

# **Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement**

## **Note on Approach and Methods**

**2 October 2015**

**Mokoro Limited**

*in partnership with*



Mokoro Ltd, The Old Music Hall, 106-108  
Cowley Road, Oxford, OX4 1JE, UK  
Tel: +44 (0) 1865 403179  
Email: [mokoro@mokoro.co.uk](mailto:mokoro@mokoro.co.uk)  
Website: [www.mokoro.co.uk](http://www.mokoro.co.uk)

**Valid International**



**FEG Consulting**





### **Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement**

The evaluation was undertaken by Mokoro Ltd in partnership with Food Economy Group (FEG) Consulting and Valid International. The team of evaluators and researchers comprised Stephen Lister (team leader), Stephanie Allan, Zoe Driscoll, Alta Fölscher, Christopher Leather, Lilli Loveday, Mirella Mokbel Genequand, Robrecht Renard, Stephen Turner, and Muriel Visser.

Additional support was provided by Philip Lister, and assistance with country case studies was provided by Liv Bjørnstad (Senegal), Adriana Moreno (Guatemala), Ray Purcell (Bangladesh), Gregory Rooney (Indonesia), and Chris Tanner (Mozambique).

The internal quality assurance panel comprised Stephen Anderson and Alistair Hallam.

The Independent Comprehensive Evaluation was commissioned by the SUN Movement's Lead Group and was overseen by its Visioning Sub-Group (VSG). Three quality assurance advisors – Julia Compton, Chris Gerrard and Eva Lithman – were appointed to assist the VSG. The evaluation manager was Ruwan de Mel.

Funding for the Independent Comprehensive Evaluation was provided by the Bill & Melinda Gates Foundation.

This paper should be cited as:

Mokoro 2015, *Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement: Note on Approach and Methods*. Oxford: Mokoro Ltd, 2 October 2015.

## **Disclaimer**

This Note on Approach and Methods was prepared by the Evaluation Team as a reflection on their experience of conducting the Independent Comprehensive Evaluation of the SUN movement. The opinions and the recommendations offered should not be ascribed to any of the other parties mentioned in this note.

The draft of this note was shared with the Quality Assurance Advisers. They communicated that they did not wish to provide any detailed feedback, but that this should not be construed as their agreement with the views expressed. While not wishing to comment in detail, they did however note (in relation to the discussion of the evaluation process in Section 5) that they do not agree with the contention that the QAA should have been more engaged rather than coming in after the fact. Consequently they see their positioning in the TOR as appropriate, giving them the necessary independence they require.

The QAA also suggested that the short turn-around times for both the Evaluation Team and the Quality Assurance Advisors should be emphasised. Table 3 in Section 5 now does this.

## **Contents**

Summary of Lessons and Recommendations .....	iv
1. Introduction .....	1
2. Overview and known challenges .....	2
3. Issues in ICE scope and subject-matter .....	5
4. Reflections on evaluation instruments .....	5
5. Reflections on the evaluation process .....	31
6. Recommendations .....	38
Bibliography .....	40
Abbreviations .....	44

## **Boxes**

Box 1	QAA assessment of the final report.....	4
Box 2	SUN ICE objectives and scope.....	5
Box 3	Uses and users of the evaluation.....	6
Box 4	The Concept of a Movement .....	8
Box 5	ToC investigation of the SUN concept of a movement .....	1
Box 6	ICE recommendation on the DBM .....	4
Box 7	ICE summary verdicts on ToC assumptions.....	11
Box 8	Role of EQs and Evaluation Matrix: QAA comments and ET response .....	16
Box 9	Use of interviews: QAA comments and ET response .....	19
Box 10	QAA reflections on ICE M&E findings .....	22
Box 11	Selection criteria for case study countries .....	24
Box 12	Statistical Methods used in the e-Survey Analysis .....	30
Box 13	Evaluation Manager comments on time and budget.....	32
Box 14	Comments on first draft ER – interchange between QAA and ET .....	37

## **Tables**

Table 1	Main Evaluation Questions.....	13
Table 2	Implementation of Country Case Studies .....	25
Table 3	Schedule for ICE Report Finalisation.....	31

## **Figures**

Figure 1	SUN ICE Global Theory of Change: Foundational Diagram .....	8
Figure 2	SUN ICE detailed Global Theory of Change .....	9

## Summary of Lessons and Recommendations

### Background

S1. This note provides the evaluation team's reflections on the Independent Comprehensive Evaluation (ICE) of the Scaling Up Nutrition (SUN) movement. The evaluation commenced in mid-June 2014 and its final report was circulated in mid-January 2015. The evaluation was intended to support a thorough review and planning process for SUN.

S2. Independent quality assurance advisers (QAA) judged that "the evaluation was generally of good quality, objectives-based and evidence based — presenting credible links between the findings, conclusions and recommendations." (Compton et al 2015)

S3. Nevertheless, it was evident from the outset that the timetable for the evaluation was very compressed, and its resources were limited. Accordingly this note focuses on efforts at mitigation, and its main lessons are about how to avoid unnecessary handicaps for future evaluations. Lessons and recommendations are particularly oriented towards future evaluations of SUN.

### Lessons

#### *On scope and approach to evaluating the SUN movement*

S4. SUN's boundaries are loosely defined, and the ICE adopted a broad approach to what counts as part of SUN. A similarly broad approach is likely to be appropriate for future evaluations. However, this has implications for any evaluation's scope of work, and should be explicitly addressed at terms of reference (TOR) stage as well as in the evaluation itself.

S5. SUN is insistent on its status as a movement, but it is not self-evident what this means. The ICE carefully investigated SUN's origins and how the idea of "a movement" has played out in practice. Future evaluations should also take SUN's status as "a movement" as an empirical issue, not an axiomatic one. It is always relevant to consider (a) whether, and to what extent, a partnership conforms to its own self-description, and (b) the practical effects of its basic assumptions.

S6. Some efforts to restrict the evaluation's scope were helpful, but others were futile. Restricting the scope of an evaluation can make the evaluation team's workload more manageable, but there may be trade-offs (e.g. in excluding significant issues — such as the experience of non-SUN countries, or in terms of the quality and timing of the secondary data sought). In particular:

(a) Multiple burdens are here to stay, and will be an unavoidable issue for future SUN evaluations.

