Lee Kuan Yew School of Public Policy
National University of Singapore

469C Bukit Timah Road
Oei Tiong Ham Building
Singapore 259772
Tel: (65) 6516 6134 Fax: (65) 6778 1020
Website: www.lkyspp.nus.edu.sg

Lee Kuan Yew School of Public Policy

Working Paper Series

Design and Scaling up of Policy Experiments and Pilots: Lessons for the Water Sector

Michael Howlett
Department of Political Science
Simon Fraser University
And
Lee Kuan Yew School of Public Policy
National University of Singapore
Email: Howlett@sfu.ca

And

Sreeja Nair
Institute of Water Policy
Lee Kuan Yew School of Public Policy
National University of Singapore
Email: sreeja.nair01@nus.edu.sg

November 25, 2014

Working Paper No.: LKYSPP 14-35 (IWP)

Abstract

The use of experimentation by practitioners and resource managers as a policy instrument for effective policy design under complex and dynamic conditions has been well-acknowledged both in theory and practice. For issues such as water resource management policy experimentation, especially pilot projects can play an important role in exploring alternate courses of action when faced with long-term uncertainty. While the political aspects of experimentation have been alluded to by several policy scholars, there is lack of empirical evidence that explores their interplay with other factors that may also be critical for scaling up of policy experiments. This paper presents a critique of factors that can influence scaling up of policy experiments, including pilots and draws lessons for experimentation in the water sector through a review of selected examples of water policy experiments and a Qualitative Comparative Analysis of pilots in multiple sectors. The analysis reveals that the design of policy experiments apart from being a technical process it is also highly driven by the interests, behavior and attitudes of the stakeholders, building on the argument that scaling up is a "craft rather than science" (Spicer et al, 2014). Presence of strong political support, synergies with ongoing policies and programs and regular monitoring and evaluation are found to be factors necessary for scaling up of pilots. When in combination, these three factors are revealed to also create a sufficient condition for scaling up.

1. Role of experimentation in policymaking

Effectively managing water resources is increasingly becoming a major challenge for policymakers and water managers given that multiple stresses are adversely impacting water resources worldwide. These impacts are being manifested in the form of conflicts over water allocation and use, inadequacy of current water distribution systems, presence of multiple stakeholders and their varied interests leading to competing demands for resource use, environmental stressors such as climate change and its impact on the water cycle (Moore et al, 2014). In addition to current challenges, policymakers and resource managers need to consider how impacts of current and likely new stressors will be manifested on water resources over longer time horizons in the future, in order to undertake effective anticipatory policy planning.

Experiments form a useful policy instrument to manage such complex policy issues and operate under uncertainty by aiding in ex-ante evaluation of policies, generating learning outcomes and policy relevant information under dynamic conditions (McFadgen, 2013). In the development sector, experimental projects have been frequently used to assess alternative courses of action. These include (1) projects that focus on problem definition by assessing evidence of "dissatisfaction or existence of a problem", (2) projects that focus on problems which are partly or wholly undefined, (3) projects that explore the most effective way of achieving pre-set policy goals, (4) projects that aim at identification of gaps and barriers in situations where problems and goals are already well-known and (5) Natural experiments that occur over a period of time without conscious intervention (Rondinelli, 1993). Similar projects can also be observed in the water sector.

Enhanced experimentation and consequent learning can also aid in adapting to the "dynamic drivers and expressions of risk" in a changing policy environment (O'Brien et al, 2012). Pilot projects are a common mode of policy experimentation and a widely used method to introduce major government policies or programmes in a phased manner, allowing them to be "tested, evaluated and adjusted" beforehand (Cabinet Office, 2003). However there exist several challenges in translating or 'scaling-up' of experimental projects, including pilots and their translation into policies (Stoker, 2010). Furthermore, the policy experiment concept itself remains ill-defined and there is little empirical analysis on their effects, in particular how their design influences their potential as "learning incubators" (McFadgen 2013).

While the importance of pilots as a form of experimentation for pre-testing policies and programmes is well-acknowledged, there are challenges in terms of their ability to act as a predictive method beyond the context in which these are applied i.e. when these are scaled up and when these are designed for complex issues and rapidly changing policy environments. Political factors including the influence of diverse stakeholders can impact scaling-up of policy experiments including pilots.

Spicer et al (2014) argue that "scaling up is a craft not a science" alluding to the predominant political nature of the activity compared to its technical aspects. The apparent political nature of pilots has been acknowledged earlier as well. Policy pilots came under much scrutiny during the late 1980s and 1990s as these were often seen as being 'donor-driven', dependent on external aid and less focused on local priorities and engagement. Also, the resource support provided for pilots at a smaller scale seemed to fade out when replication at a larger scale was planned. Some pilot projects may also provide policymakers with an excuse to delay critical large-scale policy reforms (PHR, 2004). Policymakers might also often be hesitant towards policy experiments or pilots owing to issues of 'accepting uncertainty' (Stoker, 2010). This paper presents a critique of factors that can influence scaling up of policy experiments, including pilots and draws lessons for experimentation in the water sector through a review of selected examples of water policy experiments and a Qualitative Comparative Analysis of pilots in multiple sectors.

2. Experimentation in the water sector

Several experiments have been undertaken in the water sector at the national and local scales to provide critical insights to the policy process. The broad classification of experimental projects by Rondinelli (1993) can also be used to characterize different types of water policy experiments and is used to guide the discussion in this paper. Results from laboratory and field experiments have often been used by scholars to provide policymakers and practitioners with evidence of the impacts of selected experimental interventions as well on their feasibility and acceptability by key stakeholders, including the intended beneficiaries. Water policy experiments that fall into the first category of experimental projects based on Rondinelli's classification for example can include *need-based assessments*. A common form of such assessments has been willingness- to-pay surveys which are often used as a proxy to assess the demand for services such as water and sanitation. Pattanayak et al (2006) conducted a willingness-to -pay experiment of 1800 households in Sri Lanka to demonstrate that demand for improvement in water and sanitation services is driven by a combination of several factors such as socio-economic status, costs,

location, means for self-provision and perceptions of stakeholders. Results of the experiment also indicated that while presence of policy incentives such as connection fee subsidy can increase the demand for piped water, the question of whether the benefits accrued by scaling up are more than the costs incurred still remained. Behavioural variables at the level of the individual thus form key decisive factors in influencing the overall outcome of such policy experiments. While behaviour can be regulated with incentives to some extent there are limitations to how observations at the local level can be considered to be a good indicator for the overall success or failure of the experiment when it is scaled up.

