	providers of healthcare services (health workers and facilities), sub-national organisations (health administrations, non-governmental organisations or local governments), national governments and combinations of these. All sectors: public, private and private not-for-profit

Inclusion principles (2)

Criteria	What we included
Type of intervention	<ul style="list-style-type: none">• Conditional cash payment• Conditional provision of material goods• Target payments (payments for reaching a certain level of coverage, which can be defined in absolute terms or relative to a starting point)
Type of comparison	Any alternative (including non-conditional financial incentives and different levels of conditional financial incentives); includes alternatives where there are differences in ancillary components/P4P designs
Type of outcome	<p>Primary: health outcomes, changes in targeted measures of provider performance, unintended effects, changes in resource use</p> <p>Secondary: provider motivation, satisfaction, absenteeism and acceptability; patient satisfaction and acceptability; overall financing or resource allocation; management or information systems; equity of service delivery/utilization.</p>

What is excluded

- Studies where P4P run in parallel with with a demand-side intervention without explicit untangling of effects
- Demand-side interventions (CCT)
- Payment to health workers or facilities not explicitly linked to changing patterns of performance (e.g. for coming to work; salary increases; routine increases in activity-based payments such as Diagnosis-Related Groups (DRGs) or fees for service; or changes to budget flows which are routine or intended to motivate, but without being conditional on specific activity or output measures)



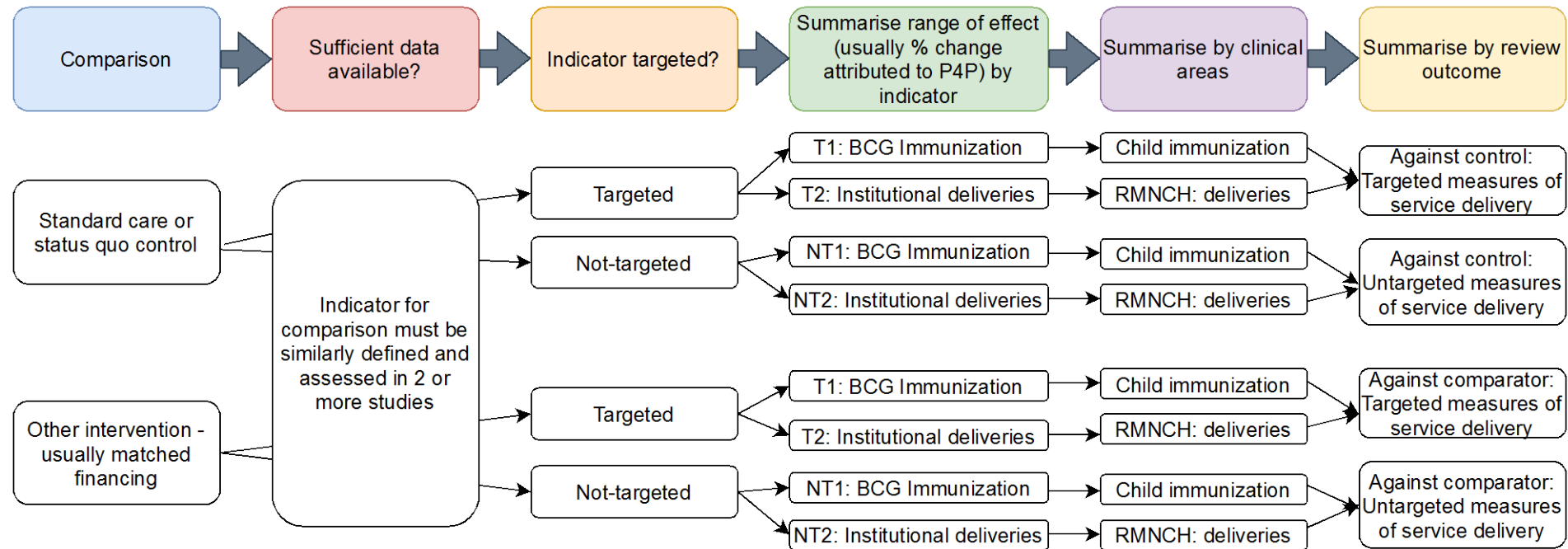
Data extraction

One reviewer extracting based on pre-determined form, full check by second reviewer.

Data extracted on:

- P4P scheme: design, targeted sectors and level, scope, funding source, incentive magnitude, verification and ancillary components
- Study setting, design & methods (unit of allocation, analysis method, data source, power calculations)
- Study participants: targets of the P4P scheme and for the impact evaluation
- Outcome measures and associated results (at indicator level as reported in evaluations)

Narrative synthesis (reported per SWiM guidelines)



We do not produce meta-estimates, instead indicating range of effect and judgment on overarching direction of effects

Is P4P yielding desirable, neutral, undesirable or uncertain effects?

