Chapter 3
Knowledge, Framing, and Ethics in Programme Design and Evaluation

Suraj Jacob

Abstract This chapter will explore ethical issues surrounding the design and evaluation of public health programmes. For programme design, the chapter will argue that programme choice often occurs with solutions already in mind and that these solutions reflect "off-the-shelf" thinking (for instance, ubiquitous "training workshops"), implying little real "choice" in programme design. Further, at a broader level, programme choice is influenced by implicit ideological and epistemological positions that may be ethically dubious especially if they are not problematised and made transparent. On programme evaluation, the chapter focuses on ethical aspects of three key elements: participatory evaluation, the use of evaluation results and the place of impact evaluation. The chapter concludes with a discussion of the role of ethics in relation to epistemology. While it may be relatively uncontroversial to note the problematic ethics of research that comes up short when benchmarked against its own research / methodological paradigm, it is worth asking to what extent the choice of research / methodological / epistemological paradigm is itself an ethical one.

After independence, health policymaking in India gravitated to controlling infectious diseases and family planning through a techno-managerial approach rather than focusing on the primary health system (Amrith 2007).[1] From the 1990s, public health budgets were cut, private sector players encouraged, and a programmatic, verticalised approach to disease control became a feature of government policy. Rao (2017:21) notes: "Complex health problems were simplified into single-line technical solutions – DOTS [Directly Observed Treatment, Short Course] for TB (Tuberculosis), immunization for infant mortality, early diagnosis and

---

[1] This went against earlier attempts at an inclusive model emphasising community engagement and environmental hygiene for positive health and wellbeing (for instance, see the ideas of the Bhore Committee). There were also occasional (but failed) attempts at an integrated approach later on, for instance, the National Health Policy of 1983.

S. Jacob (✉)
Vidya Bhawan Society, Udaipur, India

Visiting Faculty, Azim Premji University, Bengaluru, India

distribution of chloroquine tablets for malaria, and cataract surgeries for blindness". Although many of these programmes were integrated into a broader platform from 2005 (National Rural Health Mission, later National Health Mission), health *yojanas* (programmes) continue to be an important feature of policy. They target a variety of matters, such as institutionalised delivery (e.g. Janani Suraksha Yojana), health insurance (e.g. Rashtriya Swasthya Bima Yojana or Yeshasvini), adolescent health (Rashtriya Kishor Swasthya Karyakram), and so on.

In the backdrop of the importance of programmes in healthcare and health policy in India, this chapter explores ethical issues surrounding the design and evaluation of such programmes. In this chapter, a "programme" is taken to be an organised effort to address a public health need. Programme "design" refers broadly to how a public health approach is chosen and a programme constructed and justified with its expected processes and impact specified.

The chapter highlights a common tendency to design programmes without seriously problematising programme choice and assumptions, and it illustrates this through a case study (Sect. "Programme choice and design"). Programme "evaluation" refers broadly to the systematic investigation of the roll-out, processes and impact of a programme. The chapter highlights participatory approaches to evaluation (Sect. "Participatory and group-sensitive evaluation"), problematises the emphasis on impact evaluation in most mainstream evaluations (Sect. "The place of impact evaluation"), and explores what becomes of evaluation results (Sect. "Use of evaluation results"). The presentation reveals that discussions of ethical issues in public health programme design and evaluation inevitably confront the fundamental question of knowledge construction and power, and the chapter concludes with some thoughts on ethics and epistemology (Sect. "Conclusion").

Programme Choice and Design

"Off-the-Shelf" Programmes

How do policymakers choose and construct programmes? A careful approach would start with a wide-ranging and genuine "needs assessment" to identify the nature of the need/problem and the context within which it is located. A given need/problem can be potentially addressed in a variety of programmatic ways. Nevertheless, in reality, programme choice often occurs with solutions already in mind. Such solutions inevitably reflect "off-the-shelf" thinking rather than "out-of-the-box" thinking – for instance, ubiquitous "training workshops" are a ready solution – implying little real choice in programme design.[2] As Watkins et al. (2012:17) note, in designing programmes often "needs" are conflated with solutions in terms of inputs and activities, such as in the following statements: "What we really *need*

---

[2] Further, where needs assessment is attempted, it is often reduced to a cursory "baseline survey" generating descriptive statistics rather than deeper probing of the situation to construct an understanding of the "how" and the "why" of the need/problem.

is to hire more staff members", "I *need* more resources", "They *need* more training". Rossi, et al. (2004:156) note:

> The human service professions operate with repertoires of established modes and types of intervention associated with their respective specialty areas. As a result, program design is often principally a matter of configuring a variation of familiar "off the shelf" services into a package that seems appropriate for a social problem without a close analysis of the match between those services and the nature of the problem.

A careful approach would brainstorm with intended programme beneficiaries, programme personnel, and others regarding different alternative solutions and their corresponding requirements and implications. Consider, for instance, the policy reaction to the tragic 2012 rape incident in Delhi ("Nirbhaya") that set off a large storm of protest. An immediate programmatic response by the government was to try to establish "women police stations".[3] While this may or may not be an appropriate response, it is not clear that needs were carefully articulated or alternative solutions gauged. Ubiquitous public service messages are another example of off-the-shelf thinking. In the context of family planning programme, Joseph (2004) writes:

> The Post & Telegraph Department made marriage greetings in attractive envelopes that curtly said, "For happy married life please be in touch with the Family Welfare Centre". In the late '60s the government pleaded, "Two or Three Children...Then Stop". When the fertile '60s ended, the campaign cancelled the kind option of the third child. "We Two Ours Two, Horn Ok Please," the highway trucks began to say. But all that came to nothing.

