

Systematic Review Protocol

Main title	The Impact of Vouchers on the Use and Quality of Health Services in Developing Countries: A Systematic Review
Review group	Venture Strategies for Health and Development
Authors IN ORDER OF CREDIT (Please include first and surnames, institutions. Include titles - Dr, Prof - if you want them to be used.)	Carinne Meyer, MPH, MA, <i>School of Public Health, University of California, Berkeley</i> Nicole Bellows, PhD, <i>Independent Consultant</i> Martha Campbell, PhD, <i>Venture Strategies for Health and Development</i> Prof. Malcolm Potts, MB, BChir, PhD, FRCOG, <i>Bixby Center for Population, Health and Sustainability, University of California, Berkeley</i>
EPPI-Centre reference number	[To be completed by EPPI-Centre]
Month/year of publication	[To be completed by EPPI-Centre]
This report should be cited as...	--
Contact details (address, phone number, email)	Venture Strategies for Health and Development 2140 Shattuck Avenue, Suite 1110 Berkeley, California 94704-1234 USA, Tel +1 510 665 1880
Institutional base	
Review Group (with institutions)	<u>Voucher Expert:</u> Anna C. Gorter, MD, MSc, PhD Project Director - Vouchers Instituto Centro Americano de la Salud (ICAS) <u>Methods Expert:</u> Andrew Anglemyer, PhD, MPH Cochrane HIV/AIDS Group Editor Institute for Global Health, University of California, San Francisco
Advisory group (with institutions)	--
Conflicts of interest (if any)	See Section 4
Acknowledgements	--

1. Background

1.1 Aims and rationale for review

Measures of infant mortality, maternal mortality, and disease-specific deaths such as those related to malaria and HIV vary widely between countries [1]. Developing countries bear 93% of the world's disease burden and account for only 11% of the world's health spending [1]. As a result of this gap between burden of disease and funding, the health sector in developing countries has been an important recipient of international aid. Over the last three decades, the world has seen an increased focus on global health partly due to the identification of health as a key determinant of economic growth and poverty reduction [2, 3]. As a result, official development assistance from bilateral and multilateral agencies towards health has increased from \$4.5 billion in 1996 to \$7.9 billion in 2004 [4].

In addition to allocating increased funds to address health inequalities, donors have emphasized the need for efficient and transparent spending of aid funds [5]. A variety of strategies exist for distributing health aid. One strategy that is growing in popularity is the use of voucher programs, where vouchers are distributed to a targeted population for free or subsidized health goods and services. While there is much discourse in the literature on how voucher programs work and why they are potentially important, the literature lacks a systematic assessment of the existing evidence on whether vouchers yield value for donors in the form of efficient spending of health aid. As such, the overall objective of this systematic review is to assess whether voucher programs thus far have been successful in achieving their objectives and should therefore be considered as a mechanism for further health aid. Additionally, this review aims to identify conditions in which voucher programs are more or less successful and to specify gaps in the literature that require further research.

1.2 Conceptual issues and hypotheses

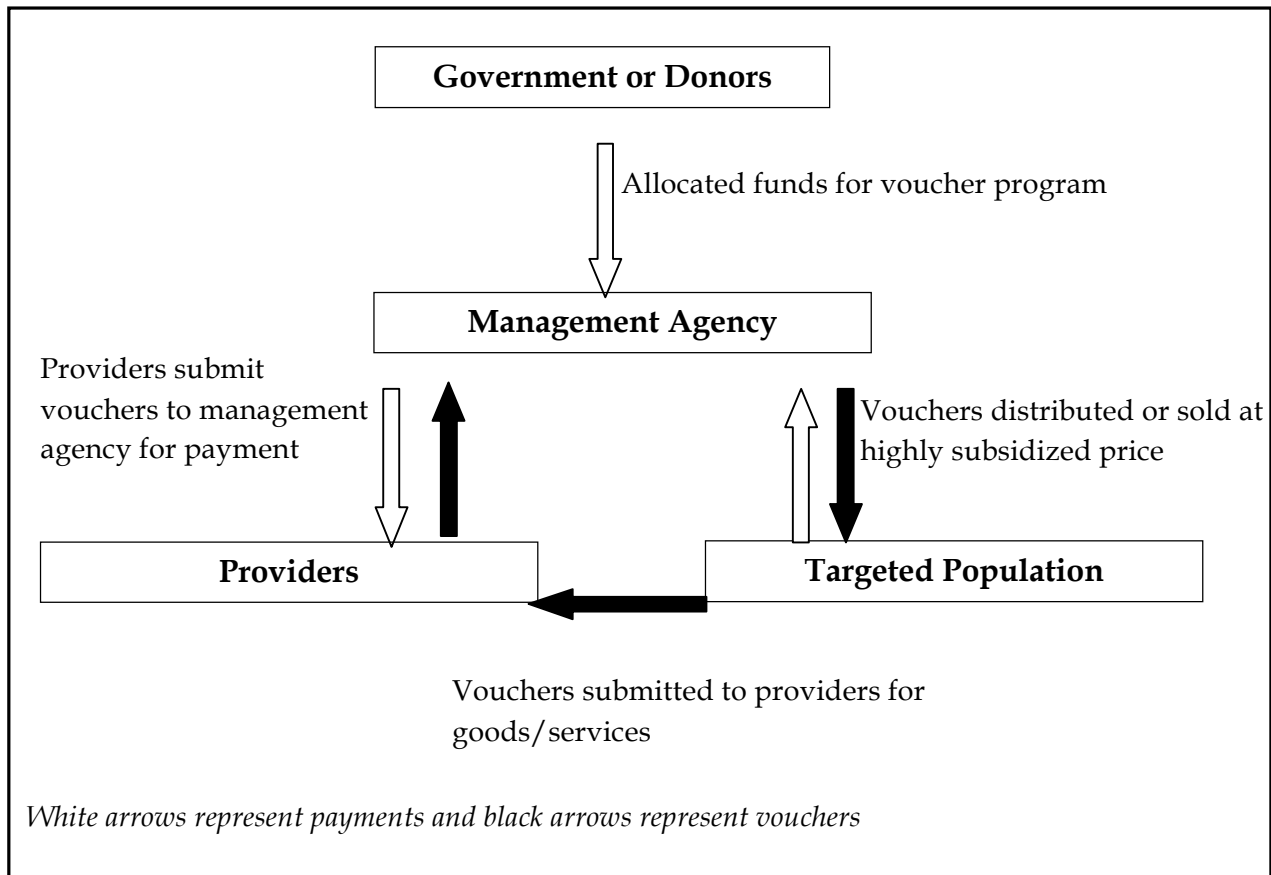
1.2.1 *How voucher programs work*

In a voucher program there are typically four major actors: (1) the government or donors who provide the funding, (2) a management agency that administers the program, (3) providers who deliver the health goods and/or services, and (4) the voucher recipients who are in need of health goods and/or services. Vouchers are

typically competitive with multiple providers; however, it is possible for them also to be non-competitive as well.