(b) TOR for future SUN evaluations will need to consider carefully the scope of work that is envisaged on SUN's place in the international architecture. It

would make sense to commission a review of this issue in advance of the main evaluation of SUN's next phase.

S7. It is impossible to anticipate everything a complex evaluation will encounter. In this case, for example, the importance of reviewing the previous history of multi-sectoral planning for nutrition became apparent only after the inception stage. More time spent on the TOR (or the inception report) will not necessarily prevent surprises, and may well eat into the time needed for the evaluation itself. There must be reasonable scope for adapting an evaluation's work programme, even after the inception report stage.

#### *Theory of change*

S8. Reconstructing SUN's theory of change was a strong foundation for the evaluation. The theory of change developed for the ICE should be useful as a starting point both in formulating SUN's future strategy and in its subsequent evaluation.

#### *Evaluation questions*

S9. ICE experience suggests: (a) that drafters of terms of reference should not allow the number of evaluation questions to proliferate; (b) however, the scope of an evaluation's work is determined more by its overall purpose than by the number of subsidiary questions posed; (c) the evaluation matrix should be recognised as a pragmatic tool: it is unlikely that any complex evaluation will collect complete answers to every sub-question – the aim is to collect sufficient evidence to make credible judgments on the major issues.

#### *Interviews*

S10. Interviews are key to a comprehensive evaluation, and it is important to ensure that interview evidence is systematically recorded and shared across the evaluation team. The quality of interview evidence should be assessed directly by quality assurance advisers.

#### *Secondary data*

S11. Reliance on secondary data made sense, and the data collected for the Global Nutrition Report (GNR) were extremely valuable. However the ICE was locked into the timetable for publication of the GNR and of SUN's annual report, which further compressed the time available for data analysis and report writing.

S12. SUN's M&E system, in the form reviewed by the ICE, does not reliably track performance either within or across countries. Both the performance of SUN 2.0 and its eventual evaluability will be seriously prejudiced if the ICE recommendations on M&E are not effectively taken forward.

### *Country case studies*

S13. The country case studies were crucial for a country-focused movement. The ICE had to adopt second-best approaches to fit the case studies into the time and resources available. The next ICE should incorporate in-depth country case studies, conducted over months rather than weeks, and arranged to precede the main phase of the evaluation, with published CCS reports feeding into the main phase of the evaluation. They should be linked to a more systematic comparison of SUN and non-SUN countries.

### *E-survey*

S14. The e-survey was a valuable component of the ICE methodology. Conducting it towards the end of the investigative process enabled it to focus on the most relevant emerging issues. Liberal use of text boxes for open-ended responses provided rich material but added to the time required for analysis and write-up.

### *Evaluation process*

S15. The compressed timetable for the ICE had foreseeable consequences in limiting the evaluation team's consultations with stakeholders around its conclusions and recommendations. An unforeseen consequence was to lower the quality of the post-report debate, because the evaluation team was not given the opportunity to defend the report's conclusions and recommendations.

S16. The arrangements for contracting and independent quality assurance were crucial in safeguarding the independence of the evaluation.

S17. The quality assurance arrangements also contributed to strengthening the evaluation, but would have contributed more if the QAA had relied less on ex post review of draft reports and engaged more fully with the evaluation team throughout the process.

## **Recommendations**

S18. We expect that the next phase of SUN (SUN 2.0) will also require an independent comprehensive evaluation. Some of the design criteria for ICE2 should be:

- (a) The contract for ICE2 should be let not later than 2 years before the end of the SUN 2.0 period. If necessary there should be scope for modification of the overall TOR and the TOR for component studies by negotiation after the contract is let.
- (b) The evaluation plan for SUN 2.0 should provide for component studies to be completed ahead of the main overall evaluation. These would certainly include full country case studies (to involve deeper studies than were feasible for ICE1, complemented by systematic consideration of the lessons for SUN that may be drawn from non-SUN countries), plus other thematic studies (e.g. of nutrition architecture and lessons from other partnerships, gender, the double burden

of malnutrition) that would provide a stronger platform for the main evaluation.

- (c) It should strengthen the use of QAA by (a) engaging them earlier (while taking care that their involvement does not serve to elongate the pre-contracting process); (b) making it clear that the QAA are expected to engage fully with the evaluation team (after it has been competitively selected) and not rely solely on post-hoc review of draft reports in assessing the adequacy of an evaluation.

# 1. Introduction

## *Purpose of NOAM*

1. During the second half of 2015, Mokoro Limited (in collaboration with Valid International and FEG Consulting) undertook the Independent Comprehensive Evaluation (ICE) of the Scaling Up Nutrition movement (SUN). Although it was not a requirement of the Terms of Reference (TOR), Mokoro planned from the outset to prepare a note on approach and methods (NOAM) to harvest lessons that could be useful for future evaluations, including future evaluations of SUN itself.

2. Prior to preparing this note, Mokoro undertook a thorough internal review of the evaluation experience. An experienced Mokoro consultant (Anthea Gordon), who had played no part in the evaluation itself, collected reflections from all team members (see list on inside cover) and from Mokoro's management in writing and/or in conversation. To encourage frankness, confidentiality was maintained. A summary report (internal to Mokoro) was shared with all involved. We also had sight of draft reflections by the Evaluation Manager (de Mel 2015) and the final *Statement by the Evaluation Quality Assurance Advisors* (Compton et al 2015).<sup>1</sup>

3. This note aims to be a commentary, not an encyclopaedia. We do not discuss every aspect of the evaluation but focus on what we see as more important issues (particularly those that will be directly relevant to the next evaluation of SUN). We assume readers who are interested enough will refer to the relevant evaluation documents, which are all available on line.

## *Main Topics*

4. The NOAM has the following main sections:
- a) overview of the evaluation and known challenges;
  - b) issues in the evaluation's scope and subject;
  - c) reflections on evaluation instruments;
  - d) reflections on the evaluation process;
  - e) summary of recommendations.

---

<sup>1</sup> However, the evaluation team (ET) was not given an opportunity to comment on the latter prior to its submission to the VSG.