The argument here is not that it was wrong to have such exhortative messages. Rather, it is that the approach followed likely did not weigh alternative solutions in a participatory manner, going instead for off-the-shelf solutions that probably had limited impact.[4] The question then is: why are programme choices and assumptions not questioned more often? Part of it may be due to a disconnect between planners and programme designers on the one hand and programme managers and implementers on the other, and part of it may be due to a status quoist tendency (Kaplan and Garrett 2005). Further, at a broader level, programme choice is influenced by implicit ideological and epistemological positions that may be ethically dubious especially if they are not problematised and made transparent. The subsection below illustrates this in the context of a specific programme.

---

[3] "All Delhi police stations to have women officers: Shinde", *Times of India*, December 29, 2012; "Only 442 women police stations across India: Police research data", *The Hindu*, December 25, 2012.

[4] More recently, the Health Ministry launched a "high-octane campaign with a three-in-one message of family planning, child spacing, and safe sex practices" ("What's the family plan", *The Hindu*, April 17, 2016).

Case Study: Tamil Nadu Integrated Nutrition Project (TINP)

The Tamil Nadu Integrated Nutrition Project (TINP) emphasised nutrition education, growth monitoring, and supplementary nutrition for pregnant and lactating mothers as well as infants. It ran as a large-scale intervention (in about 20,000 villages) in Tamil Nadu state over a long period (phase I in the 1980s, phase II in much of the 1990s, and phase III merging with the Integrated Child Development Service programme from the late 1990s). The World Bank considered the programme impactful/successful (World Bank 1994) and designed a similar programme in Bangladesh, the Bangladesh Integrated Nutrition Project (BINP) (Cartwright and Hardie 2012; White and Masset 2007).

Consider the logic (and implicit ethics) of programme choice in TINP. The programme was put in place after extensive studies and needs assessments as well as discussions of the state government with the World Bank. Heaver (2002:7) notes:

> TINP I was designed following a 1970s analysis, the Tamil Nadu Nutrition Study, which was probably the most elaborate nutrition study undertaken in the developing world by that time. In addition to carrying out anthropometry, it studied the consumption aspects of nutrition, food production and processing, and the relationship between income and employment growth and nutrition in the state. The study's broad scope led to wide-ranging nutrition strategy discussions between the state government and the Bank, and in turn to the initial identification of a complex, multi-sectoral project with components for growth monitoring, behavioral change, food supplementation, health care, food production, and food processing. In the course of a 16-month dialogue during project preparation, this initial broad menu of activities was reduced to the more limited TINP focus on growth promotion and improvement of maternal and child health services.

TINP was designed keeping in mind results from previous studies of nutrition in Tamil Nadu. Specifically, the focus on child nutrition came from studies of the strong relationship between child mortality and child nutrition. The focus on pregnant and lactating women and pre-school children also came from previous studies. The focus on nutrition education for mothers and growth monitoring of infants came from studies showing that child malnutrition occurred in many families with income levels that could ensure adequate diet. And the focus on supplementary feeding (i.e. only for children below a specified growth threshold and only until they crossed that threshold) was to help reduce dependency on the programme as well as to sharpen focus.[5] Targeting specific women and children, as well as emphasising supplementary rather than regular feeding, also helped with managerial and financial viability.

Based on fieldwork exploring how TINP played out in four caste communities, Sridhar (2008, 2010) provides empirical evidence to question assumptions partly based on which the programme was designed – for instance, that specific behavioural

---

[5] That the programme was designed explicitly for supplementary rather than regular feeding cannot seriously be considered an ethical problem – feeding was not designed to directly redress diet or calorie problems and with carefully justified reasons. Further, in practice, programme personnel did not withhold food from children who expressed hunger even when they did not qualify based on the threshold (Sridhar 2008).

practices related to breast-feeding, washing, and balanced diet were obstacles to better nutrition and health and that behavioural change could occur through appropriate counselling and education. Sridhar (2010:128) points out instances not of ignorance and inappropriate or even uncaring practices but rather of structural limitations created by poverty and the labour market: "a tribal woman was told that she should breastfeed her child. However, she noted that she was unable to comply with this advice during the day because she was a daily labourer". Further, structural limitations were not just external to the household. Sridhar reports that in a majority of instances, women did not have control over household expenditure. Although many respondents knew about the importance of protein, vegetables, and a balanced diet, there was a substantial problem of diversion of household income by males for alcohol consumption. Overall, Sridhar (2010:130–1) notes:

> … women thus have knowledge about childcare, yet are constrained by lack of control over income and time. Women often mentioned lack of time, money, or control over household expenditure to explain why their child was not healthy. For example, when a woman stopped breastfeeding before the World Health Organization's recommended twelve months period, it was usually because she had limited time, insufficient milk, or because she was sick. Many of the low caste and tribal women had to work in the fields all day and upon returning felt that since they had been separated for more than eight hours from their child, their breasts were engorged and the milk had become sour. Instead of asking the health workers for more formula, out of fear of being chided, mothers would hide the fact that they were not breastfeeding. As a result, they would end up giving the baby sugar water, or cow or water buffalo milk. (van Hollen 2003)

From the perspective of ethics, what are we to make of this structural critique? Despite the exemplary due diligence displayed by programme designers in crafting programme components in response to what was known and understood about malnutrition in Tamil Nadu, given the training and inclinations of the economists and public health experts at World Bank and in the Tamil Nadu government, the fact is that programme designers were operating with particular biomedical and health economics models. Sridhar (2010:123) argues:

> Within the Bank, the discipline of economics is hegemonic... an economic approach reduces problems, such as nutrition, to their core elements so that the professional expertise can digest them and prescribe solutions. … So the Bank framed nutrition using biomedical and economic inputs because it ultimately had to construct a problem that its own instruments could address. The Bank is in the lending business. It makes time-bound, repayable loans. Any Bank actions have to fit within the overall Bank goal of lending for growth.