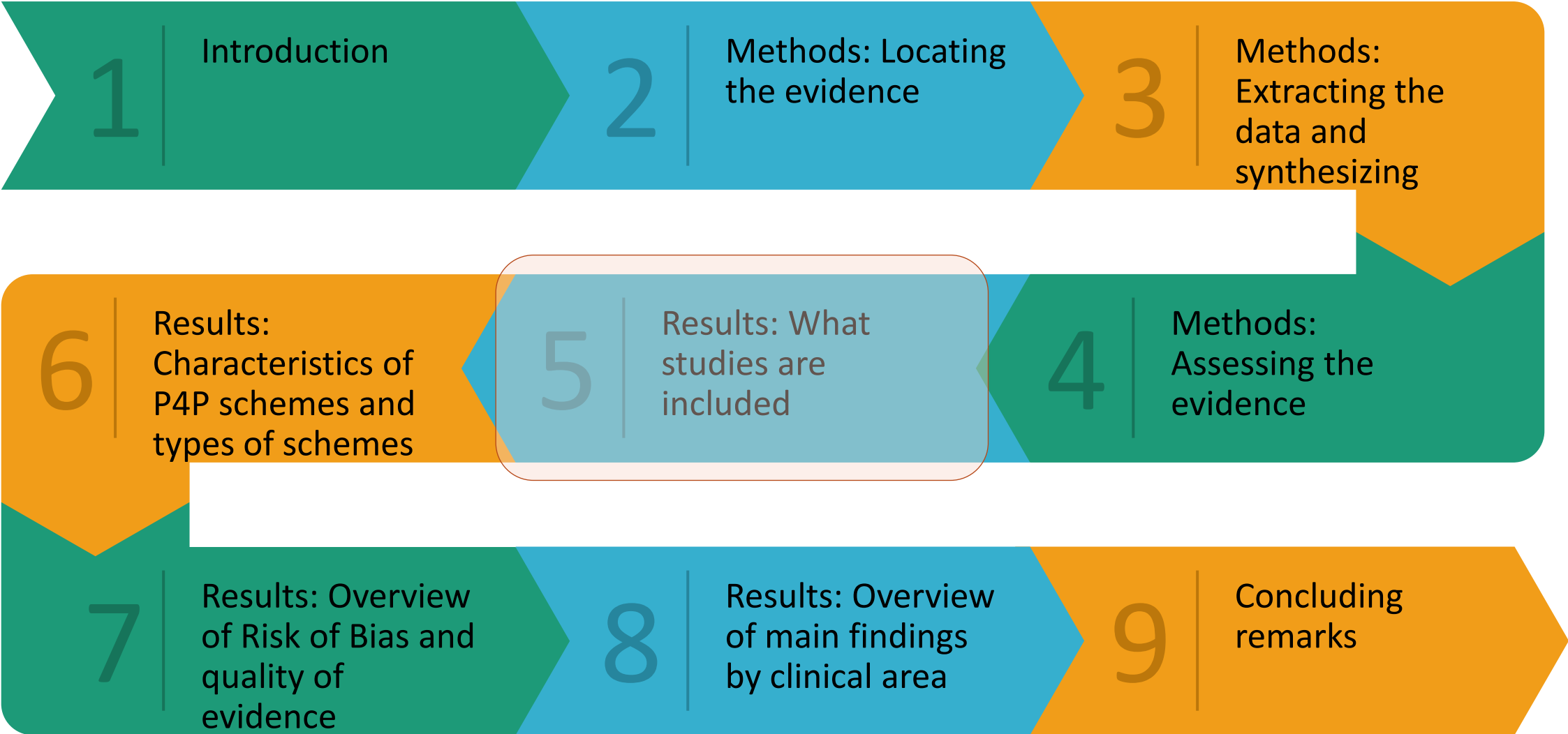
Narrative synthesis (reported per SWiM guidelines)

- Indicate range of effect and judgment for each outcome on whether effects of the intervention are:
 - Desirable: consistently positive and over 5%
 - Neutral: under 5%
 - Undesirable: consistently negative and over 5%
 - Uncertain: where either the quality of the evidence or the effects themselves are too varied to judge
- 5% threshold is contextualized – i.e. for health outcomes we do not adopt this, but for other measures (e.g. utilization) we do