The management agency plays an important role in contracting providers to deliver goods and/or services to voucher holders, distributing the vouchers to the targeted population, and overseeing the delivery of care by the providers. The targeted voucher recipients may be based on income status, living in a geographic region, having certain risk factors, or other relevant characteristics depending on the program. Once vouchers are distributed, recipients bring the vouchers to participating providers. After the specified goods and services are delivered by the provider, the provider submits the vouchers to the management agency for reimbursement. Figure 1.2.1 describes how monies and vouchers flow between the primary roles in voucher programs.

Figure 1.2.1 Flow diagram of payments and vouchers



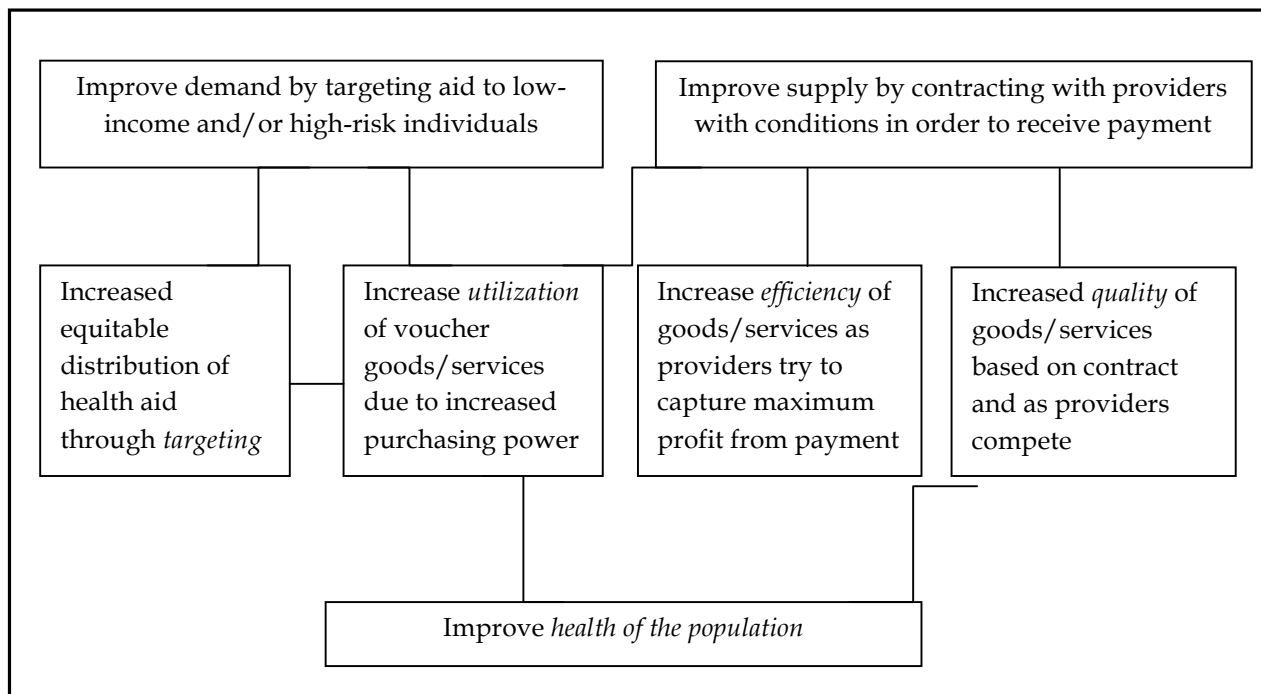
1.2.2 Theoretical context for vouchers

The theoretical context for voucher programs can be found in the basic economic theories of supply and demand where voucher programs aim to inject market mechanisms into the delivery of health aid in order to improve efficiency [6]. One rationale for subsidizing health care is the inequitable distribution of wealth and health [7]. Low income individuals may have the need for health goods/services; however, without financial resources they do not have the purchasing power to access health goods/services, particularly in the private sector. Voucher programs are a form of demand-side financing and output-based aid, where aid monies are used to stimulate demand for goods and/or services, contrasting more traditional supply-side strategies, which often focus on providing the inputs for health aid such as construction of facilities or provision of supplies. One advantage of voucher programs is the potential to place purchasing power in low-income individuals that might otherwise be ignored in the market due to their lack of funds [8]. By *targeting* the benefit towards low-income and/or high-risk individuals, voucher programs can increase demand among those most in need.

In addition to enhancing demand through targeting, voucher programs also aim to improve the supply of goods/services available. The theoretical basis for supply enhancement in voucher programs can be found in the principal-agent model of the economics literature where the principal delegates a task to an agent via an inducement embedded in a contract [9]. Inducements (sanctions or incentives) are designed to ensure that the task is completed satisfactorily. In voucher programs, the funder or management agency serves as the principal and the providers (either public or private) are the contracted agents. Providers are given a financial incentive for delivering health goods/services of sufficient quality, which is therefore expected to yield increased *utilization* of goods/services and increase *quality* of the goods/services provided. Additionally, providers have the incentive to meet the obligations of their contract in the most efficient manner in order to capture more of the set payment established in vouchers, thereby potentially yielding gains in *efficiency*. Quality and efficiency may also increase as providers compete for patients/consumers in health voucher programs.

By enhancing demand and supply, the ultimate aim of health voucher programs is to improve the *health of the population*. Figure 1.2.2 details the process by which health voucher programs are expected to improve the delivery of health aid.