The economic-biomedical approach did not sufficiently acknowledge the socio-economic positionality and structural disadvantage of the women and children that TINP was designed to address. In fact, a nonstructural approach implicitly creates a "framework of blame" where "a mother is admonished by health workers who simply assume that an undernourished child is the result of a mother's inadequacy" (Sridhar 2010:131). If this happens, it is inherently contradictory and ironic from an ethical perspective since a programme intended to improve the wellbeing of women and children may actually end up reducing the wellbeing of some.

Programme Evaluation

For programme evaluation, the chapter explores three aspects: participatory evalua-
tion, the place of impact evaluation and the use of evaluation results.

Participatory and Group-Sensitive Evaluation

One principle of humanistic research and development practice, which extends to
evaluation, is that of genuine, democratic engagement with individuals and groups
encompassed by the research/evaluation. This extends both to data collection and to
analysis – for instance, the participatory techniques spearheaded by Robert
Chambers and associates (Chambers 2007) – as well as adopting an equity-sensitive
lens to explore impact processes and outcomes associated with an intervention.
Crishna (2006) notes four dimensions of participatory evaluation: those involved in
the programme have some control over evaluation processes, evaluation objectives
are jointly set by them, difficulties faced within this group are addressed as part of
the evaluation, and the process is collectively empowering.

There is an insightful literature on community-based health planning, monitor-
ing, and evaluation in India (for instance, Khanna 2013; Shukla et al. 2014).
Chouinard (2013:238) distinguishes the sensibility of technocratic and participatory
approaches:

> Evaluation, as it is envisioned in the current governance context, is intended to serve pri-
> marily an instrumental role… as a way to legitimize government activities, ensure cost-
> effectiveness, and enhance managerial decision making… The contrast between approaches
> that are more sensitive and responsive to community needs, such as participatory or col-
> laborative approaches to evaluation, and accountability-driven technocratic approaches, is
> stark… [The mainstream, technocratic] approach to evaluation may well fall short of cap-
> turing the range of local views, contextualized meanings, and culturally relevant perspec-
> tives that are increasingly relevant today, and that participatory and collaborative approaches
> to evaluations are intended to capture.

Besides sensibility, such approaches are more explicit about seeing interven-
tion processes and outcomes from a perspective of justice (Khanna 2013). For
instance, equity-focused evaluation focuses on "assessments of what works and
what does not work to reduce inequity" with an emphasis on "difficult-to-reach
socially marginalized groups" (UNICEF 2011:9). Gender-sensitive evaluation is a
particularly important aspect, especially in a development context where gender-
related catch phrases have become the norm.[6] Emphasising specifics such as wom-
en's control over their own bodies, political and legal spaces, private and public

---

[6] Govinda (2012) notes several: "analysing gender inequality", "promoting gender equity", "main-
streaming gender", "engendering development", and "gender sensitisation".

resources, and labour and income, gender-sensitive evaluation highlights the processes and extent to which an intervention changes and transforms gender relations (Chigateri and Saha 2016). UN Women (2015) notes:

> Gender-responsive evaluation has two essential elements: what the evaluation examines and how it is undertaken. It assesses the degree to which gender and power relationships—including structural and other causes that give rise to inequities, discrimination and unfair power relations, change as a result of an intervention using a process that is inclusive, participatory and respectful of all stakeholders (rights holders and duty bearers).

Besides gender, and more generally, evaluation has sought to take perspectives related to human rights (Kemp and Vanclay 2013), social impact (Esteves et al. 2012) and environmental impact (Morgan 2012). In the Indian context, besides class and gender, analyses based on social groups structured by caste, tribe and religion are particularly important from an equity perspective.

The Place of Impact Evaluation

Programme evaluation is about understanding how a programme unfolds (including in comparison with original expectations based on planning and the theory of change), why it unfolds as it does, and with what consequences (intended and unintended). The latter component – namely, consequences of programme roll-out or "impact evaluation" – has gained increased importance in recent times relative to other components such as "process evaluation". Since it is about an intervention and its consequences, impact evaluation has attracted the attention of those working on causal methodology in positivist social science. This literature has produced interesting methodological innovations centred around the logic of a "treatment group" and its counterfactual ("control group") where difference across otherwise "equivalent groups" is interpreted as the causal impact of the intervention (Cook and Campbell 1979; Dunning 2012; Duflo et al. 2007).

Not only is evaluation increasingly focused on impact evaluation, but impact evaluation is itself increasingly focused on the experimental RCT (randomised controlled trial or clinical trial) model, to which I turn below.

Equipoise and the Randomised Controlled Trial

An important ethical challenge of RCTs is that a potentially beneficial intervention is given to one group of individuals and deliberately denied to another group.[7] And the ethical justification is that RCTs can potentially generate new knowledge that can be used for future decision-making regarding programmes and policies that can improve wellbeing. That is, a necessary condition to ethically justify an RCT is that

---

[7] Here, I am not taking up the more obvious ethical issue of doing harm, even unintentionally, as in the famous Stanford prison experiment (Zimbardo 1973).

it should reduce uncertainty about whether and how much a programme is impactful; this is referred to as "equipoise" (Freedman 1987).[8]

There are prominent studies in which the equipoise principle was violated, for instance, the Tuskegee Study of Untreated Syphilis (1932–1972) in which poor African-American men with untreated syphilis were followed even after the effectiveness of penicillin was established and it became widely available (Caplan 2001).[9] In such situations, sometimes a counterfactual ethical justification/argument is made: if the study participants were unlikely to have received the known effective treatment (in this case, penicillin) in the absence of the study, then the study by itself is not compromised ethically. However, there are in fact well-known ethical protocols – importantly, the Helsinki Declaration originating in 1964 (Riis 2003) – that require implementation of the "established effective intervention" in control groups (CIOMS 2002, Guideline 11). Such protocols protect vulnerable populations and especially those in the developing world in today's globalised research context (Angell 1997).