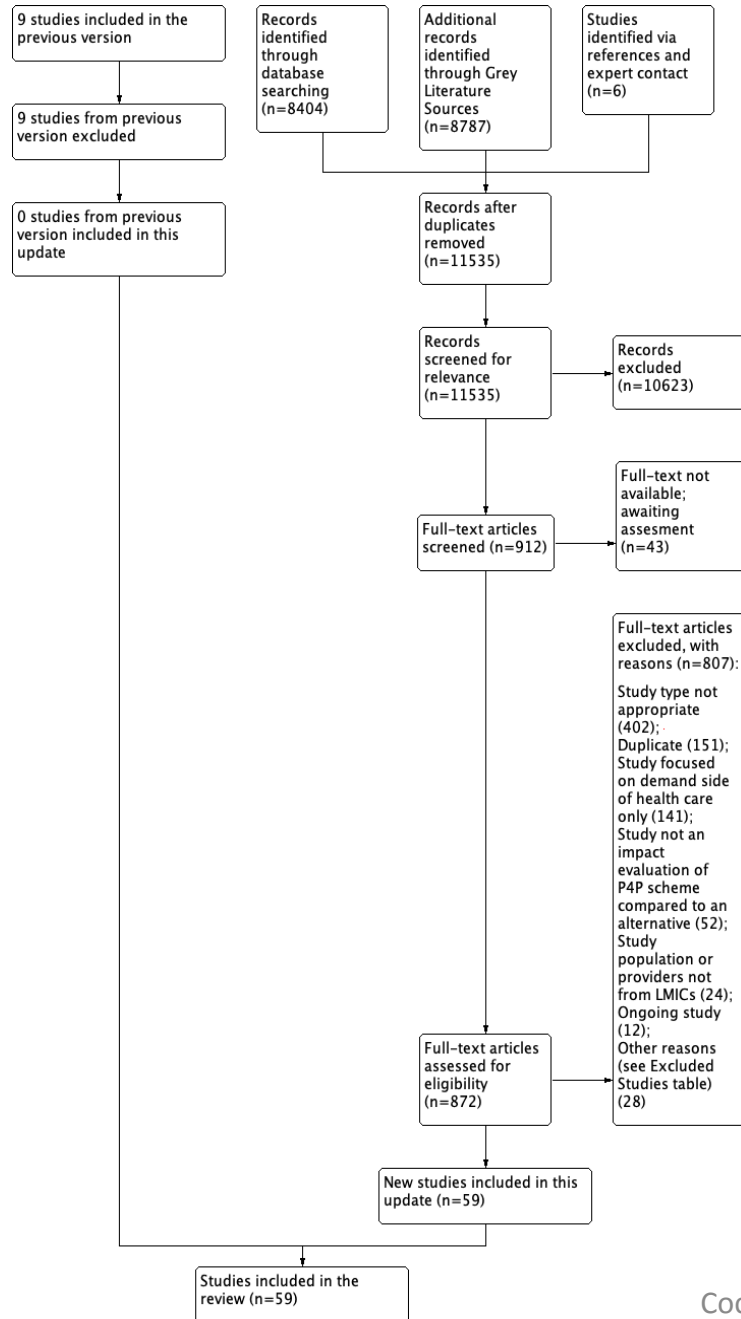


Risk of Bias assessment and GRADE

- Each individual study and outcome assessed for RoB as per Cochrane
- GRADE: assessed evidence as high, moderate, low, and very low -> we proceeded stepwise and downgraded from high
- Established criteria (risk of bias, inconsistency of results, imprecision, indirectness, and publication bias)
- Our GRADE assessment corresponds to an assessment of certainty in the overall direction of effect of the intervention.



Results of the search



Screened 11,535 unique documents for inclusion

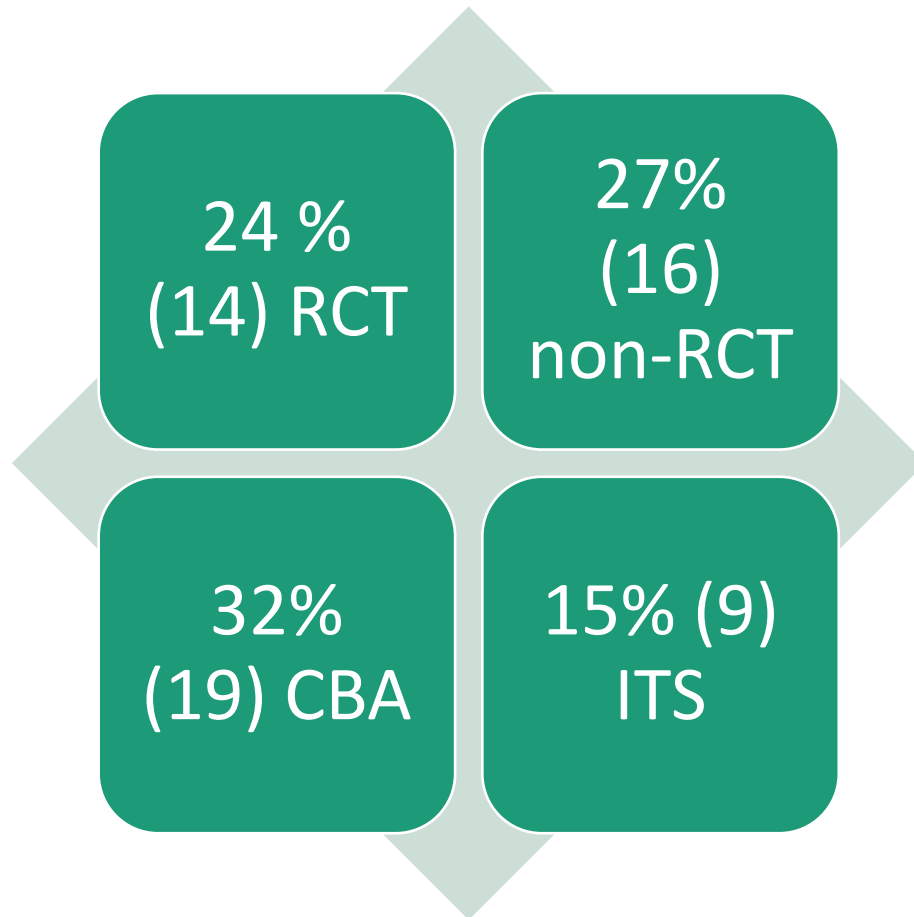
10,623 immediately not relevant, but 872 full text screened

Exclusions:

- 402 where study type did not meet criteria
- 151 duplicates that crept in
- 141 where intervention focused on demand-side only
- 52 where P4P was not evaluated against any alternative
- 24 not in LMICs
- 28 did not meet the detailed criteria for study type (e.g. not adjusting for clustering, too few clusters etc.)