Figure 1.2.2 Improving health aid through voucher programs



1.2.3 Hypotheses examined in review

For the purpose of this systematic review, the claims around voucher programs that are of high interest to policy makers include: whether voucher programs effectively target low income individuals, whether the increase utilization of health goods/services, whether the costs of voucher programs are more efficient than other forms of health aid distribution, whether vouchers result in improved quality of care, and whether voucher programs improve the health of populations. These questions can be stated as hypotheses to be tested in voucher program evaluations. The systematic review will examine the evaluation research to test the following hypotheses:

1. Voucher programs deliver health goods and services to targeted low-income and/or high populations.
2. Voucher programs increase utilization of health care goods/services.
3. Voucher programs allow for more efficient distribution of services compared to traditional aid programs.
4. Voucher programs improve the quality of services offered at a facility.
5. Voucher programs result in improved health of a population.

1.3 Research background

To date, vouchers have been used for basic health services, maternal and child health services, vaccinations, STI treatment, HIV testing and referral, sexual and reproductive health services (family planning, gender based violence and post abortion care and safe abortion). Health voucher programs have been used throughout the developing world including countries in Asia, Africa, and Central and South America.

Much of the literature on voucher programs describes the potential benefits of delivering health aid via vouchers, as described in the previous section. There is not, however, a definitive consensus on whether voucher programs achieve their goals or potential benefits. Within the peer review literature there is some evaluation information on individual voucher programs. For example, voucher programs for reproductive health services in Nicaragua have been evaluated in several papers examining and data on costs, utilization, quality measures, and population health impact [10-12]. In Tanzania, numerous studies detail the results of a voucher program to increase the use of insecticide treated bednets [13-17]. In general, these studies have found favorable results for the Nicaragua and Tanzanian programs on the variables presented.

One systematic review by [18] examined private for-profit interventions for the poor and examined three voucher programs, bednets in Tanzania and Zambia and reproductive health services in Nicaragua. A number of health voucher programs have been implemented in more recent years, with evaluation data available in both the peer-review and grey literature. It is particularly important to include the grey literature, such as agency and funder reports that may show more outcomes of limited effects than those selected for peer-review publications. To date, no systematic review has assessed whether voucher programs have achieved their specified goals.

1.4 Objectives

The primary objective of this review is to summarize the evidence on the effectiveness of voucher systems for health services in developing countries. The literature on health voucher systems will be evaluated to determine the extent that voucher systems in the

private sector have improved the quality, efficiency, and use of health services by poor persons in developing countries.

2. Methods used in the review

2.1 User involvement

This review is intended to aid policymakers within donor government agencies and recipient government ministries of health and finance, program staff at non-governmental organizations working, and health care providers and academics involved in health access and health systems strengthening programs. We aim to create a user-friendly report that synthesizes the results of our search and offers conclusions that can be translated into policy decisions. This report aims to provide stakeholders with an impartial review of the current body of evidence on health vouchers which will aid in making more informed decisions about whether health voucher programs should be considered for future application and if so, under what circumstances will they most likely be successful. We plan to disseminate this review to major funders of health aid to include both multilateral funders (World Bank, UN agencies) and bilateral funders (USAID, KfW, DFID, etc).

2.2 Identifying and describing studies

2.2.1 Defining relevant studies: inclusion and exclusion criteria

Seven different criteria will be used to determine whether a study will be included or excluded from the systematic review.

1. Language – included studies will be limited to those with an abstract published in English. Non-English publications with an English abstract will be reviewed for relevance and an appropriate translation will be sought when necessary. Only publications with an English abstract will be included in this review due to the limited availability of publications and search engines in other languages and the language capacity of the team. Our systematic search strategy can be replicated in the future to include updated evidence and can be expanded to include additional foreign language databases.

2. Population – included studies will be limited to those located in developing countries at the time the voucher program was operating. The Human Development Index (HDI) will be used to determine development and voucher programs located in a country assessed as “very high human development” by HDI will be excluded from the analysis.
3. Time frame – included studies will be limited to those published from 1960 to 2010. The 1960 cut-off date was chosen because the background literature indicates that the earliest health voucher programs for which there are evaluation data (reproductive health care in Taiwan and Korea) occurred during the 1960s. Studies prior to 1960 will be excluded from the analysis. Additionally, reports on future voucher programs not yet implemented will be excluded.
4. Type of study – included studies will be limited to those that are intended to evaluate some aspect of a health voucher programs and contain some quantitative evidence. Study designs such as before and after with and without controls, time series, and cross-sectional analyses with a comparison group will be included. General descriptions and opinion pieces on voucher programs will not be included in the analysis.
5. Type of voucher program – included studies will be limited to voucher programs that are providing health goods and/or services. Examples of relevant health goods and services are: skilled provider care, hospital and clinic services, health insurance, pharmaceuticals, family planning products, bed nets for the prevention of malaria, and vaccinations. Studies on voucher programs delivering food, clean water, and non-health education will not be included in this review, even though they may have a health impact.
6. Voucher characteristics – included studies will be limited to voucher programs that operate where health aid is distributed to a population of potential users (either for free or at subsidized price) through a physical voucher or a voucher-

like targeting mechanism, such as a “poverty card” and vouchers are used for provider reimbursement. Additionally, the value of the voucher must equal at least 25% of the total cost of the health goods/services for which the voucher is being used. Excluded from the analysis will be studies on voucher programs where no physical voucher exists, where vouchers are not used to reimburse providers, and where the value of the voucher is below 25% of the cost of the goods or services.

7. Study designs – included studies evaluating voucher programs will have an observable contrast such as time (e.g., before and after program implementation), control group (e.g., non-voucher control areas, or non-voucher patients), or control program (e.g., supply-side program delivering the same goods or services). Relevant study designs include: random control trials, non-randomized trials (time series), case-control, cohort, pre-post with and without controls.

2.2.2 Identification of potential studies: search strategy

The search will involve two phases. In the first phase the following sources a start date of 1960:

- *Bibliographic databases:* PubMed, EMBASE, MEDLINE, POPLINE, ELDIS, Jsotr, Inter-Science (Wiley), ScienceDirect, TheCochraneCentral Register ofControlledTrials (CENTRAL), The Database of Abstracts of Reviews of Effectiveness and the EPOC Register, Web of Science, Bioline International, CAB-Direct (Global Health), World Health Organization Library Information System (WHOLIS), African Healthline.
- *Publishers' pages of Key Journals:* Journal of Development Effectiveness Health Policy, The Lancet, Global Public Health, Health Policy and Planning, International Journal of Health Planning and Management
- *Organization and network websites:* UNICEF, USAID, DFID, KfW, the World Bank, Partnerships for Health Reforms, Abt Associates, Management Sciences for Health (MSH), Oxford Policy Management, the Private Sector Partnerships-One, Marie Stopes International, Population Council, GPOBA, RBFHealth, RHVouchers, International Initiative for Impact Evaluation (3ie), and Research for Development.