## **2. Overview and known challenges**

### **Evaluation timetable and outputs**

5. From the outset, the timetable for the ICE was recognised to be very tight. The contract was awarded on 20 June 2014, and in the following six months the evaluation team had to complete three formal deliverables:

1. An Inception Report (IR) (Mokoro 2014b), dated 1 August 2014, which set out a clear methodology (with detailed justification in Annexes) for the evaluation, as well as a work plan for the evaluation process itself.
2. An Interim Progress Report (IPR) (Mokoro 2014c), dated 15 September 2014, which presented some initial observations from the evaluation as had emerged at that point, in addition to an interim assessment of the SUN movement secretariat (SMS).
3. The Final Report,<sup>2</sup> dated 15 January 2015, presenting the evaluation's full findings, conclusions and recommendations (Mokoro 2014e).<sup>3</sup>

6. In addition to these reports, the evaluation team produced a fourth output, in the form of an Options Discussion Paper (Mokoro 2014d). This was published in November (between the IPR and Final Report), as a means of presenting emerging findings on the potential future directions of SUN, drawing on the country case studies that took place in September and October, and informing the ICE-related discussions at the SUN Global Gathering.

7. Other substantial elements of the work (though not separate deliverables) included eight country case studies (CCSs) and an e-survey; these are among the instruments reviewed in Section 4 below.

### **Organisation of the evaluation**

#### *Contracting and management of the evaluation*

8. The TOR for the ICE were prepared with assistance of consultants (Bezanson et al. 2014<sup>4</sup>). The TOR included terms of reference for quality assurance advisors (QAA). Three QAA were appointed, and were involved first in appraising tenders for the evaluation and then in QA of its deliverables. The Bill and Melinda Gates Foundation (BMGF) agreed to finance and contract the evaluation on behalf of SUN partners. The evaluation was openly advertised, with a financial limit of USD 500k for the bids, and a deadline to deliver the final report by end of 2014 (this deadline was linked to the timetable for a planned visioning process to develop SUN's next

---

<sup>2</sup> Referred to in this note as the Evaluation Report (ER).

<sup>3</sup> There were minor amendments, mostly to some annexes, in an edition dated 1 May 2015, but this involved no changes to conclusions or recommendations.

<sup>4</sup> The full TOR were also annexed to both the Inception Report and the Final Report.

phase).<sup>5</sup> The QAA themselves recommended employment of an evaluation manager, who was appointed shortly after the contract was awarded to Mokoro.

#### *Evaluation team and approach*

9. Mokoro assembled a team of experienced evaluators, almost all of whom had previously worked together. Core team members were given thematic areas of focus and also led case studies; a research team at Mokoro HQ provided support for the thematic work and case studies. The allocation of responsibilities to each team member was summarised in the IR (Annex O). Budgeted days for each team member (also shown in IR Annex O) were limited and therefore had to be carefully husbanded; all team members nevertheless worked more days than budgeted. The team were brought together for a three-day workshop in Oxford in July, at which the principal evaluation instruments, including the Theory of Change, were developed. Subsequent team meetings were by conference call. Team dynamics are further discussed in Section 5 below.

#### **Known challenges and approach to mitigation**

10. The objectives and scope for the ICE were set very broadly (more on this in Section 3 below); this was inherent in the role the evaluation was intended to play in the summative review and forward planning of SUN, and was made explicit in the TOR references to a comprehensive evaluation approach.

11. Comprehensive evaluations are known to be ambitious and both methodologically and practically difficult (see, for example, Isenman 2012b, IEG 2007). In this case there were particular challenges in a constrained budget and, especially, a constrained timetable. These challenges were evident from the outset, so the main theme of this note is about efforts at mitigation, and the main lessons are about how to avoid unnecessary handicaps for future evaluations.

12. Among the mitigations envisaged in the TOR were maximum reliance on secondary material, including M&E undertaken by the SUN movement secretariat (SMS), but the TOR also made clear that fieldwork in selected case study countries was regarded as crucial, as was the process of review and discussion with stakeholders that is characteristic of comprehensive evaluations.

13. At the same time, there was pressure for interim findings relatively early in the evaluation process, so as to feed into a scheduled Lead Group Meeting and the SUN Global Gathering in November 2014. Interim findings on the SMS were also required so as to meet the review schedules of the aid agencies funding the SMS (SMS 2014a). This added significantly to the pressure on the team, since such findings, in order to be credible, required a great deal of work at an accelerated pace.

---

<sup>5</sup> Hence, within SUN the ICE process was overseen by a Visioning Sub-Group (VSG).

14. The mitigations planned by Mokoro included;
- preparatory work at proposal stage (e.g. Mokoro's technical proposal had already assembled much country data and developed preliminary criteria for CCS selection);
  - squeezing as much as possible from the limited time available to senior evaluators by:
    - providing strong research support;
    - trying to balance subject-wise division of labour with pooling of team assessments;
    - giving the core evaluators primary responsibility for the case studies (and where possible assigning them to countries with which they were already familiar);
  - drawing on the data sets being compiled for the inaugural Global Nutrition Report (GNR);
  - limiting the scope of the ICE to the extent possible (but see Section 3 below).
15. Methodology is discussed in 4 below. The challenge throughout was to make the evaluation as thorough and as credible as possible despite the complexity of the subject-matter and the constraints on the budget and the timetable.
16. In the judgement of the independent QAA, the evaluation team succeeded in delivering a credible and useful report (see Box 1 below). Nevertheless there are important lessons on how to achieve better results with less stress in future.

**Box 1 QAA assessment of the final report**

"In the judgement of the QAA, the evaluation was generally of good quality, objectives-based and evidence based — presenting credible links between the findings, conclusions and recommendations." (Compton et al 2015)

### **3. Issues in ICE scope and subject-matter**

#### **Scope and purpose as defined by the TOR**

17. Box 2 and Box 3 below demonstrate the TOR's high ambitions for the ICE and the breadth of its scope. The TOR were based on extensive discussions with stakeholders and included an Annex ("Indicative Listing of Issues/Questions to be Addressed in the Independent Comprehensive Evaluation of the SUN Movement") with over 70 questions for the ICE to address. Nonetheless, they also included a requirement that "all deliverables will be as concise as possible" (TOR ¶37).

#### **Box 2 SUN ICE objectives and scope**

The central objective of the Independent Comprehensive Evaluation (ICE) is to inform the Lead Group's Visioning Exercise on the future of the SUN movement. To do this, the evaluation is expected to analyse past and present processes and activities of the movement, and present findings, conclusions and targeted recommendations that would allow the Lead Group and all stakeholders to chart the way forward. In doing so, the ICE should help to strengthen the sense of unity among stakeholders and make SUN fit for purpose, and prepared for the challenges ahead.