In another example of a behavioural experiment in the water sector in the United States, the study of individual level behaviour at the local level provided valuable insights for successful scaling up of the experiment. An experiment was conducted in the state of Georgia to capture 'bidding behaviour' in an "auction-like process", in advance of a similar process to be conducted by the government to pay some farmers to withdraw irrigation in drought years. The results from the multi-site local experiments with farmers were used by Georgia's main water management agency Environmental Protection Division to understand farmer behaviour and accordingly formulate strategies for pricing and closing rules for the actual auction including coordination of the bidding process at multiple locations and reporting of results (Cummings et al, 2002).

The second type of policy experiments focus on problems that are partly or wholly undefined. Typically under such conditions of uncertainty, *policy pilots* are undertaken. Some examples of water management pilots from Europe are discussed here. A pilot on flood control through natural flooding was conducted in Germany by the State government in collaboration with researchers to restore a stretch of the river Rhine to increase the resilience of the ecological system and reduce vulnerability from floods. An integrated water management policy (Integrated Rhine Programme) was designed in which 13 sites were to be restored. One site (Altenheim) was used as a pilot. While the policy makers considered the pilot as being very successful, citizens opposed the pilot due to concerns related to drowning of wildlife and breeding of mosquitoes. These differences in perspectives lead to a very long implementation period of the other sites, even though the pilot was well embedded in the policy programme (Vreugdenhil, 2010).

In another pilot on recharging groundwater wells with local water in Switzerland, university researchers in cooperation with a water production company aimed at ecological enhancement and cost- savings by recharging groundwater wells with local water instead of water from an external source. While the research university considered the results as promising, the water company was not keen on scaling up. The latter is the land owner and has large political support and thus they have the decisive power for scaling up. During the pilot the stakeholder relationship was strained due to mismatch of expectations. The water company thought the pilot was only for research purposes, whereas the university wanted to use it for policy purposes. The water company considered the pilot to be non-representative for scaling up (Vreugdenhil et al. 2009). Another project in the Netherlands called INSIDE[1] involved the national government and commercial partners for developing innovations that allowed strengthening of sea dikes from inside, without having to reconstruct the landscape. No scaling up happened in this case owing to two reasons- the technology was not mature enough for testing and the actor relationships with the consortia were troublesome.

[1]Retrieved from http://www.snellerinnoveren.nl/voorpagina.aspx?id=home_en

The third category of policy experiments involves projects that explore the most effective way of achieving pre-set policy goals. Under dynamic conditions this can involve *innovations and transition experiments.* In the past decade the field of transitions management has gained prominence to explore "a range of possible pathways for change" (Farelly & Brown, 2011). Transitions can be defined as 'a gradual, continuous process of structural change within a society or culture' and are complex, spread over long timeframes, involve multiple actors and occur across multiple levels (Rotmans et al., 2001). Transitions require "steering, facilitation and coordination" and experimentation and learning form important concepts (Farrelly & Brown, 2011).

In the context of urban water sustainability Farelly and Brown (2011) examined eleven local-scale experiments in Australian cities and found sustainable transitions to urban water management required changes in underlying culture and beliefs along with structural reforms. The role of 'bridging organizations' was found to be critical to collate insights from local-scale experiments and inform future policy and practice. In another study from Hyderabad city in India, Nastar (2014) explores the impacts of ongoing legislative, technical, managerial, and social aspects of the urban water regime on the citizens' access to water. The study finds that scaling-up of innovative 'niche experiments' that aim towards transitions in urban water management is often impeded due to system lock-ins and tendency of donor agencies as well as current water policy and urban development initiatives to preserve status-quo. Thus niche experiments cannot successfully scale-up without a facilitating policy environment or space.

A successful case of scaling up of a local urban development project municipality of Egedal in Denmark followed by its integration into a new sewage plan also highlights the key role of practitioners as facilitators of transition to overcome 'lock-ins', challenge existing policy regimes and enabling the aggregation of knowledge at the local level and its generalizability as the project scales up (Zhou et al, 2013). Similarly, taking the specific case of scaling up of efficient water-based technologies, Turton and Bottrall (1997) argue that these need "well-informed individuals' for their scaling up. The scaling-up of such technologies is often impeded owing to the need for collaborative effort and lack of results that are visible in the short-term. In addition, once the pre-requisites for scaling up are understood, these need to be matched with the biophysical and socio-economic context of the areas where scaling up would occur.

The fourth category of policy experiments involves those that aim at *identification of gaps* in current policy practices. While pilots form a useful means to investigate gaps, this intended purpose is not met if errors or gaps identified in the pilot phase are not corrected before scaling up of these pilots. For example, privatization of urban water services provision in Kenya began on an experimental basis followed by large-scale expansion. The objective of these privatization efforts were to decentralize water governance structure to alleviate problems such as unaccounted for water losses, unmetered water usage and uneconomic water usage fee and rates that were linked to a highly centralized water governance. However, the experiment as well as scaling up has not achieved its intended outcomes because both at the local and city scale the privatization efforts were unable to avoid intrusion of central and local government authorities in its functioning. A second cycle of privatization was also attempted but errors/ gaps identified in the first phase were not considered in subsequent efforts, thereby rendering the experiments futile (Akumu and Appida, 2006).

The final category of policy experiments is *natural experiments*. While studying from history i.e. natural experiments in the water sector is helpful however their applicability as a 'blueprint' for similar outcomes in the future is limited for dealing with policy issues such as climate change that face a high degree of uncertainty. The key challenge is that under 'surprise' (Walker et al 2010, Lempert et al 2003) these experiments offer little or no scope for decision-makers to respond from history or experience.