The basic search terms used will be: (voucher* OR coupon* OR output-based* OR “output based” OR “result based” OR “results based” OR results-based* OR “performance based” OR performance-based* OR pay-for-performance OR "pay for performance" OR “demand side” OR demand-side) AND (developing countr* OR "poor countr*" OR "low-income countr*" OR "low-resource countr*" OR "low and middle income") modified as necessary according to database.

In the second phase of the search, we will conduct a supplemental keyword search in google.com based on leads generated by the search described above. For example, if a search in PubMed identified an article describing (but not evaluating) a voucher program for maternal health services in India called *Chiranjeevi Yojana*, a search of the google.com and google.scholar will use a search of “India AND maternal AND voucher” and “*Chiranjeevi Yojana*” to determine whether there is additional information on the voucher program that may include evaluation information relevant to the analysis.

Another search component in phase two is to review the citations of all included studies and contact the lead authors or corresponding authors from the included studies with the request that they review the list of studies and make further suggestions for consideration, particularly for unpublished studies.

A record will be maintained describing the databases searched, the keywords used, and search results. This information will be included in the EPPI-reviewer database, along with all studies that pass the abstract screen for retrieval.

2.2.3 *Screening studies: applying inclusion and exclusion criteria*

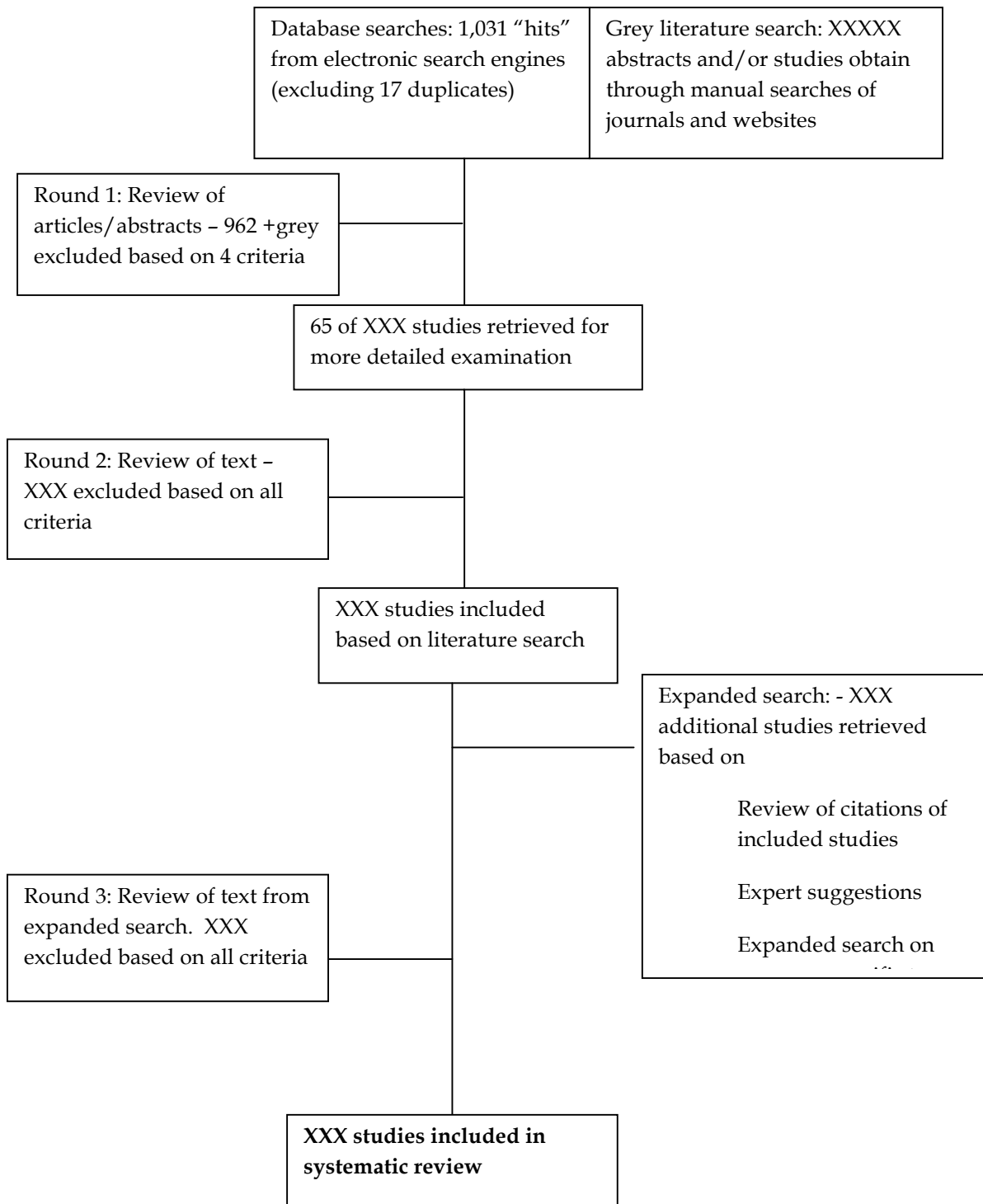
The application of the inclusion and exclusion criteria will take place in three rounds. In the first round, the abstracts will be reviewed by one team member using the first five criteria specified in section 2.2.1: language, population, time frame, type of study, and type of voucher program. If no abstract is provided, studies will be sought for retrieval if the study title contains the words “voucher(s)” or “output-based” or if the study title contains the name of a known voucher program. Studies that were not excluded in round one will be retrieved for further analysis.

In round two, two team members will independently apply all of the specified criteria listed in section 2.2.1 and determine whether the study should be included for analysis based on the full text of the study. In the case of a discrepancy between the two reviewers’ assessments, the case will be discussed with a third team member for a decision and further refinement of inclusion/exclusion criteria for further transparency.