In terms of scope, the ICE is to consider all aspects of SUN – its institutional structure, objectives, working model(s), decision processes, role within the wider architecture of international development, relevance, value-added, efficiency and effectiveness. It will address how effective SUN has been in carrying out its objectives – concerned with accelerating the reduction of undernutrition – and to pose options for evolution of the SUN movement to build on strengths and address weaknesses. It will provide an independent assessment of what SUN has accomplished and is accomplishing, the efficiency and effectiveness of its different components (its governance, networks and secretariat), its current functioning and to the extent feasible, its contribution at country, regional and global levels. It will examine the extent to which SUN is helping national governments, and other stakeholders, to contribute to transformations in the way nutrition is being addressed. And it will assess the role of SUN in increasing attention to women's empowerment and gender equality and in catalyzing nutrition-sensitive approaches in agriculture, health care, water and sanitation and other sectors.

*Source: adapted from TOR, ¶7, 13, 14*

### **Box 3    Uses and users of the evaluation**

In analysing past and present processes and activities, therefore, the evaluation is expected to present findings, conclusions and targeted recommendations that would allow the Lead Group and all stakeholders to chart the way forward for the SUN Movement. Consequently, the evaluation should be regarded as a milestone for SUN and nutrition, reinforcing SUN's potential to meet the overarching purposes for which it was established. That purpose entails helping the SUN countries themselves – which are at the centre of the SUN movement – to accelerate and maximize progress toward eliminating the scourge of malnutrition. The ICE should help to strengthen the sense of unity among stakeholders to achieve that purpose and to help make SUN fit for the challenges ahead.

*Source:* TOR ¶14

**18.    The TOR also highlighted the unusual character of SUN:**

SUN, however, is not a new institution or financial mechanism. It is a very broad multi-stakeholder partnership to support national plans to scale up nutrition. It is a voluntary movement that has no legal charter or legal status. It does not directly furnish financial or technical resources, but seeks to catalyze their availability in response to country needs. SUN is open to all countries whose governments commit themselves to scaling up nutrition and to all stakeholders committed to providing support. (TOR ¶4)

**19.    In this section, we discuss several issues ICE had to grapple with arising from the nature of SUN and the breadth of the TOR, namely:**

- defining what the evaluation should consider as part of SUN;
- dealing with the concept of SUN as "a movement";
- limiting the scope of the evaluation (but addressing unanticipated topics).

### **Defining the subject of the evaluation**

**20.    As noted in the ER:**

The nature of the SUN movement posed its own challenges, in terms of identifying clearly what are the appropriate results against which to assess its performance. (ER¶1.11)

The ICE has taken the view that it would be unrealistic, and unhelpful for the formative element of the evaluation, to take a very narrow view of what counts as part of SUN. There are a number of initiatives and interventions which do not formally come under the direction of SUN bodies (the LG, the SMS, the support networks) but which do come under the auspices of principal collaborators in SUN and contribute, more or less explicitly, to the same objectives.

**21.    The ER gave the examples of MQSUN,<sup>6</sup> UN REACH<sup>7</sup> and the Nutrition For Growth (N4G) Summit,<sup>8</sup> all of which the ICE counted – with appropriate caveats where necessary – as broadly part of SUN activities (|ER ¶2.28 and Box 4).**

---

<sup>6</sup> The Maximising the Quality of Scaling up Nutrition Framework (MQSUN) is a DFID-funded consortium of research and technical institutions, which provides technical services to improve the quality of DFID's nutrition

22. A related issue came up in QAA review of the ICE draft recommendations; the QAA queried drafts of Recommendation 10 ("The partners in SUN should pursue further rationalisation of the international architecture for nutrition.....") on the basis that it might be beyond the scope of the SUN movement as such. We felt this was an important issue that the ICE could not sensibly ignore, while making clear it was primarily addressed to the international aid partners in SUN, UN agencies and multilateral and bilateral aid partners who simultaneously participate in SUN and drive many other international initiatives.

23. **Lessons:** *a similarly broad approach to what counts as part of SUN is likely to be appropriate for future evaluations. However, this has implications for any evaluation's scope of work, and should be explicitly addressed at TOR stage as well as in the evaluation itself.*

### Dealing with SUN as a "movement"

24. SUN is very insistent on describing itself as a movement. However, it is not self-evident what this entails, and the ICE therefore considered what this terminology has meant in practice. ICE conclusions were:

We [have noted] that SUN's character as a loose and voluntary partnership means that it does not have sharp boundaries. The terminology of SUN as a movement that is country-driven is used as a way of emphasising that it seeks to avoid imposing top-down solutions on countries that suffer from undernutrition and instead to rally support behind government strategies and priorities. However, SUN is not a movement in the conventional sense reviewed in [Box 4 below], and historically it was not initiated by the governments of the countries on which it focuses. As Nisbett et al 2014a put it: "Emerging from the UN and donor network, SUN has worked hard to be seen as country led and to shift responsibility for its actions to its growing number of country leads". **The terminology of a country-driven movement is best seen as a powerful metaphor, and a simplified perspective on the complex dynamics of how SUN operates in practice** (which is analysed further in the discussion of the networks and overall governance ... . (ER¶3.10, emphasis added)

---

programmes, and can be considered part of the portfolio of service providers which the SUN movement can call on to respond to country requests for support. (MQSUN 2014) Annex M of the ER included a detailed review of the interplay between SUN and MQSUN around plans and costing.

<sup>7</sup> Renewed Efforts Against Child Hunger and Undernutrition (REACH) is a consortium of five UN agencies (FAO, IFAD, WHO, UNICEF and WFP) operating at the country level to build national capacity for multi-sectoral nutrition governance. It coordinates the UN's work in nutrition in the areas of advocacy and consensus building, support for development of national policies and plans, capacity building, and M&E (REACH 2013).

<sup>8</sup> The N4G Summit was held in June 2013, co-hosted by UK government, the Children's Investment Fund Foundation (CIFF), and the Government of Brazil. It resulted in the endorsement of a compact that included donor commitments of up to GBP 2.7 billion (USD 4.15 billion) to directly tackle undernutrition, and an estimated GBP 12.5 billion (USD 19 billion) for nutrition-sensitive investments, for the period 2013 – 2020 (N4G 2013a). The event was fairly deliberately not held under the banner of the SUN Movement – indeed the commitments emerged from a wider group of countries than SUN members, – but SUN was still actively involved in the run-up to the event, by supporting the engagement of member countries, lobbying them on their commitments, and helping with costing of plans through MQSUN.