59 studies included

Types of studies included



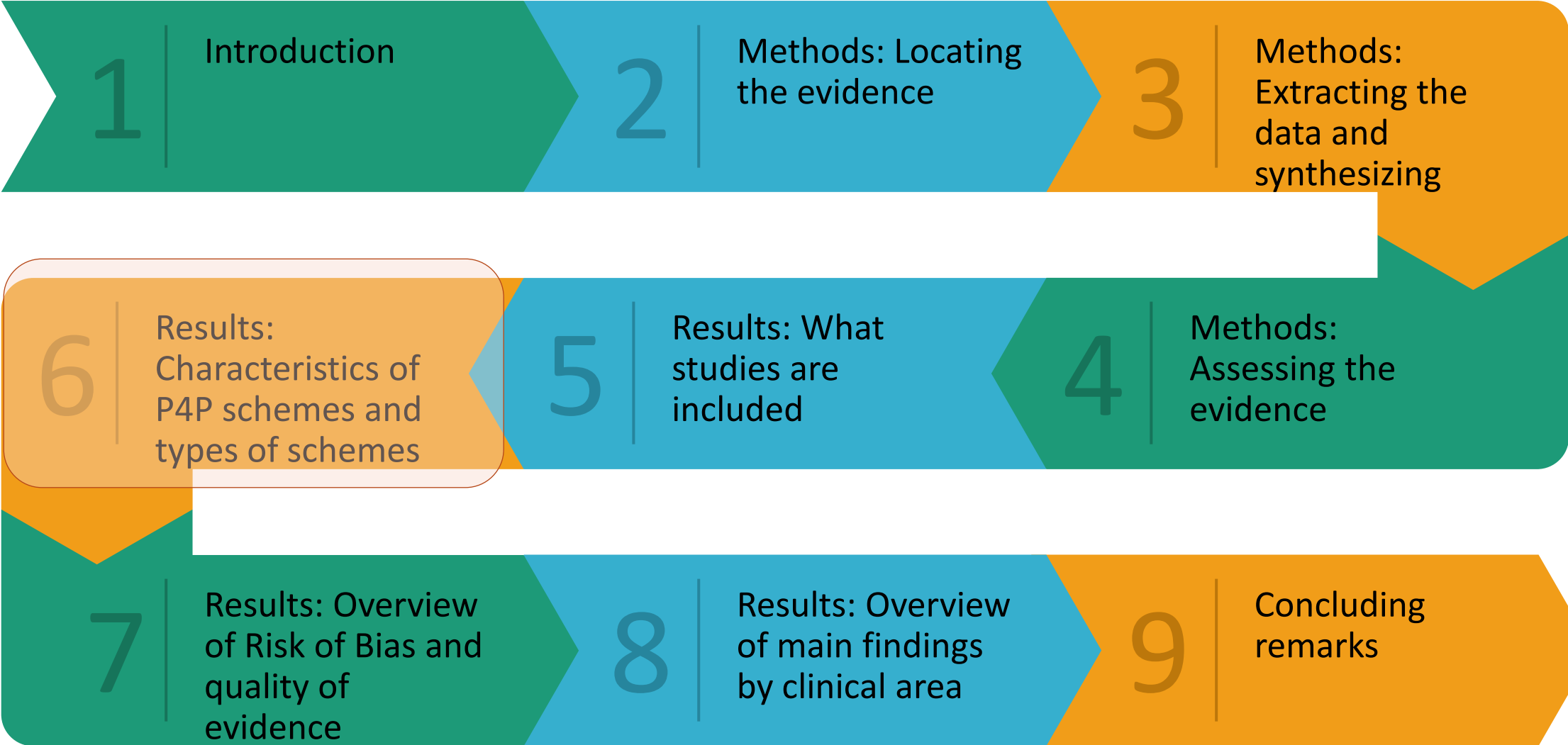
RCT = randomized controlled trial;
CBA = controlled before and after study;
ITS = interrupted time series;
1 study additionally both ITS + CBA

42 studies reporting effects against standard care or status quo, no change

13 report effects against an enhanced financing control / other financing modality or alternative

4 report against both standard care and enhanced financing

On average, studies report effects of the P4P scheme at 3 years, but this varies widely (from 1-17 years in cases)



Geography, clinical area and location of care

- Geography:
 - Included P4P schemes across 25 countries, majority Rwanda (17%, n=10), China (12%, n=7) and Tanzania (8.4%, n=5).
 - Location both urban and rural in 29% (n=18), with 2 studies focused on urban environments only.
- Clinical area and location of care:
 - Approx. half schemes focused on reproductive, maternal and child health services only; eight schemes differed (e.g. TB and HIV)
 - Location of care: 61% (n=36) of schemes operating at both in- and out-patient levels; 15% focused on outpatient care (n=9) or inpatient care (n=9) respectively; 2 studies community-based care

Participant facilities and scale of P4P schemes

- 91% (n=54) of studies reported on P4P schemes involving public or not for profit facilities (usually faith-based).
- Scale of intervention: highly variable
 - 21% (n=13) national roll-out
 - 42% (n=26) implemented across a range of districts (e.g. Cameroon)
 - 20% (n=12) focused on one province (e.g. Ningxia China)
 - 13% (n=8) focused on one health facility.

Funders of schemes

- Overall trends:
 - 37% by national governments and Ministries;
 - 33% by external agencies
 - 5% co-financed national + donor
 - No schemes without some level of national support
 - No schemes only by sub-national or local funds
- The World Bank and Government of Norway supported 11 (19%) and 5 (7%) schemes respectively.
- For the majority of P4P schemes described across studies (76%, n=45), purchasing arrangements were integrated into the national purchasing functions of the relevant Ministry of Health.

Target setting and incentive payments

- On average, schemes targeted approximately 8-12 core indicators, which related to the delivery or utilization of services, but high variability, with some schemes including both service + quality targets.(some schemes over 100 indicators assessed)
- Over half of included studies (57%, n=34) did not include details on why and how indicators were chosen and set.
- Magnitude of incentives
 - Range between 0.5-10 US\$/per indicator. Indicators which require repeat contact with the health service, or imply specialist skills, priced higher – e.g. correct tuberculosis patient management at 20US\$/patient (e.g. in [Bonfrer 2014a](#)).
 - Relative magnitude: over half of studies do not report. From what is available: P4P funds = 14-50% of all facility funding (10 studies), incentives = 1-78% of health worker salary, mostly around 10% (14 studies).