In round three, further studies retrieved through the reference review, google.com search, and content experts will be reviewed by two team members, applying all eight criteria.

The final number of studies examined will be entered into a separate database in EPPI-reviewer software. The flow diagram detailing the search process is in Figure 2.2.3.

Figure 2.2.3: Flow diagram of search strategy



2.2.4 Characterizing included studies

Information from the studies included in the review will be extracted using a data extraction form and data will then be entered into a summary form. Information to be extracted will include: location of voucher program, time period of voucher program, type of health voucher program, targeted population, private or public providers, program management, relevant study outcome(s), study design, authors, and publication date. An example of the extracted data form is presented below in Figure 2.2.4:

Figure 2.2.4: Example of extracted characteristics of included studies

<p>Authors (publication date) : Meuwissen et al. (2006)</p> <p><u>Voucher Program Description:</u></p> <p><i>Location:</i> Nicaragua</p> <p><i>Type of voucher program:</i> Reproductive health care</p> <p><i>Targeted Population:</i> Low income adolescents</p> <p><i>Funders:</i> Unclear</p> <p><i>Time period of voucher program:</i> 2000 – 2005</p> <p><i>Private/public providers:</i> private and public</p> <p><i>Management organization:</i> Central American Health Institute</p> <p><u>Evaluation Study Description:</u></p> <p><i>Study design:</i> Cross-sectional community sample of voucher receivers and non-voucher receivers.</p> <p><i>Study time period:</i> September 2000 – July 2001</p> <p><i>Study outcome categories:</i> utilization</p> <p><i>Study outcomes:</i> utilization of reproductive health care, use of condoms, use of other contraceptives</p>
--

2.2.5 Identifying and describing studies: quality assurance process

Pilot testing of key word searches will be conducted to ensure that the keyword list and limits are neither too broad, thus generating unmanageable results, nor too narrow and thus excluding important studies that should be included in the review. The round one criteria for inclusion/exclusion at the abstract level is straight-forward and should not generate bias, however, any hesitation on the reviewing team member as to whether the abstract should be included or excluded will default to inclusion so that two individuals can assess the criteria in round two.

The two team members applying inclusion/exclusion criteria in round two will do so independently and come to consensus on any discrepancies by bringing in a third team member to discuss and decide on whether the study will be included or excluded. The same strategy will apply to characterizing and synthesizing the data extracted from the studies, as described in the next sections.

2.3 Methods for synthesis

2.3.1 Assessing quality of studies

Reviewers will follow guidelines recommended by the Cochrane Collaboration for assessing study quality using the CONSORT checklist for RCTs, cluster RCTs, controlled before and after and interrupted time series. We will use the Newcastle-Ottawa Scale (NOS) as recommended by the Cochrane Non-Randomized Studies Methods Working Group for assessing the quality of observational studies such as case-control, cohort or other non-randomized studies (See Appendix 1 and 2).

Reviewers will then make an overall 'Level of Quality' assessment regarding the individual study based on CONSORT and NOS criteria. Reviewers will assign each study a level quality of a high, medium, low or uncertain quality, as defined below.

Figure 2.3.1 Quality Ranking:

High Quality: appropriate and clearly described selection of participants, measurement of exposure and outcome variables, use of design and analytical methods to control confounding

Medium Quality: inappropriate or unclear use of one of the following: selection of participants, measurement of exposure and outcome variables, use of design or analytical methods to control confounding

Low Quality: inappropriate use of two or more of the following: selection of participants, measurement of exposure and outcome variables, use of design or analytical methods to control confounding

Unclear Quality: unclear description of any of the following: selection of participants, measurements of exposure and outcome, study design or analytic methods to control for confounding

2.3.2 Selection of outcome data for synthesis

The unit of analysis for data synthesis will be the outcome variable since one publication may contain multiple relevant outcome variables. Based on the aims of vouchers programs as described in section 1 of this protocol, the following five categories of outcomes will be examined in the review:

1. Targeting - the extent that vouchers reach the intended recipients (e.g., low income individuals). Relevant outcome variables in this category may include: income level of voucher users, high risk-status of voucher users, and health status of voucher users.
2. Utilization - the extent that voucher programs change the utilization of health goods and services. Relevant outcome variables in this category may include: vaccination coverage, use of health good (e.g. bednet), and use of health service (e.g. prenatal care).
3. Efficiency - the extent that voucher programs deliver health goods and services efficiently. Relevant outcome variables in this category may include costs of goods/services delivered (compared to competing programs and standards) and administrative cost comparisons.
4. Quality- the extent that voucher programs increase the quality of health goods/services being provided. Outcome variables associated with quality may include several different measures. Typically, quality measures are classified as structure, process, or outcome measures. Structural measures of quality for health voucher programs may include whether contracted providers have sufficient supplies or are open at consistent hours. Process quality measures may include whether contracted providers followed a specific protocol in treating patients or whether patients were treated in a timely manner. Outcome measures of quality may include patient perceptions of quality of care or percentage of complications post care.

5. Population health impact – the extent that voucher programs improve the health of the population. Relevant outcome variables in this category may include disease rates, fertility rates, and mortality rates.

Additional outcome categories may be necessary to include if the review of the literature identifies other relevant outcomes that cannot be classified in one of the five categories described above. A list of other outcome variables will be maintained in the EPPI-reviewer database and further categories will be created if the same type of outcome is examined more than three times in the studies analyzed.

2.3.3 Process used to combine/synthesize data

Within each outcome category, individual outcomes will be described according to the following information and entered into a summary table. Table 2.3.3 provides the framework for the individual outcome table with examples from the utilization category with a few examples included. The quality of the study will be assessed as low, medium, high, or unclear, according to criteria described in 2.3.1. Additionally, direction of effect will be either positive, negative, or no effect. Positive effects indicate good results, such as a decrease in disease rate or an increase in utilization of needed services. As such, the “positive” and “negative” results reported in our synthesis may not directly correspond to the coefficient in the regression results depending on how the outcome variable is specified.

Table 2.3.3: Summary Table for Utilization of Voucher Goods/Services

Voucher Program	Study Citation	Outcome Variable	Quality of Study	Direction of Effect	Notes

These data will be aggregated within each outcome category for deriving conclusions. The next section describes the process for deriving conclusions.