There is a further ethical nuance to consider. Outside of clinical medicine per se, and more properly in the domain of public health and social policy, many RCTs are conducted primarily to understand the *extent* of impact rather than to gauge *whether* or not the corresponding intervention will have positive impact. For instance, in the case of an RCT involving a vaccination camp or provision of midday meals, the logic of equipoise and established effective intervention are not at work – unlike, say, with a potential curative drug. Rather, the vaccination camp or midday meal RCT seeks primarily to gauge the extent of impact (on health, nutrition, and other indicators) in specific socio-economic and geographical contexts long after the clinical or health "value" of such interventions has been established. For instance, Banerjee et al. (2010) frame their RCT on vaccination camps by noting explicitly

---

[8] Freedman (1987:141) notes that equipoise is "a state of genuine uncertainty … regarding the comparative therapeutic merits of each arm in a trial". For him, "clinical equipoise" is when there is "genuine uncertainty" on the part of the "expert medical community – not necessarily on the part of the individual investigator –about the preferred treatment". However, this is not without its critics. For instance, Miller and Brody (2003: 20) critique equipoise for viewing "clinical trials through a therapeutic lens". Relatedly, Clayton (1982) distinguishes between an "individual ethic" (avoid harm, provide equal benefit to each individual) and a "collective ethic" (acquire new knowledge so that individuals may benefit in the future). In this rendering, RCTs can be justified ethically when the latter counters the former.

While equipoise is about uncertainty regarding knowledge, there is also the related matter of clinical trials with methodological failures that have ethical consequences. May (1975: 25) notes that "one of the most serious ethical problems in clinical research is that of placing subjects at risk of injury, discomfort, or inconvenience in experiments where there are too few subjects for valid results, too many subjects for the point to be established, or an improperly designed random or double-blind procedure".

[9] The study had other ethical violations as well, for instance, deliberate deception of participants. Subsequently, a larger literature and consensus have developed around ethical dos and don'ts regarding research on human subjects, and these have been institutionalised in specific research contexts (for instance, in Institutional Review Boards).

that "[i]mmunisation is a highly cost effective way of improving survival in children in developing countries" and that yet there is little consensus on the best (institutional or other) way for implementing it. In other words, equipoise is invoked, but it is "implementation equipoise" – to coin a related term – and not clinical equipoise.[10]

Unfortunately, the world of "what works" in policy implementation is sparse in settled, context-appropriate knowledge. For instance, in the case of vaccination, Banerjee et al. (2010) find that small material incentives may be more cost-effective than simply improving reliable supply. Yet we know little about what aspects of the specific context of the RCT (NGO intervention among poor tribal villagers of southern Rajasthan) may have contributed to the results and whether and how the results carry over to other contexts (nontribal villages even in the same region or peri-urban areas). Indeed, the absence of a sizeable literature on the context-specific nature of implementation (Konkipudi and Jacob 2017), and therefore also impact, provides a relatively strong case for implementation equipoise even though the literature does not articulate it in such terms. This holds in broader terms for all "pilot projects" where an intervention is tested out on some subset of a target population.

Humanistic Critique

RCTs and other interventions with an evaluative component (such as pilot projects) sometimes tend to have other characteristics that have attracted critiques. One critique relates to the ethics of experimentation that disproportionately falls on the marginalised, of which the Tuskegee study is an example. Reddy (2012: 68) points out:

> … the relative disorganisation of poor people and poor communities, which makes it possible to experiment upon them without facing much resistance… It would very likely be politically unpalatable to provide a recognisable benefit randomly to middle-class or upper-class individuals or communities, and, at a minimum, to do so would involve complex political negotiation.

A second critique is that evaluation-centric interventions tend to take a non-humanistic, engineering approach where intervenors stand in a privileged position outside the community/system in which the intervention occurs, as suggested by the TINP case study. Sridhar (2010:128) suggests that programme instructions to frontline workers took an economic-biomedical approach that did not encourage autonomy or empathy in programme personnel: "at the front line, the community workers did not have enough time to discuss caring practices with women… more emphasis [was] placed on producing written educational materials than on actually speaking to women about health and nutrition".

---

[10] However, this has come in for criticism. For instance, Glewwe et al. (2012) report an RCT gauging the extent to which students with eyesight problems do better at school if they wear corrective eyeglasses. For a critique from the perspective of clinical equipoise, see Ziliak and Teather-Posadas (2016).

Further, the overall approach of TINP may also have tended to view programme personnel, especially frontline workers, as cogs in the intervention-evaluation machinery. Sridhar (2010:127) notes: "The health workers routinely collected massive amounts of data, usually not fully understanding why they were doing it or what the information would be used for". Similarly, Mishra (2014:966) notes that frontline health workers in Odisha "ensured that all state health services were moulded to accommodate local aetiologies of illness and remedies… However, these efforts were confined to the community space and never shared with senior health officials, lest this might be seen as promoting quackery".