The design of policy experiments thus apart from being a technical process, is also highly driven by the interests, behavior and attitudes of the stakeholders. Compared to the earlier works however that focused more on the *content* of the experiments itself, the more recent literature on experimentation has shifted its focus to the *process* of experimental policy design, including the role of various stakeholders therein (van der Heijdin, 2013). This new wave of "experimentalist governance" presents an iterative process of "provisional goal setting" with the intention of revising the goals based on the learning derived from trying out alternate modes of goal achievement in different contexts (Sabel and Zeitlin, 2012). The next section specifically discusses the characteristics of the scaling up process of policy pilots in general as a mode of policy experimentation.

3. Scaling up of policy pilots

The empirical evidence on the composition of effective policy pilots and the process of their diffusion i.e. continuation or expansion is lacking (Vreugdenhil et al, 2009). Many factors influence the pilot dynamics, including the pilot design and the context. These can include factors such as, the stakeholders involved, that further influences the availability of knowledge and resources, choice of scale and the choice for pilot sites, the mode of governance that influences the nature of stakeholder engagement and learning, the level of innovativeness of the pilot and how it converges or diverges from the current policy context, flexibility to make changes to adapt to local conditions and finally the timing of the strategy for pilot diffusion. Pilot diffusion can face impediments in case a strategy for diffusion management is entirely absent, poor or there is widespread opposition from some critical stakeholders (Vreugdenhil, 2010). Additionally, if the policy change involves significant costs it is likely to motivate policymakers to change and thus increase the 'stickiness' of existing policies (Callander, 2011). As evident from multiple rural development including water resource development initiatives, scaling up of pilots and their sustenance beyond pilot sites operates in conjunction with sustained efforts towards empowerment and capacity building of local communities and beneficiaries of the pilot (Turton and Bottrall, 1997).

Hartmann and Linn (2007) define scaling up as "expanding, replicating, adapting and sustaining successful policies, programs or projects in geographic space and over time to reach a greater number of people". Scaling up occur when a program increases in size, its geographical spread or budget (quantitative); increases in its range of activities and interaction with related programs (functional); increases in political power and engagement with wider political processes (political) or increases in organizational capacities and processes (organizational) (Gillespie, 2004). Such scaling up, whether in space or time, often runs the risk that the initial project objectives and outcomes become less appropriate or relevant for the new context (Simmons et al,

2007). For example, Margerum (2012) presents the successful case of watershed management at the state level in Oregon, United States and argues that success at the watershed level may be rather fragmented and thus may not uniformly translate or scale-up to a larger i.e. river basin scale owing to limitations in stakeholder capacities and quality of coordination efforts. Furthermore, successful small-scale, often non-regulatory approaches such as water management efforts taken at a watershed level might not always be scaled-up successfully to address issues at the larger scale such as river-basin flooding.

The key challenge is thus to identify both context-specific as well as universal elements contributing to scaling up and to ensure that the universal elements are maintained while leaving scope for context-specific changes to take shape through adaptation and learning (Hartmann and Lin, 2007). Hartmann and Lin also identify seven elements as being critical for scaling up of developmental interventions. These factors are obtained from a review of literature and local experiences to identify the driving factors, paths and enabling environment for scaling up inter alia. These factors include, (i) applying leadership, vision and values; (ii) managing political constituencies; (iii) ensuring supportive policies; (iv) developing institutional capacity; (v) creating incentives and accountability; (vi) practicing evaluation, learning and feedback; and (vii) planning for success (see Table 1).

Table 1: Key factors for scaling up (Hartmann and Lin, 2007)

Factors	**Description**
Leadership, vision and values	Presence of leaders driving the scaling up with a clear vision, enabling institutions to exemplify a set of values for achieving scaling up, to avoid "short-termism" of programs and "fragmentation of effort".
Presence of political constituencies	Scaling up is supported by political constituencies. This entails the active engagement of political players in the scaling up process and its placement on their agendas, driven by "need and appropriateness" rather than any personal interest and guarded from "elite capture".
Presence of supportive policies, programs and projects	Presence of a supportive policy framework (laws, regulations, norms and linkages with related policies, programs and projects) for scaling up.
Strong institutional support and capacities to facilitate change	Scaling up requires adequate institutional and human capacities and additional training and development and institutional capacity building. These efforts also need to be constantly evaluated in their performance relative to appropriate benchmarks, while ensuring accountability.
Incentives and accountability	Incentives for stakeholders form a critical factor for enabling leadership, political support and institutional capacity for scaling up. Accountability for scaling up on the other hand is essential to ensure that incentives are in sync with some shared objectives of the stakeholders.
Effective monitoring & evaluation (M & E)	Monitoring and Evaluation focusing on scaling up as a key indicator of success can assess the impact of the program and obtain feedback for improvement, and thus build a case for garnering political and stakeholder support and sustainability of the program.
Scaling up benefits from an orderly and gradual process of planning	A systematic and gradual process, careful planning, and clear demarcation of roles and responsibilities of partners and strong communication channels are important factors for scaling up.

The Hartmann and Lin framework is applied to selected cases of policy pilots in multiple sectors to further investigate the characteristics of factors that can influence scaling up of these pilots. In the absence of similar information for detailed analysis using water policy pilots, a Qualitative Comparative Analysis (QCA) from multiple sectors is presented in this paper with the aim of providing a methodological framework and insights for conducting similar analysis for the water sector. Another objective is to extend the argument of scaling up being a "craft rather than science" and study it in multiple sectors. The term policy pilots as used in this paper refers to pilot projects initiated by governments for policy purposes, including testing potential policies, implementing policies that have difficulties in being implemented and evaluating new policies at an early stage. Using a fuzzy-set Qualitative Comparative Analysis this paper identifies sets of factors that are conditional and sufficient for the scaling up of the selected policy pilots.

4. Insights from Qualitative Comparative Analysis of policy pilots in multiple sectors

The direct motivation for conducting a QCA in different sectors came from the review of an ongoing pilot in India. This is the Weather Based Crop Insurance Scheme (WBCIS) that was launched by the Government of India (GoI) in 2007 as a pilot in selected areas across India to assess it as an alternative to the ongoing National Agricultural Insurance Scheme (NAIS) and to bring more farmers under the aegis of crop insurance (AFC, 2011).