#### **Box 4    The Concept of a Movement**

The growing frequency of social movements since the 1960s means that their study has become a well-established field in the social sciences, particularly within sociology. Della Porta & Diani 2006 define social movements as “distinct social process, consisting of the mechanisms through which actors engaged in collective action are involved in conflictual relations with clearly identified opponents, are linked by dense informal networks, [and] share a collective identity” (p.20).

With regards to being “involved in conflictual relations with clearly identified opponents”, in other movements, such as the American civil rights movement (1954–68), the opponents are clear (federal, state, and local governments) and the conflictual relations well-defined (centring on the legally-enforced segregation and discrimination against black Americans). Similarly so for the UK’s jubilee debt relief movement, where the opponents are the creditors of poor countries debt, and the conflictual relations concern the cancellation of those debts. For the SUN movement, the idea of the opponent is more ambiguous: those responsible for tackling (or failing to tackle) malnutrition are the members of the movement – government, development partners, the private sector and civil society, and as such the “conflictual relations” would be with members of the movement itself.

Relatedly, what would be counted as success – in terms of a favourable outcome to the conflictual relations – whilst being relatively straightforward in the case of the American civil rights movement or the jubilee debt campaign (legislative reform and debt relief respectively), is markedly less so in the case of SUN, where addressing malnutrition is long term endeavour dependent on a multitude of factors, and which requires constant monitoring and follow-up (and the structures to support that).

In terms of a movement’s members being linked by “dense informal networks”, the structure of multi-stakeholder networks at the global and country level is the *modus operandi* of SUN, but the extent to which these can be considered informal is questionable.

With regards to collective identity, Della Porta and Diani assert that “a social movement process is in place only when collective identities develop, which go beyond specific events and initiatives” (Della Porta & Diani 2006). They argue that this is highly dependent on mutual recognition and a sense of connectedness between actors, to define who is and who is not part of the network. In terms of the SUN Movement, this sense of connectedness is clearly bolstered by efforts such as the annual Global Gatherings,<sup>9</sup> but it is less clear to what extent this collective identity is recognised by actors external to the movement or trickles down to country level, where networks may not be “branded” as SUN.

*Source:* appeared as Box 3 in the final ER.

---

<sup>9</sup> As witnessed by members of the ICE team attending the 2014 Global Gathering.

25. The utility of the "movement" approach was included as an underlying assumption of the theory of change, with the findings summarised in Box 5 below.

**Box 5 ToC investigation of the SUN concept of a movement**

**Assumption 10: The SUN concept of a Movement will overcome institutional rivalries, inertia and competition, at global and country level.**

*Issues*

Institutional rivalries, inertia and competition are ubiquitous, so it might have been better to phrase this assumption in terms of mitigating rather than overcoming them. As discussed in the main report (see section 3.2, Box 3) SUN is not historically a movement in the usual sense, but the terminology has served to differentiate it from more formal global partnerships, emphasising the stress it places on country ownership, and also implying that it (despite the existence of a funded secretariat) is not an agency along the lines of (or in competition with) the UN agencies that form the UN network (although a significant part of the motivation for launching SUN was to compensate for the perceived collective failings of the UN system).

*Assessment*

By operating as a rather loose partnership, and styling itself as a movement, SUN has not abolished the rivalries and competition from which it emerged (see for example discussion of the UN System Network in Annex J) but it was able to bypass them to a useful degree and thus overcome inertia. We therefore consider this assumption adequately validated.

*Source:* ER, Annex S, ¶28–29.

26. The ICE also noted that deference to the concept of a movement could have perverse effects:

... the language of a "movement" is a powerful metaphor (chapter 3, section 3.2). It has been valuable in protecting the principles that SUN should be inclusive and country-driven in its approach. It has helped SUN to avoid the top-down prescriptive approaches of some other global partnerships. However, while it is right to respect countries' own determination of priorities, SUN has not been sufficiently specific about the criteria costed plans and CRFs must fulfil if they are to be useful instruments of accountability at country level – where the challenge is to hold all stakeholders, not just the government, accountable for their commitments towards tackling undernutrition (sections 4.5, 5.4 and 5.5) ( ER ¶7.13)

27. ***Lessons:*** Future evaluations should also take SUN's status as "a movement" as an empirical issue, not an axiomatic one. It is always relevant to consider (a) whether, and to what extent, a partnership conforms to its own self-description, and (b) the practical effects that may arise from its basic assumptions.

**Efforts to limit the practical scope of work**

28. At the TOR stage and subsequently, there were efforts to moderate the evaluation's demands, recognising the constraints of time and resources the evaluation team would face, but there was also unanticipated work that pulled in the opposite direction.

*Mitigations built into the TOR*

29. The main mitigations built into the TOR were:

- a) Recognition that the ICE could not include a full evaluation of the Multi-Partner Trust Fund (MPTF). Although the ICE work on the MPTF was not negligible (see ER Annex K), this was a restriction of scope that it was practical to observe.
- b) Proposed reliance on secondary data, in particular the data sets collated by the SMS as part of SUN's M&E systems. Again, this was a useful mitigation, inasmuch as it avoiding duplication by the ET of work that the SMS would do anyway. And similarly, the cooperation of the GNR team in making available their draft data sets (and emerging drafts of the GNR itself) was very valuable. However, in both cases (for the GNR and the SUN annual report prepared by the SMS) completion was scheduled to feed into the Global Gathering and subsequent International Conference on Nutrition (ICN2) in mid-November. This meant that key data were not available until very close to the deadlines for delivery of the ICE draft final report. This timetabling issue is further discussed in Section 5 below. There is further discussion of the use and quality of secondary data in ¶74ff below.
- c) Restriction of case studies to SUN countries. Again, this was clearly a necessary time-saver, but it excluded important issues that might be explored by considering the performance of non-SUN countries and their reasons for remaining outside SUN. A similar exclusion in the next major evaluation of SUN would be hard to justify.