Impact Evaluation: Causation Without the "Why"

The discussion regarding ethical complexities with RCTs and the non-humanistic approach of evaluation-centric interventions more generally is not applicable to impact evaluation approaches that are not intervention-based. Such approaches, often called natural experiments or quasi-experiments, have ethical and logistical advantages over experimental/intervention-based approaches (Cook and Campbell 1979; Dunning 2012). Nevertheless, both intervention-based and quasi-experimental approaches can be critiqued for an epistemology that is ethically problematic, as discussed below.

In impact evaluation, impact is "the positive and negative, intended and unintended, direct and indirect, primary and secondary effects produced by an intervention" (definition from the OECD Development Assistance Committee; see Rogers 2012). Almost the entire focus on impact evaluation is to legitimately estimate the size/extent of impact that can be causally attributed to an intervention. Given this, experimental or quasi-experimental evaluation cannot explain how impact occurs. Addressing the "how" question would require understanding the processes and contexts within which interventions play out. Nevertheless, positivist methodologies of impact evaluation are framed explicitly around their ability to produce estimates of the causal impact of an intervention on an outcome. The irony is that these approaches claim to estimate empirical causal impact without actually empirically explaining why the impact is what it is estimated to be – in other words, causality which does not address the "why"!

Why is it important for evaluation to understand impact processes – how and why a particular impact obtains – rather than only estimate causal impact? Dalkin et al. (2015) note:

> …it is necessary to establish what goes on in the system that connects its various inputs and outputs… In pharmacology, the term 'mechanism of action' refers to the specific biochemical interaction through which a drug substance acts on the body to generate its curative effect. Programme evaluators do not suppose that CCTV (the intervention) causes a fall in crime rates (the outcome). It does so, when it does so, by persuading potential perpetrators of increased risks of detection (the mechanism). In all cases, science delves into the 'black box'.

Impact evaluation studies typically specify the theoretical mechanism(s) linking intervention and projected outcomes, but even if it is empirically established that the outcomes occurred as projected, it does not sufficiently establish the validity of the proposed theoretical mechanism as empirical causal explanation (Woodward 2017). Rather, what is needed is empirical tracing out of the process(es) through which the intervention in a particular context produced the outcomes causally attributed to it (Jacob et al. 2015). As Smith et al. (2002:72) note, we often intuitively assume processes without empirically establishing them:

> It makes intuitive sense that better working conditions would improve productivity. In the same way it is clear that street lighting or closed-circuit television (CCTV) should work by increasing surveillability. Plague reduction should work by drowning witches, and fever should be reduced by the extraction of overheated blood. We have been too ready to assume that how crime prevention *should* work is the way crime prevention *does* work.

Biomedical studies show the enormous variation in impact and point the way to the importance of understanding process and context. For instance, studies based on mice and meant to throw light on human physiology may not even be generalisable to all mice contexts since it turns out that the specific type of mouse used in laboratory experiments (Black 6) is "a teenaged, alcoholic couch potato with a weakened immune system, and he might be a little hard of hearing" (Engber 2011). In the case of TINP, based on its perceived success (George et al. 1993; World Bank 1994), the World Bank conceived and funded a similarly massive programme in Bangladesh. Nevertheless, BINP has been judged to be a failure even apart from implementation weaknesses (White and Masset 2007; World Bank 2005). Cartwright and Hardie (2012) note that an important contextual difference – and one that programme designers and implementers ignored because of their a-contextual approach to "scaling up" – was that in Tamil Nadu decisions on infant nutrition revolve around the mother, but in Bangladesh they revolve around the paternal grandmother. This implies that programmes such as BINP focused on changing the knowledge and behaviour of mothers are less likely to create large positive impact.

Writing about public health in India, Prashanth et al. (2013:36–37) note that the challenge is about "focusing on 'how' interventions to improve health services work in addition to understanding if they work". They scrutinise all published evaluations of public health programmes in India over a 5-year period and note that few trace processes of impact and indeed only 34 of 93 even reported the context within which the intervention produced the outcomes that were being reported. They explicitly note a sad reality in the literature (Prashanth et al. 2013:44): "In spite of several health and related social initiatives taken up under the National Rural Health Mission (NRHM) and various national-level schemes in the country, the policy-relevant question what worked for whom and under what conditions remains largely unanswered".

Although impact evaluation studies seldom undertake process tracing, in fact, the impact evaluation approach offers a potentially excellent starting point for doing so, namely the "heterogeneity effect": estimated causal impact of an intervention is

an average that may be distributed widely.[11] Deaton (2010:430) notes: "heterogeneity is not a technical problem calling for an econometric solution but a reflection of the fact that we have not started on our proper business, which is trying to understand what is going on". If the estimated average impact is relatively higher among some individuals or groups compared to others, it forms an excellent starting point to explore causal processes and contexts for explaining why this is so. Consider the impact evaluation study by Lim et al. (2010) of the Janani Suraksha Yojana (JSY), a conditional cash transfer scheme for promoting institutional delivery. The study uses a quasi-experimental methodology to establish that on average JSY reduced perinatal and neonatal mortality rate. However, it also finds considerable heterogeneity around this average based on region, socio-economic group, and so on. Prashanth et al. (2013:39) note:

> It concludes that conditional cash transfer schemes such as JSY, through incentivizing institutional delivery, could 'somehow' improve particular health outcomes. However, crucial information for policymakers and decision-makers at district and sub-district level is related to why there were differences among beneficiaries, and why within districts some women did not avail the scheme, or did not benefit from the scheme in spite of availing the scheme.