Scheme classification (1)

Scheme classification	Details on scheme	Countries (n)	Study types (n)	Comparators (n)
Capitation and PBF	Payment reforms including capitation and PBF elements	China (2)	RCT (1) and quasi-non randomized trial (1)	Fee for service (1) and global capitated budget only (1)
Conditional provision of material goods	Conditional provision of material goods alongside supervision and quality improvement strategies	Tanzania (1)	Quasi-non randomized trial (1)	Unconditional gifts (either immediate or delayed) as alternative interventions and control (all receive a standard encouragement intervention) (1)
Financial and non-financial incentives + clinical decision guide	Mix of financial and non-financial incentives, alongside clinical decision guide and supervision/technical support	Burkina Faso, Ghana and Tanzania (all in 1)	CBA (1)	Control as standard care (1)
Performance related pay	Performance-related pay (results-based management) involving different types of agreement according to province implemented (ranging from multi-level agreements with strategic targets to not specified)	Brazil (1)	ITS (1)	Comparison of impact over time in implementing provinces. (1)
Performance based contracting or service agreements	Service agreements introduced as part of reform and in case of contracting, with indicators for performance chosen at year end to avoid distortion	Cambodia (2), Haiti (1)	CBA (2), ITS (1)	Routine practice as control (2) and comparison of indicators over time. (1)
Hybrid scheme	Payment per output and for target	China (1), Peru (1)	Quasi/non randomized trials (2)	Control as standard care (2)
Results based aid	Fixed element alongside a targeted element as part of results based aid	El-Salvador (1)	CBA (1)	Control as status quo (1)

Scheme classification (2): payment per output

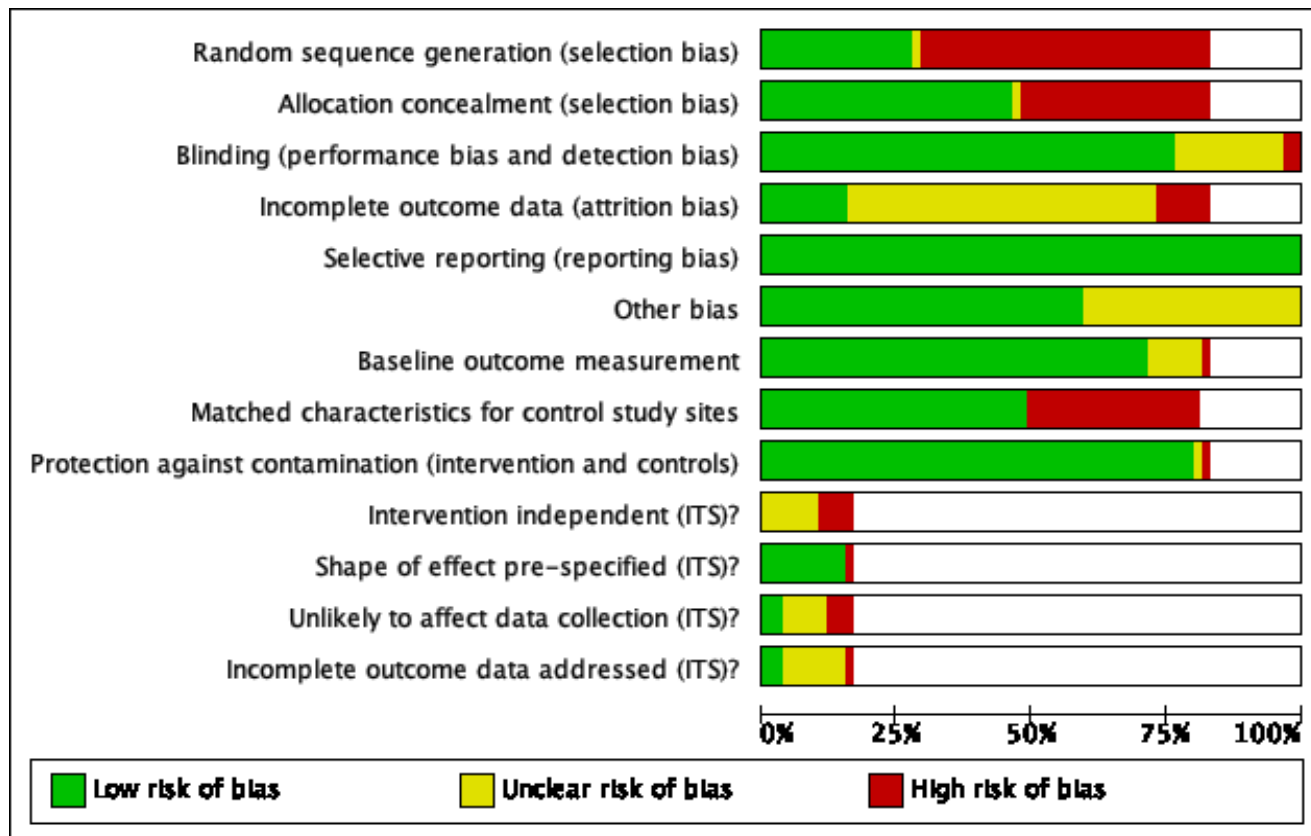
Scheme classification	Details on scheme	N.	Countries included (n)	Study types (n)	Comparators (n)
Payment per output	Payment for each output	9	Afghanistan (1), Argentina (1), China (1), Cambodia (2), DRC (1), Swaziland (1), Rwanda (2)	RCT (4), Quasi/non-randomized (2), ITS (2), CBA (1)	Control as status quo/standard care (4), comparison over time in implementing locations (2), comparator of matched funding or background PBF programmes into which experiments nested (3)
	Payment per output with income potentially withheld	1	China (1)	ITS (1)	Comparison of impact over time in implementing hospital. (1)
	Payment per output including revenue	1	China (1)	ITS (1)	Comparison over time in implementing provinces (1)
Payment per output modified by quality score	Payment per output with quality as multiplicative adjuster (between 0-1)	11	Congo (1), Zambia (1), Benin (1), Rwanda (8)	Quasi/non-randomized trial (8), CBA (1), ITS (2)	Control with standard care (2), Over time comparison in implementation areas (2), Comparator of matched funding (7)
	Payment per output with quality bonuses (quality adjuster an additional but not detracting component)	7	Burundi (4), Zambia (2)	RCT (2) and CBA (4)	Control as standard care (5), Comparator of enhanced matched financing (2)
	No description of payment equation - quality adjustment noted	1	Afghanistan (1)	RCT (1)	Control with standard care (1)
Payment per output modified by quality and equity score	Modification to payment equation based on population equity or remoteness of facilities	5	Burkina Faso (1), Cameroon (2), DRC (1), Zimbabwe (1)	Quasi/non randomized trials (2), CBA (3)	Control as standard care (4) and comparator including equipment and other in kind support (1)
Payment per output modified by quality and satisfaction score	Modification to payment including bonuses for enhanced patient satisfaction	2	Malawi (1), Zimbabwe (1)	CBA (2) and ITS (1) (one study does both)	Control as standard care (2)