To include diverse sectors in this analysis, a narrative review is undertaken to explore the quantitative scaling up of pilots i.e. increase in geographic spread and coverage of beneficiaries owing to limited online availability of detailed documentation on different pilot projects. A narrative review "summarizes different primary studies from which conclusions may be drawn into a holistic interpretation contributed by the reviewer's own experience, existing theories and models" (Kirkevold, 1997). Results are of a qualitative meaning and it helps synthesize the "diversities and pluralities of understanding around scholarly research topics" (Jones, 2004). A narrative review is undertaken for this paper because there is lack of theoretical frameworks and synthesis of evidences from the large number of operational/ abandoned pilots to glean common factors that make pilots scale up.

A narrative review of published articles on pilot projects in different countries and sectors was conducted using Googlescholar. The search was limited to a 10 year time-period (2003-2013) and yielded a large number of articles that were reduced to ten cases after a review of the abstract that were selected based on the following criteria:

- Articles should be first-hand documentations of individual pilot projects with detailed analysis
- Articles should refer to pilots that were consciously launched with the objective of scaling up and guiding future policy development.

The factors that influenced scaling up of the pilots were identified. A combination of the following keywords was used to conduct the online search for articles: 'scaling up', 'policy pilots', 'success', 'failure', 'diffusion of policy pilots', 'evaluating policy pilots', 'replication'. Ten cases of policy pilots spanning health care, poverty alleviation and agriculture risk

management, with different levels of diffusion and designed and governed in different ways are discussed in this section.

4.1 Fuzzy set Qualitative Comparative Analysis

Qualitative Comparative Analysis (QCA) has been designed by Prof. Charles C. Ragin and colleagues at the University of Arizona, United States and is an analytical technique based on Boolean algebra to allow for comparison of qualitative cases that are often large enough to do in-depth qualitative analysis and too small to do variable-oriented quantitative analysis[2]. The variables in QCA are either presented as Crisp sets i.e. binary sets that denote presence or absence (1 or 0 respectively) of "membership" in a specific category. A "fuzzy set" splits this all-or-none categorization into further categories using scores from 0 to 1 (Ragin, 2006). The objective of QCA is to enable causal interpretation in addition to detailed qualitative information that is obtained from case studies, in order to understand the different combination of plausible factors that could lead to a specific outcome (Ragin, 2008).

Fuzzy set Qualitative Comparative Analysis (fsQCA) is "a program that uses combinatorial logic, fuzzy set theory and Boolean minimisation to work out what combinations of case characteristics may be necessary or sufficient to produce an outcome" (Kent, 2008). fsQCA was specifically designed for analysis of small-n and medium-n datasets. There are specific cases where fsQCA is particularly helpful. This includes instances where there is an hypothesis regarding the underlying causal factors affecting the outcome being studied (scaling up in this case), when different combinations of these plausible causal factors could give rise to the outcome and conditions are sufficient only when they are in combination, when results need to be interpreted as "necessary and sufficient conditions", when the number of cases is very low for conventional quantitative methods to be applied, when a good deal is known regarding the cases, and when the key concepts are clearly defined and measured (Ragin, 2008).

a. Constructing a data matrix

The first step in fsQCA is to construct a data matrix, which lists certain characteristics of the cases as variables. These characteristics denote the "degrees of membership" of a defined category. Fuzzy set can allow for scores in between 0 and 1 to denote various degrees of membership. For this study a 4-point fuzzy set has been used where the membership in a particular category has been denoted in the following way:

1= fully in, 0.67= more in than out, 0.33= more out than in, 0= fully out……(1)
Here the data matrix consists of the 10 case studies as rows and the 7 factors to be tested as columns, including an additional column called 'scaling up'. This column marks whether scaling up has happened or not and to what extent based on categorization discussed in (1).

[2] For further details see http://www.u.arizona.edu/~cragin/fsQCA/index.shtml.

b. Constructing the truth table

Next, the truth table is constructed marking 'scaling up' as the 'outcome' that the paper wishes to assess based on membership scores or 'conditions' (causal factors) that may be necessary or sufficient for the outcome to happen. The truth table considers each case as a combination of the characteristics selected. Normally, four kinds of result can be expected in the truth table:

- Combination of specific characteristics lead to positive outcomes,
- Combination of specific characteristics lead to negative outcomes,
- There are contradictory cases i.e. a specific combination leads to positive outcomes in some cases and negative in others, and
- No cases for specific combinations: This is likely for small-n studies, wherein there will be many combinations of characteristics that are possible but not observed in any of the cases (due to the small sample size). Hence in these cases it is also not possible to say whether the outcome occurred or not (termed 'remainders' in fsQCA).

This study is a small-n type hence the remainders are excluded from the analysis. There are also no contradictory cases that were found in this study. Studying the truth table can give a big picture of the variety of combinations of characteristics that are common or those that happen often or seldom. It is difficult to observe this diversity in the small-n analysis conducted for this paper though.

c. The analysis of sufficient and necessary conditions

The 'truth table' presents the different combinations of causal factors that have met specified criteria of sufficiency for the outcome to occur. This suggests that the membership score on the outcome is always higher than the membership score of the causal combination. The analysis of necessary conditions in fsQCA assesses individual causal factors that may be necessary for the outcome to occur. This suggests that the membership score on the outcome is usually always lower than the membership score of the causal factor being investigated. In other words, when X (causal factor) is considered as a necessary condition for Y (outcome) to occur, it means that Y cannot occur without X, i.e. Y (outcome) is a subset of X (causal factor). On the other hand, when X is considered as a sufficient condition means that if Y is present X must be present too. This however does not mean that X by itself will cause Y (i.e. there may be other factors influencing Y too). In other words, in this case X (causal factor) becomes a subset of Y (outcome) (Ragin, 2008).

d. Consistency and Coverage

Some other results that are provided by the QCA include Consistency i.e. degree to which the cases sharing a specific combination of causal factors share the same outcome and coverage, i.e. the degree to which a specific causal combination accounts for occurrence of an outcome. Raw coverage measures the "proportion of memberships in the outcome explained by each term of the solution". Unique coverage measures "the proportion of memberships in the outcome explained solely by each individual solution term (i.e. memberships that are not covered by other solution terms)" (Ragin, 2006).

e. Description of case studies and assigning fuzzy values

The fuzzy values for the 10 cases are assigned after reviewing the cases and understanding the case context. Fuzzy values are assigned for the degree of presence of each of the 7 plausible causal factors for scaling that is being tested in the hypothesis here. Table 2 presents an overview of the 10 cases that have been considered in this paper.