30. **Lessons:** *Restricting the scope of an evaluation can make the evaluation team's workload more manageable, but there may be trade-offs (e.g. in excluding significant issues, or in terms of the quality and timing of the secondary data sought).*

*The double burden of malnutrition*

31. In their comments on the draft IR, the QAA advised that the ET should limit its review of issues related to overnutrition and the associated "double burden of malnutrition" (DBM):

In some cases it seems unrealistic to do much in this evaluation –this particularly applies to [the sub-question] about whether SUN should address obesity and dual burden issues, which is an important question, but needs a study in itself – it could have major implications for choice of study countries and choice of stakeholders to interview (and moreover is a question not specifically required in the evaluation ToR). For example, the evaluation country case studies can document obesity and dual burden problems and interests, and the evaluation can summarise key points which would need attention in a focused study of this area, but the evaluation cannot realistically do a full study. Second, it would be better to manage stakeholder expectations down to a realistic level at the inception phase of the evaluation – when there is still time for the VSG to commission additional studies of particular areas if required. (from QAA comments on draft IR)

32. In fact, the TOR did specifically raise this issue:

The ICE will need to take into account the rapid changes occurring in the landscape of international development and new realities and challenges in nutrition. Overnutrition, obesity and their associated non-communicable diseases are now widespread and increasing so rapidly that the World Health Organization refers to this phenomenon as a new pandemic. Moreover, obesity is growing in all developing regions, even in countries beset by high levels of poverty where increasingly there is a double burden on the healthcare system from under-nutrition and obesity. The focus of SUN thus far has been almost exclusively on the challenges of under-nutrition. A central question for the future will be whether the next stage in SUN's evolution should include a broader nutrition objective that would also specifically address overnutrition. (TOR ¶11)

Should SUN broaden its overall objective of accelerating reduction in undernutrition to include reduction in overnutrition, with its consequences for Non-Communicable Diseases, as well? (TOR Annex E)

33. Nevertheless, the IR noted the following limitation:

The TOR ask whether reduction in overnutrition should be added to SUN's existing objectives. We will note this as a strategic option for the future, and we will examine whether and how obesity is being addressed in the policies and strategies of our case study countries. But it is beyond the scope of this evaluation to undertake a full and thorough review of this topic. (IR ¶8.2a)

34. In practice, it quickly became apparent that the DBM is an inescapable issue for SUN, and not one that could be partitioned off to a separate study and left out of ICE considerations.<sup>10</sup> The ER noted:

over SUN's lifetime it has become increasingly apparent that problems of undernutrition and overnutrition are interrelated, both in physiological and other dimensions, and that national nutrition policies and strategies need to take account of the so-called double burden of malnutrition ... ); too narrow a focus on stunting should be avoided.

35. SUN's future approach to DBM became the subject of a main recommendation (see Box 6 below). Although we would not claim to have undertaken "a full and through review of this topic", the idea that the ICE workload could be reduced by marginalising this topic proved illusory.

36. ***Lessons:*** *Multiple burdens are here to stay, and will be an unavoidable issue for future SUN evaluations.*

---

<sup>10</sup> Multiple burdens of nutrition as a "new normal" were a principal message of the GNR (IFPRI 2014).

**Box 6 ICE recommendation on the DBM**

*Recommendation 3: SUN's primary focus should continue to be on undernutrition and the elimination of stunting. However, there should be more acknowledgement that good nutrition plans will need to address all forms of malnutrition, including the double burden of undernutrition and overnutrition.*

SUN's advocacy efforts around stunting have been found to be highly relevant and effective, and with global stunting levels still unacceptably high, will continue to be relevant. At the same time, it must be recognised that most SUN countries already suffer from multiple burdens, that this trend will continue, and that national nutrition strategies must address all dimensions of malnutrition. SUN 2.0 should continue to focus on countries where undernutrition is a significant burden, while recognising that stunting must be tackled in tandem with other dimensions of malnutrition. (ER ¶S34)

*International architecture for nutrition*

37. The international architecture for nutrition was bound to feature prominently in ICE, not least because the Lancet's forthright criticism of "fragmented and dysfunctional" global institutional architecture for nutrition (The Lancet 2008) was part of the motivation for SUN. One of the questions posed in the TOR was "Are SUN's mandate and role appropriate, in relation to the numerous international organizations and global partnerships involved in closely related areas (e.g. food security and maternal and child health)?"

38. The ET undertook to map this architecture and SUN's place within it. This turned out to be a much bigger task than anticipated, as reflected in the ER's Annex H. Even though we confined ourselves to nutrition and food security initiatives, beginning with the first International Conference on Nutrition (ICN) in 1992, we found more than 20 relevant initiatives, and clarity around SUN's precise place in the architecture remained elusive, as illustrated by the following interchange with the QAA:<sup>11</sup>

Comment: It would be very helpful to the reader to have a diagram and summary of where SUN fits in the international architecture (the reader is referred to a very long descriptive annex of chronology other initiatives). (QAA comment on ER draft v0)

Response: We did not come across such a diagram or summary, which itself says something about how clear the system is. It has probably not been in SUN's interest to prepare anything so clear because it would reopen debates about responsibilities and mandates which SUN, with some success, has tried to bypass. The point of the inventory in Annex H was simply to demonstrate the profusion of initiatives.... (ET response to QAA comment)

---

<sup>11</sup> In addition much of the review of the UN network was concerned with the respective roles and responsibilities of UN agencies.

39. **Lessons:** *TOR for future SUN evaluations will need to consider carefully the scope of work that is envisaged on SUN's place in the international architecture. It would make sense to commission a review of this issue in advance of the main evaluation of SUN's next phase.*

#### *Previous history of multi-sector nutrition planning*

40. There was an even more significant example of unanticipated work. Neither the TOR nor the Inception Report made reference to previous experiences of multi-sector nutrition planning., but the IPR noted:

The assessment to date has underscored the need to address in more detail and more systematically the history and challenges of multi-sector nutrition planning at the country level. The IPR highlights that in reviewing the quality and practicality of SUN country nutrition plans, their integration with other in-country policy frameworks and the challenges of monitoring plans' financing, implementation, and results will need to be considered. The CCSs will need to compare country plans with same-country precursors, in order to understand whether and how what is happening at country-level under SUN auspices differs from unsuccessful efforts in the past. (IPR ¶S12)

41. This proved to be an extremely important strand of work: the ICE assessment of SUN hinges on whether SUN is repeating the mistakes that undermined multi-sector nutrition planning in the 1970s, 1980s and 1990s (see ER ¶3.12–3.17, and the extended treatment of the issue in ER Annex I).