Scheme classification (3): Target payment

Details on scheme	N.	Countries included (n)	Study types (n)	Comparators (n)
Potential for income gain only	12	Argentina (1), Kenya (1), Philippines (4), Tanzania (4)	RCT (5), CBA (5)	Control as standard care/status quo (12)
Potential for income withheld	1	China (1)	ITS (1)	Over time (1)
Target payment or payment per input	1	India (1)	RCT (1)	Control as status quo (1)



Results overview of RoB (excerpt)



	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding (performance bias and detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias	Baseline outcome measurement	Matched characteristics for control study sites	Protection against contamination (intervention and controls)	Intervention independent (ITS)?	Shape of effect pre-specified (ITS)?	Unlikely to affect data collection (ITS)?	Incomplete outcome data addressed (ITS)?
Basinga 2011	●	●	●	?	●	●	●	●					
Bernal 2018	●	●	●	?	●	?	●	●	●				
Binyaruka 2015	●	●	●	?	●	●	●	●	●				
Binyaruka 2017	●	●	●	?	●	●	●	●	●				
Binyaruka 2018	●	●	●	?	●	●	●	●	●				
Bonfrer 2014a	●	●	●	?	●	?	●	●	●				
Bonfrer 2014b	●	●	●	?	●	?	●	●	●				
Brock 2018	●	●	●	●	●	?	●	●	●				
Celhay 2015	●	●	●	●	●	●	●	●	●				
Chang 2017			?	●	?	●				●	●	●	?
Chansa 2015			●	●	●	●				?	●	●	●
Cruzado de la Vega 2017	●	●	●	?	●	●	●	●	●				
Das 2017	●	●	?	●	●	●	●	●	●				
de Walque 2015	●	●	●	?	●	●	●	●	●				
de Walque 2017	●	●	●	?	●	●	●	●	●				
Duysburgh 2016	●	●	●	?	●	?	?	●	●				
Engineer 2016	●	●	●	?	●	●	●	●	●				
Falisse 2015	●	●	●	●	?	●	●	●	●				
Friedman 2016a	●	●	●	●	●	●	●	●	●				
Friedman 2016b	●	●	●	?	●	●	●	●	●				
Gertler 2012	●	●	●	●	●	●	●	●	●				
Gertler 2014	●	●	●	●	●	●	●	●	●				
Hullery 2017	●	●	●	?	●	●	●	●	●				
Ir 2015			?	●	●	●				●	●	?	?
Khim 2018			●	●	●	?				?	●	?	?
Kliner 2015	●	●	●	?	●	?	●	●	●				
Lagarde 2015	?	?	?	?	●	?	●	●	●				
Lannes 2015	●	●	●	?	●	●	?	●	●				
Lannes 2016	●	●	●	?	●	●	●	●	●				
Liu 2005			●	●	●	●				●	●	●	●
Matsuoka 2014			?	●	?	●				?	●	?	?
Mayumana 2017	●	●	?	●	●	●	●	●	●				
McMahon 2016			?	●	●	●				?	●	●	●

Comments on certainty of evidence

Differs by indicator but overall still judged as low to moderate (latter for sensitivity analysis – see paper)

Table 1. GRADE certainty ratings

Certainty	What it means
Very low	The true effect is probably markedly different from the estimated effect
Low	The true effect might be markedly different from the estimated effect
Moderate	The authors believe that the true effect is probably close to the estimated effect
High	The authors have a lot of confidence that the true effect is similar to the estimated effect