Table 2: Overview of cases[3]

S. no.	Country	Pilot title	Objective	Scaling up	Fuzzy score	Reason for score
1	China	Methadone Maintenance treatment (MMT)	Increasing the coverage of MMT, its beneficiaries and improving accessibility of services	The project moved from being a pilot in 8 sites in 2004 to a nation-wide programme covering 27 provinces by the end of 2009.	0.67	Though the pilot spread geographically, its coverage of beneficiaries is low
2	India	Kudumbashree	A multisectoral poverty alleviation program initiated by the Government of Kerala, India to eradicate poverty in the state by 2008.	In 1991, the Govt. of Kerala (GoK) state, India and UNICEF initiated Community-Based Nutrition Program (CBNP) in Alleppey town to improve the health and nutritional status of children and women. CBNP facilitated collective action by forming community development societies for women. Based on the positive experiences, GoK scaled up the program to the entire state in 1998 under the name Kudumbashree.	1	Complete scale-up in terms of spread and coverage of intended beneficiaries
3	Vietnam	Injectable contraception and quality of care	Scaling up introduction of the injectable contraceptive depot-medroxyprogesterone acetate (DMPA) as part of health intervention packages to improve the quality of care in the family planning programme.	After a strategic assessment of the need for contraceptive introduction and pilot testing of the interventions in three provinces of Vietnam, these interventions were scaled up to 21 of 64 provinces in the country.	1	Complete scale-up in terms of spread and coverage of intended beneficiaries
4	Pakistan	Rural support	Bottom-up, community driven development using politically neutral approach.	The program started in 1982 in remote, rural parts of Northern Pakistan and by 2004 it covered almost all Northern districts	1	Complete scale-up in terms of spread and coverage of intended beneficiaries
5	Zambia	Urban poverty	Alleviate urban poverty	The PROSPECT ended in 2004 as a 6-year	0.33	Full scale-up not

[3] The list of references of the case study articles has been provided separately in the reference list

S. no.	Country	Pilot title	Objective	Scaling up	Fuzzy score	Reason for score
		alleviation-Programme of Support for Poverty Elimination and Community Transformation(PROSPECT)	through empowerment of poor communities and enabling their participation in decision making and building collective capacities to act.	follow-up to two previous projects that operated one after the other from 1992. PROSPECT operated in only 13 of Lusaka's total 37 compounds, reaching 300,000–400,000 of the estimated population of 800,000 in Lusaka's informal settlements.		achieved despite being a 6 year program; low coverage of intended beneficiaries
6	Cambodia	Health Equity Funds (HEFs)	Use Health Equity Funds for translation into health policies for the poor to promote equity	HEFs pilots were initiated in 2000 in two urban slums and were translated into a national health policy; and scaled up to 50 HEF schemes based in 51 hospitals in Cambodia.	1	Complete scale-up in terms of spread and coverage of intended beneficiaries
7	China	Quality of care	Promote family planning and limit births as part of China's sustainable development goals	Initiated in 6 counties in 1995 and scaled into a national reform effort in China by 2004	1	Complete scale-up in terms of spread and coverage of intended beneficiaries
8	Ghana	Community-based Health Planning & Services	Guide national reforms for supporting community-based primary health care	Pilot launched in 1994 in three villages and by 2003 the initiative had become a national initiative for district planning process and community-based health care.	0.67	Scaled up geographically, but coverage of intended beneficiaries is low.
9	Thailand	100% condom programme	Control Sexually-Transmitted Diseases in sex workers	Initiated in one province in 1989 and it was agreed to be implemented in all provinces in 1991 owing to the success of the pilot.	1	Complete scale-up in terms of spread and coverage of intended beneficiaries
10	India	Weather-based insurance scheme (WBCIS)	Improving crop insurance cove rage to farmers in India in addition to traditional crop insurance	As of 2013, WBCIS has been scaled up to 21 states in India starting from one district in 2007.	0.33	Full scale up to national level not achieved; low coverage of intended beneficiaries.

The results following the fsQCA and in-depth analysis of the cases are presented in this section in three parts:

(1) Results of the truth table which is constructed by putting the values of occurrence of the outcome i.e. scaling up and degree of presence of the seven potential causal factors that are being tested.
(2) Results from assessment of the necessary conditions
(3) Results from assessment of the sufficiency conditions

f. Assessment of the Truth table

A 4-point fuzzy set is used to mark the outcomes and plausible causal factors in the truth table (Table 3).

Table 3: 4-point fuzzy set for outcome i.e. scale-up

0	0.33	0.67	1
No scale up	Scale up to some extent	Scale up to a large extent	Full scale-up

Scaling up as referred to in this paper refers to quantitative scale-up i.e. increase in spread of the pilot and coverage (of beneficiaries). The plausible causal factors are also ranked based on the 4-point fuzzy set (Table 4):

Table 4: 4-point fuzzy set for causal factors

0	0.33	0.67	1
Causal factor not present	causal factor present to some extent	causal factor present to a large extent	causal factor present significantly

The truth table is presented in Table 5. It shows the cases on the left-hand side. The next column shows the outcome, i.e. in this case the degree to which scale up has occurred (or not) in the particular case. These gradations of the 7 causal factors (that are being tested) are used to indicate the degree to which the causal factor was present in that case. These are determined based on details regarding the case study provided in the article, and the presence of the specific factor and attribution of scaling up to the presence of that plausible causal factor.