2.4 Deriving conclusions and implications

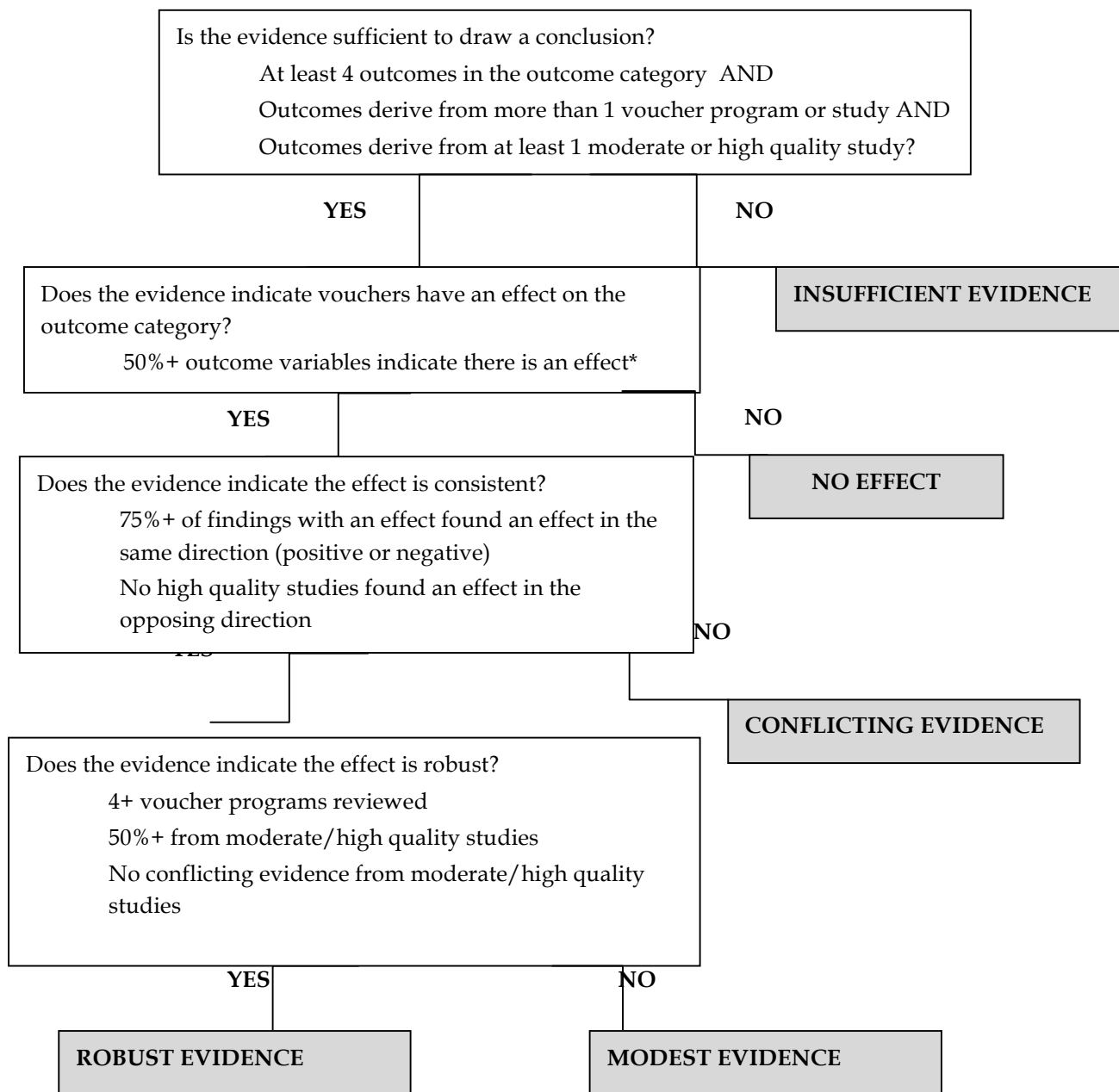
2.4.1 Conclusion categories

In deriving the conclusions and implications, we will summarize the findings for each category of outcomes, as discussed in section 2.3.5. Based on the aggregated data for each outcome category, we will conclude that the evidence supports one of five conclusion categories:

1. Insufficient evidence - indicates that there is not enough evidence available to determine the relationship between voucher programs and the outcome categories. A conclusion of insufficient evidence will be made if there are fewer than four variables in a particular outcome category, if all outcomes only derive from one voucher program (e.g. Nicaragua cervical cancer program), or if all outcomes derive from studies with a weak quality rating.
2. No effect - indicates that the evidence suggests that vouchers do not have an effect on the outcome categories. A conclusion of no effect will be made if more than 50% of outcomes within a category indicate there is no effect. An exception to this rule is when all of the “no effect” conclusions come from low quality studies and at least 25% of the outcomes from moderate/high quality studies find a significant effect.
3. Conflicting evidence - indicates that vouchers have had both positive and negative effects on the outcome category and may signal a need for sub-analysis to indicate under what conditions might voucher programs have positive or negative findings. A conclusion of “conflicting evidence” will be drawn if two different high quality studies or sets of studies (25% or greater) have findings in opposing directions.
4. Modest evidence - indicates that there is moderate evidence that voucher programs have an impact on the outcome category. A conclusion of “modest evidence” will be made if there is evidence indicating a positive or negative relationship; however, the evidence is not strong enough to be called robust. The outcomes may derive from fewer than four voucher programs or the quality of the studies may not be adequate to qualify for robust evidence.
5. Robust evidence - indicates that there is clear and convincing evidence that voucher programs have a significant positive or negative impact on the outcome category. A conclusion of “robust evidence” will be drawn if four or more voucher programs were reviewed, 50% of the findings (in the same direction) derive from medium or high quality studies, and no conflicting evidence from medium/high quality studies is found.

Figure 2.4.1 shows a decision tree that depicts how the conclusion categorization will occur.

Figure 2.4.1: Decision tree for making conclusions regarding outcome categories



* See exception described in text.
 2.4.2 Multiple study outcomes from the same study or voucher program

Since some studies may have multiple outcome variables in the same outcome category and some voucher programs may have several publications associated with the same program, it is important to consider how to control for publication bias, where one or two voucher programs dominate the findings through their multiple studies and

outcome analyses. Three steps are taken to control these biases. First, as described in 2.4.1, conclusions need to derive from more than one voucher program; otherwise a determination of insufficient evidence will be drawn.