Local specifics of caste and class matter enormously for public health processes. Mishra (2014:968) notes that in one Odisha village, "villagers even refused to lead us to the ASHA's residence and later complained about how indifferent and inactive an ASHA she was. We learned that this ASHA's recruitment had been controversial, as she was married to a rich contractor and a converted Christian, pitting her as someone who is privileged and unfit to relate to the concerns of the villagers". More generally, Mishra (2014:961) notes:

> Existing literature on global programmatic experiences with integration of health services tend to approach delivery of services as a technical and mechanistic process … The focus on supply-side health system factors, though important, assumes that community demand relies unproblematically on providers delivering services…. Ethnographic evidence suggests that the demand/uptake of health services is linked to a host of factors, such as the community's perceived vulnerability to a specific illness for which the health service is offered, previous experiences with other state health services, modes of health communication, interaction with health workers and broader political identities and perceptions of the state by the community…

Woolcock (2013:230) critiques inconsistency in approach to rigour in the typical impact evaluation approach:

> … having expended enormous effort and resources in procuring a clean estimate of a project's impact, and having successfully defended the finding under vigorous questioning at professional seminars and review sessions, the standards for inferring that similar results can be expected elsewhere or when 'scaled up' suddenly drop away markedly. The 'rigorous result', if 'significantly positive', translates all too quickly into implicit or explicit claims

---

[11] Interestingly, participatory and group-sensitive evaluation approaches naturally take process and context more in their stride than do conventional approaches: "Equity-focused evaluations pay particular attention to process and contextual analysis, while conventional impact evaluation designs use a pre-test/post-test comparison group design, which does not study the processes through which interventions are implemented nor the context in which they operate" (UNICEF 2011:9–10). See also Batliwala and Pittman (2010).

that the intervention now has the status of a veritable 'best practice', the very 'rigor' of 'the evidence' invoked to promote or defend the project's introduction into a novel (perhaps highly uncertain) context, wherein it is confidently assumed that it will also now 'work'.

Finally, and quite apart from what impact evaluation cannot tell us – although implicitly insisting it does so – because of absence of empirical process tracing, impact evaluation approaches are also limited as they cannot tackle several important questions, viz. those not amenable to construction of relevant control groups. Reddy (2012:63) notes that "[t]hese include questions related to the structure and dynamics of markets, governmental institutions, macroeconomic policies, the workings of social classes, castes, and networks, and so forth".

Use of Evaluation Results

Why is programme evaluation conducted? There are, of course, standard, legitimate answers to this question. Evaluation can help improve programme implementation in ongoing and later rounds, keep programme personnel accountable, decide whether a programme should be continued, and foster organisational learning (Alkin 2010). However, evaluation sometimes becomes merely "ritualistic" in nature (Rossi et al. 2004) – mandated by programme sponsors/donors and conducted desultorily, often to justify predetermined funding and policy choices and with little commitment to use evaluation findings meaningfully.[12] This turns the spotlight on the motivation of the evaluation sponsor and the purpose and credible uses of evaluation.[13]

In India, the Planning Commission took a proactive interest in programme evaluation in its early years and created an independent and technically competent agency (the Planning Evaluation Office, PEO) to evaluate plan programmes. Further, PEO reports were sent to Members of Parliament, and some public discussion occurred (Pal and Chakrabarti); they were also an important agenda item at the annual conference of State Development Commissioners (Mehrotra 2013). From the 1970s there was a decline in the importance of evaluation, partly due to decline in the role of the Planning Commission itself as an agent of development. In the late 1970s, the Dubhashi Committee on the PEO noted the low use of evaluation reports and highlighted both supply-side weaknesses (poor evaluation) as well as inadequate

---

[12] In the context of ethics, it is also worth raising the question of who the evaluator should be answerable to. In practice, typically accountability is to evaluation sponsors and programme funders alone rather than to intended beneficiaries and specific marginalised groups, and this is particularly problematic where evaluation is merely "ritualistic".

[13] Although in the text I do not discuss the problem of poor-quality evaluation reports, this is also an important reality in the Indian context, particularly when "[e]valuations are typically carried out by professionals who have neither an evaluation background nor a good understanding of how governments function", so that evaluation reports merely "contain generalised statements" rather than contextually relevant recommendations based on real-world processes and pragmatic judgements (Kumar 2010:239).

institutional follow-up of evaluation reports. However, even in the more robust evaluation period earlier, there was little evaluation capacity available outside the Planning Commission, particularly in state governments, and evaluation was seen primarily from the perspective of intra-government accountability rather than social accountability (Pal and Chakrabarti). Notes Kumar (2010:239):

> Anything passes off for evaluation…. A tradition of evaluation is yet to permeate the administrative, bureaucratic, and political cultures of most South Asian countries. Most managers are fearful of evaluation; they see it as an audit or a fault-finding exercise.

With greater pluralisation in governance over time, particularly from the 1990s, civil society groups became more active in evaluation and accountability in India. There was also a turn towards donor-imposed evaluation frameworks, but this was "to evaluate against donor criteria for donor needs—and according to the criteria and methods the donor has preselected" (Carden 2010:220).[14]

Conclusion

Much of the discussion of ethics in this chapter was interwoven with issues of knowledge construction and framing. One aspect is the ethical implication of insufficient fealty to research methodology and presentation within the approach chosen by an evaluation. Unfortunately, this applies to a large number of public health evaluation studies from India. As Altman (1980:1183) notes, "it is unethical to publish results that are incorrect or misleading…. Once published, a piece of research achieves both respectability and credibility…" While it may be relatively uncontroversial to note the problematic ethics of research that comes up short when benchmarked against its own research/methodological paradigm, it is worth asking to what extent the choice of research / methodological / epistemological paradigm is itself an ethical one. Scientism – "the conviction that we can no longer understand science as one form of possible knowledge, but rather must identify knowledge with science" (Habermas 1971:4, cited in St. Pierre 2006) – can produce a closing off of alternative knowledges, which has ethical implications.