Table 5: Truth table

Case	**scaleup**	**leadvision**	**Polsupp**	**policies**	**instcap**	**incentives**	**moneval**	**Planning**
ChinaMMT	0.67	0	1	1	0.67	0	1	0.67
IndiaKudum	1	0.67	1	1	0.67	0.33	0.67	0.33
VietnamDMPA	1	0	1	0.67	0.67	0	0.33	0.67
PakRural	1	1	0.33	0.67	0.67	0	1	1
ZambiaPov	0.33	0.33	0.33	0.33	0.33	0	0.33	0.33
CambHealth	1	0	1	1	0	0	1	1
ChinaFP	1	0.67	1	0.67	0.67	0	1	0.67

GhanaCHPS	0.67	0.33	0.67	1	0.67	0	1	0.67
ThaiAIDS	1	0	1	1	0.67	0	1	0.67
IndiaWBCIS	0.33	0	0.67	0.67	0.33	0.33	0.67	0.33

Legend: scaleup= outcome i.e. degree of scale-up; leadvision= presence of leader, vision, values; Polsupp= political support; policies= synergy with current policies/ programs; instcap= institutional support and capacities; incentives= presence of incentives and accountability; moneval= monitoring and evaluation; planning= gradual process with detailed planning, clear communication and engagement of stakeholders with clear demarcation of roles and responsibilities

Based on a review of the 10 cases, it is found that 6 cases were examples of successful scale-ups in terms of increase in their spread and coverage of beneficiaries. Two cases were scaled-up to some extent while two others were scaled up to a large extent but not completely. 8 cases were initiated by the national/ state government except for the Rural Support Programme in Pakistan which was run by a non-governmental organization, Aga Khan Foundation, and PROSPECT in Zambia which was support by the U.K. development aid agency Department for International Development.

g. Assessment of necessary conditions

The fsQCA is used to run an assessment of necessary conditions. The results are presented in Table 6.

Table 6: Assessment of necessary conditions

Factor	**Consistency**	**Coverage**
leadvision	0.375	1
polsupp	*0.9162*	*0.9162*
policies	*0.8762*	*0.8752*
Instcap	0.6687	1
incentives	0.0825	1
moneval	*0.875*	*0.875*
planning	0.7925	1

A score of above 0.8 is considered to be good for acceptance of a causal factor as a necessary condition (Kent, 2008). Table 6 can thus be interpreted in the following way:

Presence of political support is necessary for scaling up in 91.6% of the cases studied. This is closely followed by support from existing policies in 87.6% of the cases, and monitoring and evaluation which is necessary for scaling up in 87.5% of the cases studied.

h. Assessment of sufficiency conditions

The truth table is now analyzed using the fsQCA software for sufficient conditions (Table 7). Based on the hypothesis for this paper, the model used suggests that scale-up is considered to be a factor of all 7 causal factors i.e.

Scaleup= f (planning, moneval, incentives, instcap, policies, polsupp, leadvision)

Table 7: Assessment of sufficient conditions

fs/QCA

File Analyze Graphs Window

```
**********************
*TRUTH TABLE ANALYSIS*
**********************

File:  C:/Users/Sre/Desktop/Final pilots.csv
Model: scaleup = f(planning, moneval, incentives, instcap, policies, polsupp, leadvision)

 Rows:       32



 Algorithm: Quine-McCluskey
      True: 1
   0 Matrix: 0L
Don't Care: -

--- INTERMEDIATE SOLUTION ---
frequency cutoff: 1.000000
consistency cutoff: 0.885906
Assumptions:
planning (present)
moneval (present)
incentives (present)
instcap (present)
policies (present)
polsupp (present)
leadvision (present)

                                              raw       unique
                                           coverage    coverage   consistency
                                          ----------  ----------  ----------
moneval*policies*polsupp                  0.750000    0.208750    0.899550
planning*instcap*policies*polsupp         0.583750    0.042500    1.000000
planning*moneval*instcap*policies*leadvision  0.291250    0.042500    1.000000
solution coverage: 0.835000
solution consistency: 0.908844
```

Analysis of the results presented in Table 7 can be interpreted as follows:
Presence of either of the following combinations (planning*moneval*instcap*policies*leadvision) or (planning*instcap*policies*polsupp) is completely consistent i.e. are *sufficient* to ensure scale-up. This is followed by the combination (moneval*policies*polsupp) which can explain 89.9% of the occurrence of scale-up in the cases studied. Secondly, the *raw coverage* value indicates that the presence of (moneval*policies*polsupp) can explain 75% of the scaling-up that occurs. The *Unique coverage* value indicates that when only (moneval*policies*polsupp) is present it can explain 20.8% of the occurrence of scale-up in the cases studied. It should be noted that the unique coverage value is very low for all combinations because the number of cases studied here is very less and hence the diversity is very less to come across instances of observing exclusive patterns of many combinations.

5. Implications for the water sector

Policy experimentation has been well-acknowledged as a useful policy tool to deal with complex and dynamic policy issues. Different types of policy experiments including pilot projects have been conducted in the water sector and these have provided useful insights to water resource managers for policy design. Policy pilots may be scaled up in space, time or based on their purpose. Individual factors that are considered important for scaling up can be studied in combination with others to see the impact various combinations can have on scaling up. This can be particularly important when governments and other agencies operate with limited resources, and thus can invest these resources in enhancing these specific factors in a targeted manner.

The political aspects of policy experiments including pilots are not very well researched i.e. pilots might sometimes be used as an excuse to garner political acceptability, or maybe abandoned citing them as failures because the political milieu might not be conducive for it to move ahead. Presence of multiple stakeholders and their power positions can also influence the scaling up process. Many water policy experiments also depend on behavioral variables, making scaling up efforts more challenging as it requires an extrapolation of behavior observed at an individual level. While incentives can be used regulate to regulate behavior to some extent, mismatch of expectations or disagreement between stakeholders can impede the scaling up process despite successful results at the local level (Vreugdenhil, 2010). If the experiments are challenging an established water management regime by suggesting innovative policy solutions and alternate pathways for resource management and transitions, collaboration between the key stakeholders is critical to break policy inertia and system lock-ins (Zhou et al, 2013).