Second, no more than three outcome variables from the same study will be included in the conclusion of an individual outcome category. All variables will be listed so that they may be used for sub-group analyses if needed; however, they will be footnoted in the tables as not being used for conclusion purposes for the overall outcome category conclusion. In the case of more than three outcome variables in the same category, the variables included will be the broadest and/or most generalizable. For example, if a voucher program delivering family planning services is analyzed with multiple outcome variables related to utilization (any contraception, condoms, IUDs, birth control pills, sterilization, etc...), then the three most commonly used family planning strategies at baseline will be included for analysis. When this type of situations arises, the three outcome variables will be independently selected by two reviewers and a third reviewer will be used if there is any disagreement.

Third, if more than one study of the same voucher program examines the same outcome variable on the same population, then the study of the highest quality will be used for drawing conclusions. For example, if one before-and-after study of a voucher program providing prenatal care finds a positive impact, while another study of the same program but with a controlled before-and-after design finds no impact, then the second study will be included and the first will not. When this type of situations arises, the study to be used will be independently selected by two reviewers and a third reviewer will be used if there is any disagreement.

2.4.3 Additional subgroup outcome analyses

In addition to making conclusions about the general outcome categories (e.g. utilization, efficiency, quality), conclusions will also be made about the outcomes as they relate to subgroups when data are available. For example, if 20 study outcomes examine questions of utilization and eight of these outcomes derive from studies on voucher programs designed to distribute insecticide treated bednets to prevent malaria, then a conclusion will be made about the relationship between voucher programs for bednets and utilization of bednets. The same criteria described in 2.4.1 will be applied to the subgroups. The number and types of subgroup analyses examined will depend on what is identified through the search process. Subgroup analyses will be considered based on the following variables:

- Types of services – e.g., maternity, family planning, bed nets, vaccinations, health goods vs. health services
- Location – e.g., region, country, urban/rural
- Voucher program characteristics – e.g., free vs. subsidized vouchers, private vs. public participation
- Study characteristics – e.g., high-quality studies

2.4.4 *Potential quantitative analysis*

Due to the heterogeneity of interventions and outcomes likely to be reported in the included studies, our review may have too broad a scope to make use of meta-analysis. The use of meta-analysis to describe the size of effect may not be meaningful if the interventions are so diverse that an effect estimate cannot be interpreted in any specific context. If this is the case, all reported estimates of effects therefore would come directly from the original studies and be analyzed using a narrative synthesis method as described above.

In the event that there are two or more studies of voucher programs with the same evaluation design and the same outcome variable, we will investigate the possibility of statistically combining results through a meta-analysis. Our review may include multiple comparisons and meta-analyses between various matched pairs of interventions. Two researchers will independently assess what studies are appropriate for meta-analysis and any discrepancies will be judged by a third researcher. Three main criteria will be used to assess whether a meta-analysis can be conducted. The first criteria is that two or more studies are examining the same type of voucher program (e.g. maternity services). Second, these studies must have comparable study designs (e.g. controlled before and after) and third, they must report the same outcome variables (e.g. increased percentage in attended deliveries).

We will then derive estimates of effect within groups of studies in a systematic way, to measure and investigate differences among studies and to interpret the findings and conclude how much confidence should be placed in them. We can expect both dichotomous and continuous outcomes to be extracted from studies and we will obtain a standard error from a confidence interval or a *P*-value where appropriate. The next step will be to prepare a table of summary data and effect estimates for each pair or group of studies. We will also explore if there is a way to graphical represent the data, perhaps with a forest plot where each lines represents a meta-analysis rather than a study.

2.4.5 *Outline of reporting for each outcome category*

In the systematic review report, the reporting of the results on the assessment of the evidence will follow an outline for each outcome category (targeting, utilization, efficiency, quality, and population health impact).

Outline of report for

- I. Table for outcome category displaying analyzed outcome variables in row per Table 2.3.3
- II. Overall assessment on outcome based on all available evidence using flow chart and criteria in 2.4.1
- III. Sub-category assessment by type of service, depending on where data is sufficient (the relevant categories will be based on what programs are identified and reviewed) Example:
 - a. Health services overall
 - b. Health goods overall
 - c. Reproductive health
 - i. Specific types of reproductive health
 - d. Bed nets
 - e. Health insurance
- IV. Sub-category assessment by location
 - a. Asia
 - b. Africa
 - c. Latin America
- V. Sub-category assessment by voucher program characteristics (e.g. management agency, free vs. subsidized)
- VI. Sub-category assessment restricting to high-quality studies
- VII. Further sub-category analysis and sensitivity tests

VIII. Relevant quantitative estimates

3. Timeframe

The review is expected to take 6 months to complete. At present, we are aware of several voucher programs being established and evaluations that are currently taking place but will likely not be available during the timeframe of this analysis. These new evaluations may substantially alter the results of a systematic review. As such, we recommend that the review be updated one to two years after completion of the review described in this protocol.

4. Conflict of interests

Venture Strategies has previously been involved in evaluating health voucher programs in Uganda. While Venture Strategies does not have any financial interests in the results of the systematic review, it is important that steps are taken to ensure any bias in reviewing evidence generated through Venture Strategies work. As such, any reports on health vouchers included in the review that were produced by Venture Strategies or someone affiliated with Venture Strategies will be evaluated by an external individual in order to judge the quality of the research per the criteria listed in section 2.3.1.

References

1. Satcher D: Eliminating Global Health Disparities. JAMA 2000, 284:2864.
2. Greene M, Merrick T. Poverty Reduction: Does Reproductive Health Matter? World Bank, July 2005.
3. Deaton A. Health, inequality, and economic development. Journal of economic literature, 2003.
4. OECD: Recent Trends in Official Development Assistance to Health. 2006.
5. World Bank: Healthy Development: The World Bank Strategy for Health, Nutrition, and Population Results. Washington DC: World Bank; 2007.
6. Bhatia M, Yesudian CAK, Gorter A, Thankappan KR. 2006. Demand side financing for reproductive and child health services in India. Economic and Political Weekly XLI(3): 279–284.
7. World Bank. A Guide to Competitive Vouchers in Health. Washington DC. World Bank 2005.