From: <https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/>



Presentation of results in the review itself

Outcome	Indicator	Direction of relative effect and GRADE assessment for targeted and un-targeted outcomes			
		Targeted outcomes		Un-targeted outcomes	
		Direction of effect	Certainty of the evidence	Direction of effect	Certainty of the evidence
Primary: Utilization and delivery	Provision of HIV testing (%)	▲	⊕⊕⊕⊖	▲	⊕⊕⊕⊖
	Provision of ART services (%)	▼	⊕⊕⊕⊖	No evidence	
	Provision of PMTCT (%)	▲	⊕⊕⊕⊖	No evidence	
	Bednet use (% of children and households using bednets)	▼	⊕⊕⊕⊖	—	⊕⊕⊕⊖
	TB adherence rate	□	⊕⊕⊕⊖	No evidence	
	Child immunization: % at least one vaccine	—	⊕⊕⊕⊖	No evidence	
	Child immunization: % fully vaccinate	□	⊕⊕⊕⊖	No evidence	
	Child immunization: % receiving BCG	▲	⊕⊕⊕⊖	No evidence	
	Child immunization: % receiving DTP	▼	⊕⊕⊕⊖	No evidence	
	Child immunization: % receiving measles vaccine	▲	⊕⊕⊕⊖	No evidence	
	Child immunization: % receiving polio vaccine	▲	⊕⊕⊕⊖	No evidence	
	Child immunization: % receiving pentavalent vaccine	—	⊕⊕⊕⊖	No evidence	
	Mothers receiving immunizations (%)	▲	⊕⊕⊕⊖	No evidence	
	Probability of any utilization (%)	—	⊕⊕⊕⊖	—	⊕⊕⊕⊖
	Frequency of curative utilization (%)	▲	⊕⊕⊕⊖	□	⊕⊕⊕⊖
	Frequency of outpatient utilization (%)	▲	⊕⊕⊕⊖	—	⊕⊕⊕⊖
	Frequency - all visits (number of visits)	—	⊕⊕⊕⊖	—	⊕⊕⊕⊖
	Frequency - elderly visits	No evidence		—	⊕⊕⊕⊖

Example – high level of granularity, here presenting summary

Direction of effect key

- ▲ Desirable
- ▼ Non-desirable
- Neutral
- Uncertain

Certainty in evidence key

- ⊕⊕⊕⊖ Moderate
- ⊕⊕⊕⊖ Low
- ⊕⊕⊕⊖ Very low

Overview of results against standard care (1)

Outcome	Summary of impacts	GRADE
Utilization and delivery of health services	<p>Overall inconsistent picture: the intervention may improve some utilization and delivery indicators but may lead to poorer results for other indicators.</p> <p>When targeted:</p> <ul style="list-style-type: none"> • Proportion of persons receiving HIV testing (range 6-600%) and delivery of PMTCT (range 3.8 to 21%) may be affected positively; proportion of persons receiving ART and children (up to 120% decline) and households protected with bednets may decline (up to 7.3%); • effects on tuberculosis adherence are uncertain given very low certainty evidence; • effects on family planning outreach may be positive (moderate certainty evidence, increase up to 300%) • Evidence on mother and child immunizations and antenatal care utilization is mixed. <p>Effects on indicators when they are not targeted are largely uncertain or neutral.</p>	<p>⊕⊕⊖⊖ Low</p>
Quality of care (mainly assessed by score)	<p>Largely uncertain overall.</p> <p>Effects on quality of care indicators appear to be sustained only when indicators are targeted.</p> <p>Indicators for which moderate certainty evidence was found include:</p> <ul style="list-style-type: none"> • P4P probably improves quality of care scores (range 5 to 300% relative increases); • P4P probably improves the quality scores of available medicine and equipment, effects ranged from 2.7% to 220%; • Overall quality of service by specific departmental area/service: P4P probably improves the average quality of service scores in specific targeted areas (effects ranged from 39% to 15-fold increase in scores). <p>P4P may make little or no difference to staff knowledge and skills (low certainty evidence).</p>	<p>⊕⊕⊖⊖ Low</p>

Overview of results against standard care (2)

Outcome	Summary of impacts	GRADE
Health outcomes	<p>When targeted:</p> <ul style="list-style-type: none"> • P4P may reduce child mortality (range: 0.2-6.5%); • P4P may lead to a modest reduction of 2-3% in the proportion of children with reported anaemia; • P4P may increase the likelihood of tuberculosis treatment success (range: 12-20% improvement in treatment success). • Evidence on neonatal mortality is inconsistent: P4P may have desirable effects and ensure reduction in neonatal mortality in implementing clinics by up to 22% in one study, however, another study identified increases in region of 6.5% across catchment areas of P4P incentivized providers. 	⊕⊕⊖⊖ Low
Unintended effects	No distorting unintended effects.	⊕⊕⊖⊖ Low
Changes in resource use	<p>Overall certainty in evidence across indicators is low, for those where moderate certainty observed:</p> <ul style="list-style-type: none"> • P4P probably has a positive effect on human resource availability (range: 19-44%, moderate certainty evidence). • P4P probably affects infrastructure functionality and medicine availability positively. 	⊕⊕⊖⊖ Low

Overview of results against standard care (3)

Outcome	Summary of impacts	GRADE
Provider motivation, satisfaction, absenteeism and acceptability	<p>P4P probably makes little or no difference to provider absenteeism (range: 0.7-2%, low certainty evidence). Effects on overall motivation scores and satisfaction are largely neutral (low certainty evidence).</p> <p>Where these outcomes were not directly targeted, the intervention may have desirable effects.</p>	⊕⊕⊖⊖ Low
Patient satisfaction and acceptability	<ul style="list-style-type: none"> • Overall positive, with only two outcomes noting limited to no effect in relation to the intervention (satisfaction with care quality and provider communication). • When not targeted, effects may be largely positive, except for satisfaction with provider-patient contact time and facility opening hours. 	⊕⊕⊖⊖ Low
Impacts on management or information systems (if not a targeted measure of performance)	<p>P4P may positively affect facility managerial autonomy (low certainty evidence), probably makes little to no difference to management quality or facility governance.</p>	⊕⊕⊖⊖ Low
Equity considerations: evidence of differential impacts on different parts of the population	<ul style="list-style-type: none"> • P4P may increase the proportion of poor persons utilizing child immunization services, however the intervention may potentially decrease the proportion of poor persons utilizing antenatal care. • P4P may make little to no difference to the utilization of institutional deliveries by poorest groups. • If not explicitly targeted, effects are mixed. 	⊕⊕⊖⊖ Low