An ethic and ethics of *anekantavada* (multifacetedness of knowledge and truth) would imply a very different approach to public health programme design and evaluation. The routine application of off-the-shelf programme ideas would need to be rethought (and the shelf likely dismantled), as would the broader technocratic approaches that currently hold sway in programme design (as evidenced by the biomedical, health economics approach of TINP). In the field of evaluation, participatory co-production of knowledge would need to be emphasised, changes in

---

[14] Of the three types of "evaluation orientation" distinguished by Carden and Alkin (2012) – use-oriented approaches, values-oriented approaches, and methods-oriented approaches – donor-driven evaluation focuses more on the third and is particularly weak on the second (for instance, genuinely participatory methodologies, as discussed earlier), whereas both the second and the first are likely of greater relevance for the programme beneficiaries and the programme itself.

social relations highlighted, the present emphasis on impact estimation broadened to an understanding of underlying processes and contexts and the evaluation results better integrated with broader policymaking.

References

Alkin, M. C. (2010). *Evaluation essentials: From A to Z.* New York: Guildford Press.

Altman, D. (1980). Statistics and ethics in medical research: Misuse of statistics is unethical. *British Medical Journal, 281*.

Amrith, S. (2007, January 13–19). Political culture of health in India: A historical perspective. *Economic and Political Weekly*, *42*(2), 114–121.

Angell, M. (1997). The ethics of clinical research in the third world. *New England Journal of Medicine, 337*(12), 847–849.

Banerjee, A. V., Duflo, E., Glennerster, R., & Kothari, D. (2010). Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *BMJ, 340*, c2220.

Batliwala, S., & Pittman, A. (2010). *Capturing change in women's realities: A critical overview of current monitoring and evaluation frameworks*. Toronto: Association for Women in Development.

Caplan, A. L. (2001). Twenty years after: The legacy of the Tuskegee syphilis study. In *Bioethics, justice and health care* (pp. 231–225). Belmont: Wadsworth-Thomson Learning.

Carden, F. (2010). Introduction to the forum on evaluation field building in South Asia. *American Journal of Evaluation, 31*(2), 219–221.

Carden, F., & Alkin, M. C. (2012). Evaluation roots: An international perspective. *Journal of MultiDisciplinary Evaluation*, *8*(17), 102–118.

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. New York: Oxford University Press.

Chambers, R.(2007). From PRA to PLA and pluralism: Practice and theory. In *The SAGE handbook of action research: Participative inquiry and practice* (p. 297). London: Sage.

Chigateri, S., & Saha, S. (2016). *Gender transformative evaluations*.

Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation, 34*(2), 237–253.

CIOMS (Council for International Organizations of Medical Sciences). (2002). International ethical guidelines for biomedical research involving human subjects. *Bulletin of Medical Ethics, 182*, 17.

Clayton, D. G. (1982). Ethically optimised designs. *British Journal of Clinical Pharmacology, 13*(4), 469–480.

Cook, T. D., & Campbell, D. T. (1979). *Quasi experimentation: Design and analytical issues for field settings*. Chicago: Rand McNally.

Crishna, B. (2006). Participatory evaluation (I)–sharing lessons from fieldwork in Asia. *Child Care, Health and Development, 33*(3), 217–223.

Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). What's in a mechanism? Development of a k concept in realist evaluation. *Implementation Science, 10*, 49.

Deaton, A. (2010, June). Instruments, randomization, and learning about development. *Journal of Economic Literature, 48*, 424–455.

Duflo, E., Glennester, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. In *Handbook of development economics*.

Dunning, T. (2012). *Natural experiments in the Social Sciences: A design-based approach*. New York: Cambridge University Press.

Engber, D. (2011, November 15). The mouse trap (part I): The dangers of using one lab animal to study every disease. *Slate*. Available at http://www.slate.com/articles/health_and_science/the_mouse_trap/2011/11/lab_mice_are_they_limiting_our_understanding_of_human_disease_.html

Esteves, A. M., Franks, D., & Vanclay, F. (2012). Social impact assessment: The state of the art. *Impact Assessment and Project Appraisal, 30*(1), 34–42.

Freedman, B. (1987). Equipoise and the ethics of clinical research. *New England Journal of Medicine, 317*(3), 141–145.

George, S. M., Latham, M. C., Frongillo, E. A., Abel, R., & Ethirajan, N. (1993). Evaluation of effectiveness of good growth monitoring in south Indian villages. *The Lancet, 342*(8867), 348–352.

Glewwe, P., Park, A., & Zhao, M. (2012). *Visualizing development: Eyeglasses and academic performance in primary schools in China*. Center for International Food and Agricultural Policy Research, University of Minnesota, Working Paper WP12-2 (Jan.)

Govinda, R. (2012). Mapping 'gender evaluation' in South Asia. *Indian Journal of Gender Studies, 19*(2), 187–209.

Habermas, J. (1971). *Knowledge and human interests* (J. J. Shapiro, Trans.). Boston: Beacon Press.

Heaver, R. (2002)*. India's Tamil Nadu nutrition program: Lessons and issues in management and capacity development*. HNP discussion paper series. World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/13787 license: CC BY 3.0 IGO.

Jacob, S., Natrajan, B., & Patil, I. (2015). Explaining village-level development trajectories through schooling in Karnataka. *Economic & Political Weekly, L52*, 54–64.

Joseph, M. (2004, August 23). What if 'Hum do, Hamare Do' had worked? *Outlook India*.

Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning, 28*, 167–172.

Kemp, D., & Vanclay, F. (2013). Human rights and impact assessment: clarifying the connections in practice. *Impact Assessment and Project Appraisal, 31*(2), 86–96.

Khanna, R. (2013). *Ethical issues in community based monitoring of health programmes: Reflections from India*. Paper 3 in COPASAH Series on Social Accountability.

Konkipudi, K., & Jacob, S. (2017). Political pressures and bureaucratic consequences: Vignettes of Janmabhoomi implementation in Andhra Pradesh. *Studies in Indian Politics, 5*(1), 1–17.

Kumar, A. K. (2010). Shiva "a comment on 'evaluation field building in South Asia: Reflections, anecdotes, and questions". *American Journal of Evaluation, 31*(2), 238–240.

Lim, S. S., Dandona, L., Hoisington, J. A., James, S. L., Hogan, M. C., & Gakidou, E. (2010, June 5). India's Janani Suraksha Yojana, a conditional cash transfer Programme to increase births in health facilities: An impact evaluation. *Lancet, 375*, 2009–2023.

May, W. W. (1975). Composition and function of ethical committees. *Journal of Medical Ethics, 1*(1), 23–29.

Mehrotra, S. (2013). *The government monitoring and evaluation system in India: A work in progress*. ECD working paper series No. 28.

Miller, F. G., & Brody, H. (2003). A critique of clinical equipoise: Therapeutic misconception in the ethics of clinical trials. *Hastings Center Report, 33*(3), 19–28.

Mishra, A. (2014). 'Trust and teamwork matter': Community health Workers' experiences in integrated service delivery in India. *Global Public Health, 9*(8), 960–974.

Morgan, R. K. (2012). Environmental impact assessment: The state of the art. *Impact Assessment and Project Appraisal, 30*(1), 5–14.

Pal, S. P., & Chakrabarti, M.. Reforming India's evaluation architecture: The role of Independent Evaluation Office (unpublished work).

Prashanth, N. S., Marchal, B., & Criel, B. (2013). Evaluating healthcare interventions: Answering the 'how' question. *Indian Anthropologist, 43*(1), 35–50.

Rao, K. S. (2017). *Do we care: India's health system*. Delhi: Oxford University Press.

Reddy, S. G. (2012). Randomise this! On poor economics. *Review of Agrarian Studies, 2*(2), 60–73.

Riis, P. (2003). Thirty years of bioethics: The Helsinki declaration 1964-2003. *New Review of Bioethics, 1*(1), 15–25.

Rogers, P. J. (2012). *Introduction to impact evaluation*. Impact Evaluation Notes. Retrieved from http://interaction.org/impact-evaluation-notes.

Rossi, P., Freeman, H., & Lipsey, M. (2004). Monitoring program process and performance. In *Evaluation: A systematic approach*.

Shukla, A., Khanna, R., & Jadhav, N. (2014). Using community-based evidence for decentralized health planning: insights from Maharashtra, India. *Health Policy and Planning*, czu099. https://doi.org/10.1093/heapol/czu099.

Smith, M. J., Clarke, R. V., & Pease, K. (2002). Anticipatory benefits in crime prevention. In N. Tilley (Ed.), *Analysis for crime prevention: Crime prevention studies* (Vol. 13, pp. 71–88). Monsey: Criminal Justice Press.

Sridhar, D. (2008). *The battle against hunger: Choice, circumstance, and the World Bank*. Oxford: Oxford University Press.

Sridhar, D. (2010). Addressing undernutrition in India: Do 'rational' approaches work? In H. Margetts & C. Hood (Eds.), *Paradoxes of modernization: Unintended consequences of public policy reform* (pp. 119–137). New York: Oxford University Press.

St. Pierre, E. A. (2006). Scientifically based research in education: Epistemology and ethics. *Adult Education Quarterly, 56*(4), 239–266.

UN Women. (2015). *How to manage gender-responsive evaluation*, Independent Evaluation Office. Retrieved from: http://genderevaluation.unwomen.org/en/evaluation-handbook

UNICEF. (2011). *How to design and manage Equity-focused evaluations* (Michael Bamberger and Marco Segone).

van Hollen, C. (2003). *Birth on the threshold: Childbirth and modernity in South India*. University of California Press, 272 pp.

Watkins, R., Meiers, M. W., & Visser, Y. L. (2012). Needs assessment: Frequently asked questions. In *A guide to assessing needs: Tools for collecting information, making decisions, and achieving development results*.

White, H., & Masset, E. (2007). Assessing interventions to improve child nutrition: A theory-based impact evaluation of the Bangladesh integrated nutrition project. *Journal of International Development, 19*(5), 627–652.

Woodward, J. (2017). Scientific explanation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2017 edn). https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation/

Woolcock, M. (2013). Using case studies to explore the external validity of 'complex' development interventions. *Evaluation, 19*(3), 229–248.

World Bank. (1994). *Impact evaluation report. India. Tamil Nadu integrated nutrition project*. Washington: World Bank, Operations Evaluation Department. Internal report. Processed.

World Bank. (2005, December). *The Bangladesh Integrated Nutrition Project: Effectiveness and lessons*. Bangladesh Development Series – paper no.8.

Ziliak, S., & Teather-Posadas, E. R. (2016). The unprincipled randomization principle in economics and medicine. In G. DeMartino & D. McCloskey (Eds.), *The Oxford handbook of professional economic ethics*. New York: Oxford University Press.

Zimbardo, P. G. (1973). On the ethics of intervention in human psychological research: With special reference to the Stanford prison experiment. *Cognition, 2*(2), 243–256.