8. Mumssen Y, Johannes L, Kumar G. Output-Based Aid: Lessons Learned and Best Practices. 2010. The International Bank for Reconstruction and Development / The World Bank, Washington DC 20433.
9. Azam J, Laffont J. Contracting for Aid. *Journal of Development Economics* Volume 70, Issue 1, February 2003, Pages 25-58.
10. Borghi J, Gorter A, Sandiford P, Segura Z. The cost-effectiveness of a competitive voucher scheme to reduce sexually transmitted infections in high-risk groups in Nicaragua. *Health Policy Plan.* 2005 Jul;20(4):222-31.
11. McKay JE, Campbell DJ, Gorter AC. Lessons for management of STI treatment programs as part of HIV/AIDS prevention strategies. *American Journal of Public Health* 2006; 96:7-9.
12. Meuwissen LE, Gorter AC, Knottnerus JA. Perceived quality of reproductive care for girls in a competitive voucher programme. A quasi-experimental intervention study, Managua, Nicaragua. *Int J Qual Health Care.* 2006 Feb;18(1):35-42. Epub 2006 Jan 18.
13. Hanson A, et al. Household ownership and use of insecticide treated nets among target groups after implementation of a national voucher programme in the United Republic of Tanzania: plausibility study using three annual cross sectional household surveys. *BMJ* 2009;338:b2434.
14. Khatib R, et al. Markets, voucher subsidies and free nets combine to achieve high bed net coverage in rural Tanzania. *Malaria Journal* 2008, 7:98
15. Killeen GF, et al. Cost-sharing strategies combining targeted public subsidies with private-sector delivery achieve high bednet coverage and reduced malaria transmission in Kilombero Valley, southern Tanzania. *BMC Infectious Diseases* 2007, 7:121
16. Marchant T, et al. Timing of delivery of malaria preventive interventions in pregnancy: results from the Tanzania National Voucher program. *J Epidemiol Community Health.* 2010 Jun 7.
17. Mulligan J, et al. Costs and effects of the Tanzanian national voucher scheme for insecticide-treated nets. *Malaria Journal* 2008, 7:32.
18. Patouillard E, et al. Can working with the private for-profit sector improve utilization of quality health services by the poor? A systematic review of the literature. *Int J Equity Health,* 2007; 6:17.

Appendix 1. CONSORT Checklist

 CONSORT 2010 checklist of information to include when reporting a randomised trial*			
Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract	1a	Identification as a randomised trial in the title	_____
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	_____
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	_____
	2b	Specific objectives or hypotheses	_____
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	_____
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	_____
Participants	4a	Eligibility criteria for participants	_____
	4b	Settings and locations where the data were collected	_____
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	_____
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	_____
	6b	Any changes to trial outcomes after the trial commenced, with reasons	_____
Sample size	7a	How sample size was determined	_____
	7b	When applicable, explanation of any interim analyses and stopping guidelines	_____
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	_____
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	_____
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	_____
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	_____
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	_____
	11b	If relevant, description of the similarity of interventions	_____
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	_____
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	_____
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	_____
	13b	For each group, losses and exclusions after randomisation, together with reasons	_____
Recruitment	14a	Dates defining the periods of recruitment and follow-up	_____
	14b	Why the trial ended or was stopped	_____
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	_____
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	_____
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	_____
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	_____
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	_____
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	_____
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	_____
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	_____
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	_____
Other information			
Registration	23	Registration number and name of trial registry	_____
Protocol	24	Where the full trial protocol can be accessed, if available	_____
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	_____

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming; for those and for up to date references relevant to this checklist, see www.consort-statement.org.

Appendix 2. Newcastle – Ottawa Quality Assessment Scale

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

- 1) Is the case definition adequate?
 - a) yes, with independent validation *
 - b) yes, eg record linkage or based on self-reports
 - c) no description
- 2) Representativeness of the cases
 - a) consecutive or obviously representative series of cases *
 - b) potential for selection biases or not stated
- 3) Selection of Controls
 - a) community controls *
 - b) hospital controls
 - c) no description
- 4) Definition of Controls
 - a) no history of disease (endpoint) *
 - b) no description of source

Comparability

- 1) Comparability of cases and controls on the basis of the design or analysis
 - a) study controls for _____ (Select the most important factor.) *
 - b) study controls for any additional factor * (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

- 1) Ascertainment of exposure
 - a) secure record (eg surgical records) *
 - b) structured interview where blind to case/control status *
 - c) interview not blinded to case/control status
 - d) written self report or medical record only
 - e) no description
- 2) Same method of ascertainment for cases and controls
 - a) yes *
 - b) no
- 3) Non-Response rate
 - a) same rate for both groups *
 - b) non-respondents described
 - c) rate different and no designation

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability

Selection

- 1) Representativeness of the exposed cohort
 - a) truly representative of the average _____ (describe) in the community *
 - b) somewhat representative of the average _____ in the community *
 - c) selected group of users eg nurses, volunteers
 - d) no description of the derivation of the cohort
- 2) Selection of the non exposed cohort
 - a) drawn from the same community as the exposed cohort *
 - b) drawn from a different source
 - c) no description of the derivation of the non exposed cohort
- 3) Ascertainment of exposure
 - a) secure record (eg surgical records) *
 - b) structured interview *
 - c) written self report
 - d) no description
- 4) Demonstration that outcome of interest was not present at start of study
 - a) yes *
 - b) no

Comparability

- 1) Comparability of cohorts on the basis of the design or analysis
 - a) study controls for _____ (select the most important factor) *
 - b) study controls for any additional factor * (This criteria could be modified to indicate specific control for a second important factor.)

Outcome

- 1) Assessment of outcome
 - a) independent blind assessment *
 - b) record linkage *
 - c) self report
 - d) no description
- 2) Was follow-up long enough for outcomes to occur
 - a) yes (select an adequate follow up period for outcome of interest) *
 - b) no
- 3) Adequacy of follow up of cohorts
 - a) complete follow up - all subjects accounted for *
 - b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ % (select an adequate %) follow up, or description provided of those lost) *
 - c) follow up rate < ____ % (select an adequate %) and no description of those lost
 - d) no statement