#### *Implications for future evaluations*

42. **Lessons:** *The final section of this note draws implications for the scope and planning of future SUN evaluations, taking account also of lessons from the ICE's use of various evaluation instruments, as discussed next. A general lesson is that it is impossible to anticipate everything a complex evaluation will encounter. More time spent on the TOR will not necessarily prevent surprises, and may well eat into the time needed for the evaluation itself. There must be reasonable scope for adapting an evaluation's work programme, even after the Inception Report stage.*

## **4. Reflections on evaluation instruments**

### **General approach**

43. The adoption of a mixed methods approach for the ICE was not controversial.<sup>12</sup> This involved:

- an inferred theory of change (ToC) to underpin the evaluation;
- an evaluation matrix which drew on the ToC as well as the questions posed in the TOR;
- stakeholder interviews and participant observation;

---

<sup>12</sup> The QAA commented: " We consider the mixed methods approach adopted by the evaluation team to be fully appropriate and necessary to capture the complexity of the task." (Compton et al 2015)

- an extensive literature review and analysis of secondary data;
- eight country case studies;
- an electronic survey.

44. All of these components were important in developing the ICE conclusions and recommendations,<sup>13</sup> and we review each of them below.

## Theory of change

### *What was intended*

45. SUN had not developed a substantial theory of change,<sup>14</sup> and the ET devoted considerable effort to the elaboration of an inferred ToC during the Inception Stage. This was a substantial topic at the team's inception workshop. Annex E of the Inception Report included a review of the way ToCs and similar approaches had appeared in SUN contexts and (more elaborately) in nutrition analyses (including the classic UNICEF analysis and its further development in The Lancet series).

46. The IR summarised the rationale for a ToC approach as follows:

Theory-based evaluation is not a new technique, and there are many variations on the approach (Carter 2012, Vogel 2012). However, even if a programme has not adopted an explicit theory of change (ToC), it is increasingly recognised that elaborating its implicit ToC can be a valuable foundation for an evaluation. This is especially true for evaluations of complex enterprises such as SUN. Within any one evaluation different theories of change can nest to evaluate different aspects of an initiative, or evaluate the initiative at different levels.

There are some similarities between a logical framework and a theory of change, but an important distinction is that the latter also sets out *why it is expected that something will cause something else*. It opens up the black box between programmes and observed changes (or lack of change), and makes explicit the underlying assumptions or conditions on which causal chains depend. This is important for policy-relevant or formative evaluation (Clark & Anderson 2004; Carter 2012).

Theories of change consider initiatives in their context, which include the immediate technical environment, but also the social, political and economic context within which the initiative operates. This is useful for evaluating initiatives that operate in many different contexts.

In the case of the SUN Independent Comprehensive Evaluation, all of these reasons apply for using a theory of change approach to the evaluation. (IR, Annex E, ¶1–4)

47. The IR described the intended use of the ToC as follows:

The evaluation team has used the theory of change elements present in the SUN Strategy 2012–2015, the SUN Revised Road Map and the SUN Monitoring and Evaluation Framework, and has also drawn on a review of literature and on interviews with the SUN's originators, to develop a high-level theory of change to guide the evaluation.

---

<sup>13</sup> The QAA reflections imply that the case studies and the e-survey were less important: "The main sources of evidence underlying the conclusions and recommendations appear to have been the interviews, the document and literature reviews, and participation and observation at the various SUN meetings." (Compton et al 2015). We do not agree.

<sup>14</sup> Although the SUN Strategy 2012 to 2015 and Revised Road Map (SMS 2012a and SMS 2012b) contained elements of a theory of change (see Figure 12 in IR Annex E).

This theory of change is intended:

as a high level guide to reflect (and then check) our understanding of the reasoning on which the SUN movement is based; and

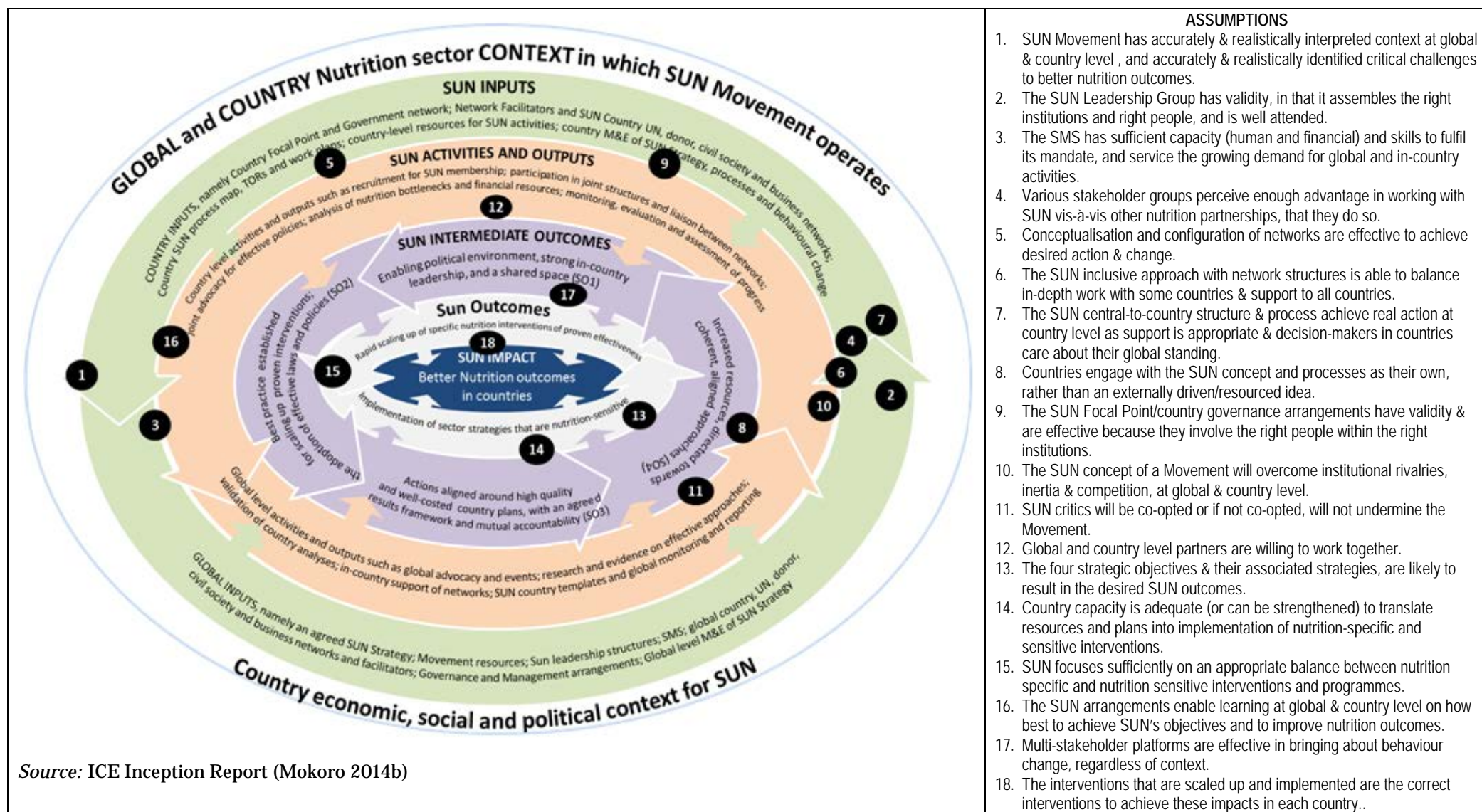
as an evaluation tool to identify and investigate key links in the logic that the theory of change depicts, in terms both of the internal causal/contributory links it proposes and of the key assumptions it sets out. (IR ¶4.7-4.8)