The second part of this paper set out to study lessons from pilots from different sectors across the world as an attempt to garner empirical evidence of why some pilots are scaled up whereas others are not. A fuzzy set Qualitative Comparative Analysis has been conducted to quantify some of the observations from these pilot studies. Presence of strong political support, synergies with ongoing policies and programs and regular monitoring and evaluation are found to be factors necessary for scaling up of pilots. When in combination, these three factors are revealed to also create a sufficient condition for scaling up. Policies and programmes that govern the management of water resources including their use, conservation and allocation among diverse users are also politically sensitive especially when the resource transcends geographical and jurisdictional boundaries (Vivekanandan and Nair, 2009). Thus political support is likely to be a very strong factor in scaling up water pilots also.

A major challenge in this analysis has been that in-depth analysis and lessons from pilots are rarely reported, and when they are these usually refer to success factors and seldom to factors leading to failure. The main sources providing the details of the case studies are the research articles only. In addition, pilot projects that have been considered under the narrative review cover different time periods, and might have changed over time. This paper only explores the process of quantitative scaling up and not the quality of the scaling up, for e.g. in terms of services offered, whether it has been scaled up equally well in all places etc. The individual factors that are considered to be critical causal factors for scaling up as obtained from Hartmann and Linn's framework for this paper may also influence each other. For the purposes of the fuzzy set QCA conducted for this paper these factors have been treated as mutually exclusive.

Most of the reported pilot studies that are reported do so in a very 'context-specific' or 'limited by context' manner that generalization of these lessons to another context remains largely unexplored. By doing a synthesis from ten different case studies operating in different sectors, scales and countries, this paper makes an attempt to deviate from this pattern and draw useful insights for the water sector. While QCA offers a very useful method to convert qualitative observations from case studies into figures that can enable comparison across case studies, it should be noted that conducting the QCA and trying to interpret the results would not make much sense without an in-depth understanding of each case study. Thus the results from QCA should always be interpreted in conjunction with a detailed analysis of each case study individually.

Additionally, 'controlled experiments' at the local level can sometimes mask the financial, social and economic risks that become evident when these are scaled up (Nordblom et al, 2011). Thrush et al (1997) for example conducted predictive modelling in marine ecosystems and argued that though projections at the large-scale are required for decision-making and resource management purposes, these often involve the risk of reduced precision or confidence in results. Thus there are limits to extrapolation from experiments at local scales. This is primarily because results from ecological phenomenon observed at lower scales cannot simply be aggregated to the larger-scale and this process is error-prone. In addition, biological processes "may have emergent properties at larger scales".

References

AFC (Agricultural Finance Corporation Ltd.), 2011. Report on impact evaluation of pilot weather-based crop insurance study (WBCIS). Submitted to Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India.

Akumu, O. A. K and Appida, P. O., 2006. Privatisation of urban water service provision: the Kenyan experiment. Water Policy 8 (2006) 313–324.

Cabinet Office, 2003. Trying it out: The role of 'pilots' in policy-making. Report of a Review of Government Pilots. Strategy Unit, Government of the United Kingdom.

Callander, S., 2011. Searching for good policies. American Political Science Review. Volume 105 / Issue 04 / November 2011, pp 643-662.

Cummings, R. A., Charles A. Holt, and Susan K. Laury, 2002. Using laboratory experiments for policymaking: an example from the Georgia irrigation reduction auction. Retrieved from http://people.virginia.edu/~cah2k/water.pdf

Farrelly, M. and Brown, R., 2011. Rethinking urban water management: Experimentation as a way forward? Global Environmental Change 21, 721–732.

Gillespie, S., 2004. Scaling up community-driven development: a synthesis of experience. International Food Policy Research Institute.

Hartmann, A. and Linn, J., F., 2007. Scaling Up: A Path to Effective Development. 2020 Focus Brief on the World's Poor and Hungry People. Washington, DC: IFPRI

Jones, K., 2004. Mission drift in qualitative research, or moving toward a systematic review of qualitative studies, moving back to a more systematic narrative review. Qualitative Report, 9 (1), pp. 95-112.

Kent, R., 2008. Using fsQCA: A Brief Guide and Workshop for Fuzzy-Set Qualitative Comparative Analysis, University of Stirling, United Kingdom

Kirkevold, M., 1997. Integrative nursing research – an important strategy to further the development of nursing science and practice. Journal of Advanced Nursing 25, 977-984.

Lempert, Robert J., Popper, Steven W. and Bankes, Steven C. (2003). Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis. Santa Monica, CA: RAND.

Margerum, R., 2012. Integrated water resources management in the United States: The Rogue and Willamette River cases. In Making Space for the River: Governance Experiences with Multifunctional River Flood Management in the US and Europe, edited by Jeroen Warner, Jeroen Frank Warner, Arwin Van Buuren, Jurian Edelenbos. IWA Publishing.

McFadgen, B., 2013. Learning from Policy Experiments in Adaptation Governance. Paper presented at the International Conference on Public Policy. Grenoble, France.

Moore, M., Suzanne von der Porten, Ryan Plummer, Oliver Brandes, Julia Baird, 2014. Water policy reform and innovation: A systematic review. Environmental Science and Policy. 38, 263 – 271.

Nastar, M. 2014. What drives the urban water regime? An analysis of water governance arrangements in Hyderabad, India. *Ecology and Society* 19(2):57. http://dx.doi.org/10.5751/ES-06570-190257

Nordblom, T. L., Reeson, A. F., Finlayson, J. D., Hume, I. H., Whitten, S. M. and Kelly, J. A., 2011. Price discovery and distribution of water rights linking upstream tree plantations to downstream water markets: experimental results. Water Policy Vol 13 No 6 pp 810–827.

O'Brien, K., M. Pelling, A. Patwardhan, S. Hallegatte, A. Maskrey, T. Oki, U. Oswald-Spring, T. Wilbanks, and P.Z. Yanda, 2012: Toward a sustainable and resilient future. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 437-486.

Pattanayak and Jui-Chen Yang, 2006. The Use Of Willingness To Pay Experiments : Estimating Demand For Piped Water Connections In Sri Lanka. In George Van Houtven, Caroline van den Berg, Subhrendu K. Policy Research Working Papers. January 2006

PHR (Partners for Health Reformplus), 2004. The Role of Pilot Programs: Approaches to Health Systems Strengthening. Bethesda, MD: PHRplus, Abt Associates Inc.