Overview of results against other interventions (1)

Outcome	Summary of impacts	GRADE
Utilization and delivery of health services	<ul style="list-style-type: none"> • P4P may positively affect the probability of persons utilizing care (range: 2 to 10%, however, evidence on immunization utilization is indicative of little to no effect or uncertain). • P4P may have little to no important effect on the utilization of any family planning services or the overall rates of antenatal care utilization; however, P4P may positively affect the timeliness of ANC care-seeking (range: 1 to 10% women accessing care earlier). • Evidence on the effects of P4P on % women utilizing institutional deliveries is uncertain (range between -9% and 23%). • P4P may have negative effects on postnatal care utilization. • Evidence on effects on non-targeted utilization outcomes is sparse. 	⊕⊕⊖⊖ Low
Quality of care	<ul style="list-style-type: none"> • P4P may lead to improved quality of care in relation to family planning (up to 500%) or antenatal care (up to 40%). • P4P may also increase procedural care quality, e.g. increasing the proportion of staff conducting appropriate patient background and physical assessments during consultations. 	⊕⊕⊖⊖ Low
Health outcomes	P4P may have little to no effect on health outcomes, both when targeted and when not targeted.	⊕⊕⊖⊖ Low
Changes in resource use	P4P may increase equipment availability by 75%, however medicine availability may be reduced by up to 160%.	⊕⊕⊖⊖ Low

Overview of results against other interventions (2)

Outcome	Summary of impacts	GRADE
Provider motivation, satisfaction, absenteeism and acceptability	<ul style="list-style-type: none"> No studies assessing when directly targeting. Little to no difference when not targeted. 	⊕⊕⊖⊖ Low
Patient satisfaction and acceptability	<ul style="list-style-type: none"> No studies assessing when directly targeting. Available evidence suggests desirable effects on cleanliness, waiting and contact time indicators, but the intervention may make little to no different to overall patient satisfaction scores. 	⊕⊕⊖⊖ Low
Impacts on management or information systems (if not a targeted measure of performance)	<ul style="list-style-type: none"> Where indicators have been targeted, paying for performance may have desirable effects. For untargeted outcomes, effects are uncertain due to very low certainty evidence. 	⊕⊕⊖⊖ Low
Equity considerations: evidence of differential impacts on different parts of the population	<p>The intervention may have little or no effect on equity or may worsen equity, for example, paying for performance may lead to increased utilization of family planning services and institutional deliveries among wealthier population groups.</p>	⊕⊕⊖⊖ Low

Highlight messages

Main outcome	Key message
Utilization and delivery of health services	Inconsistent effects overall – effects differ by indicator but no overarching trend
Health outcomes	Slight positive impacts on the majority of these when assessed against control, but not when compared to enhanced financing
Quality of care	May increase quality of care overall (especially when directly targeted) and may increase availability/ functionality of medicines, equipment and infrastructure Uncertain or limited effects on process quality.
Unintended effects	Assessed in minority of studies, probably no negative distorting effects.

Highlight messages (2)

Main outcome	Key message
Health worker motivation	Mixed effects
Facility governance, autonomy	May increase managerial autonomy, but limited effects on quality of management or governance
Equity of service utilization or delivery	Depends on the comparator intervention. When assessed against status quo, some evidence of redistributive effects, but otherwise mixed.
User fees	Effects unclear

Also considered which scheme types perform best overall:

- Performance based contracting and results based aid seem to achieve best outcome effects, but minimally assessed.
- Overall, schemes adjusting for quality + equity perform best against utilization outcomes (payment per output schemes performed best, but target payment + adjustments also).



Limitations

1. Given volume of data, restricted inclusion to those indicators assessed comparably in two or more studies (lack of harmonization on indicators incentivized so this does not help)
2. Focused on relative effects, but did not carry out a comprehensive review of absolute effects (reporting on this is patchy)
3. Had initially planned on exploring effects of ancillary components – reported in two-thirds of studies but inconsistently reported on and designs also do not accurately capture effects
4. Need for more consistent reporting on some aspects, e.g. overall theory of change of intervention, costs of programmes, how indicators/targets + payments were set, interaction with wider relevant context changes (e.g. parallel health system investments)

Some final reflections

- Growing evidence (9 to 59 studies) but still gaps, e.g. on aid modalities; also limited and mixed evidence on cost-effectiveness; need for more focus on health outcomes and longer term results
- Complexity of multiple dimensions means careful interpretation is needed
- Findings suggest some possible mechanisms of change – e.g. additional financing (hence relative lack of difference with comparators), also increased autonomy, inputs to QoC
- But surprisingly little shift on HW indicators, equity, financial access
- Lack of dominance over comparator investments raises importance of considering alternative investments, e.g. direct facility finance, depending on context and priorities
- Findings by design type interesting and suggest areas of future focus
- Also need for greater probing on heterogeneity within schemes
- Much more in the pipeline – 63 studies awaiting classification. Any interest in doing the next review, let us know!

For further details, see:

Diaconu K, Falconer J, Verbel Facuseh AV, Fretheim A, Witter S. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. Cochrane Database of Systematic Reviews 2020, Issue 12. Art. No.: CD007899. DOI: 10.1002/14651858.CD007899.pub3

Or contact us:

kdiaconu@qmu.ac.uk

switter@qmu.ac.uk