Ragin, C., 2006. Set relations in social research: evaluating their consistency and coverage. Political Analysis Advance Access. Oxford University, pp 20.

Ragin, C., 2008. Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA). In Benoit Rihoux and Charles Ragin (editors), Configurational Comparative Analysis, Sage Publications, 2007.

Rondinelli, D., 1993. Development projects as policy experiments: An adaptive approach to development administration. 2nd edition. New York: Routledge

Rotmans, J., R. Kemp.and M.B.A. van Asselt, 2001. More Evolution than Revolution. Transition Management in Public Policy, Foresight 3 (1), 15-31

Sabel, C., and Zeitlin, J., 2012. Experimentalist governance in David Levi-Faur (ed.), The Oxford Handbook of Governance, Oxford: Oxford University Press, pp. 169-83.

Simmons, R., Fajans, P. and Ghiron, L., (eds.), 2007. Scaling up health service delivery: from pilot innovations to policies and programmes. WHO, 2007.

Spicer, N., Bhattacharya, D., Dimka, R., Fanta, F., Mangham-Jefferies, L., Schellenberg, J., Tamire-Woldemariam, Walt, G. and Wickremasinghe, D., 2014. Scaling-up is a craft not a science': Catalysing scale-up of health innovations in Ethiopia, India and Nigeria. Social Science & Medicine 121, 30-38.

Stoker, G., 2010. Translating experiments into policy. The ANNALS of the American Academy of Political and Social Science 2010 628: 47.

Thrush SF, Schneider DC, Legendre P, Whitlatch RB et al, 1997. Scaling-up from experiments to complex ecological systems: where to next? J Exp Mar Biol Ecol, 216:243–254.
Turton, C. and Bottrall, A., 1997. Water resource development in the drought-prone uplands. Overseas Development Institute. Number 18.

van der Heijden, J in press (early view online 2013), 'Experimentation in policy-design: Insights from the building sector', Policy Sciences, DOI: 10.1007/s11077-013-9184-z

Vivekanandan, J. & Nair, S., 2009. Climate change and water: Examining the interlinkages In Troubled Waters Climate Change, Hydropolitics, and Transboundary Resources, Eds. Michel, D. and Pandya, A. The Henry L. Stimson Center, Washington D. C.

Vreugdenhil, H., 2010. Pilot projects in water management: Practicing change and changing practice. PhD thesis, Delft University, The Netherlands, pp 280.
Vreugdenhil, H., Frantzeskaki, N., Taljaard, S., Ker Rault, P. and Slinger, J., 2009. The next step in policy transitions: Diffusion of pilot projects. 13th Annual Conference of the International Research Society for Public Management. April 6th – 8th, 2009, Copenhagen Business School, Denmark.

Walker W, Marchau V and Swanson D, 2010. "Addressing deep uncertainties using adaptive policies". Technological Forecasting & Social Change, Vol. 77, No. 6, Special Section 2.
Zhou, Q., Maj-Britt Quitzau, Birgitte Hoffmann, Karsten Arnbjerg-Nielsen, 2013. Towards Adaptive Urban Water Management: Up-Scaling Local Projects. International Journal of Architecture, Engineering and Construction. Vol 2, No 1, March 2013, 63-72

Case references

India- Weather Based Pilot: AFC (Agricultural Finance Corporation Ltd.), 2011. Report on impact evaluation of pilot weather-based crop insurance study (WBCIS). Submitted to Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India.
100% Condom programme Thailand: Rojanapithayakorn, W., 2003. The 100 Per Cent Condom Use Programme: A Success Journal of Health Management, 5: 225.
PROSPECT Zambia: Mattingly, M., 2008. PROSPECT Lost: When a pilot project does not look to learn. Public Admin. Dev. 28, 129–137.
Pakistan: Rural Support Programme: Rasmussen, S. F., Piracha, M., M., Bajwa, R., Malik, A. and Mansoor, A., 2004. A case study from Reducing Poverty, Sustaining Growth What Works, What Doesn't, and Why: A Global Exchange for Scaling Up Success Scaling Up Poverty Reduction: A Global Learning Process and Conference. Shanghai, May 25–27, 2004. Pakistan: Scaling Up Rural Support Programs.
Health Equity, Cambodia: Por, I., Bigdelic, M., Meessenb, B. and Dammeb, W. V., 2010. Translating knowledge into policy and action to promote health equity: The Health Equity Fund policy process in Cambodia 2000–2008. Health Policy 96, 200-209.

Quality of care, China: Kaufman, J., Erli, Z. and Zhenming, X., 2006. Quality of Care in China: Scaling up a Pilot Project into a National Reform Program. Studies in Family Planning, Vol. 37, No. 1, pp. 17-28.

India, Kudumbashree: Kadiyala, S., 2004. Scaling up Kudumbashree: collective action for poverty alleviation and women's empowerment. International Food Policy Research Institute, Discussion Paper 180.

Ghana community health services: Nyonator, F., Awoonor-Williams, K., Phillips, J., A., Jones, T. C. and Miller, R. A., 2005. The Ghana Community-based Health Planning and Services Initiative for scaling up service delivery innovation. Health Policy and Planning 20 (1), 25-34.

MMT treatment China: Yin,W., Hao,Y., Sun,X., Gong,X., Li, F., Li, J., Rou,K., Sullivan,S. G., Wang,C., Cao,X., Luo, W. and Wu, Z., 2010. Scaling up the national methadone maintenance treatment program in China: achievements and challenges. International Journal of Epidemiology. 39: ii29–ii37.

Vietnam contraceptive care: Strategic choices in scaling up: introducing injectable contraception and improving quality of care in Viet Nam. Fajans, P., Thom, N. T., Whittaker, M., Satia, J., Mai, T. T. P., Dinh Can, T., Nhan, D. T. T., and Newton, N., 2007. In Simmons, R., Fajans, P. and Ghiron, L., (eds.), 2007. Scaling up health service delivery: from pilot innovations to policies and programmes. WHO, 2007.