### *What was implemented*

48. The ToC as developed by the ET is shown below. There were two presentations:

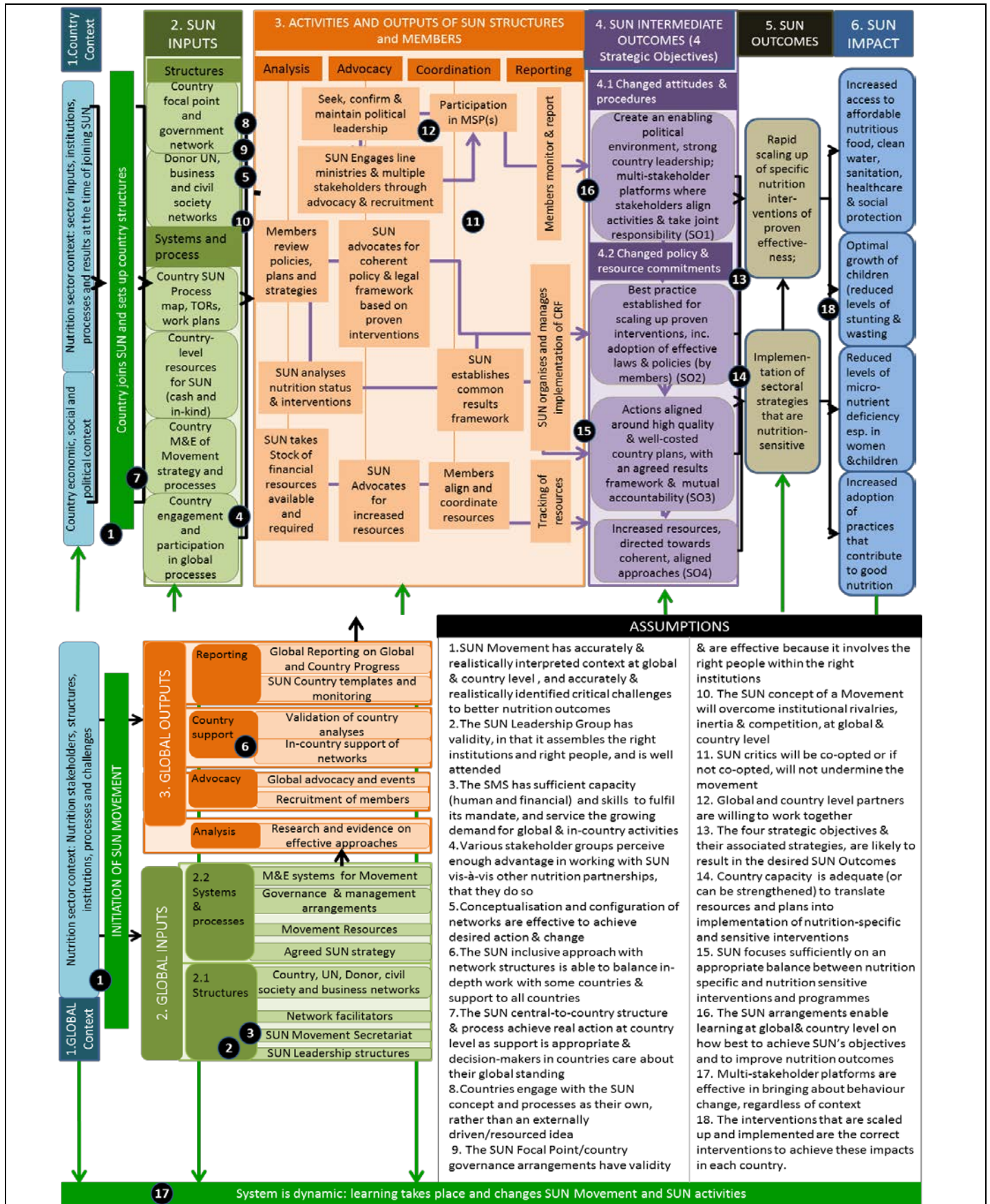
(i) a foundational diagram [Figure 1 below], which provides all the main elements and the assumed causal contributory links between dimensions of the programme in one summary diagram but does not provide detail at the context, input and output levels; and (ii) a detailed theory of change diagram [Figure 2 below] which does provide this detail as well as a more detailed mapping of contributory links. The Foundational diagram should be read from the centre outwards, insofar as a theory of change maps out how an initiative will achieve its aim (or desired final impact), starting from that aim and asking questions such as “what do I need to achieve this aim [outcome, intermediate outcome, output]” working backward. The detailed diagram unpacks this reasoning, and is best read from the bottom up (global level to country level contributory links), and then from left to right, to trace the reasoning from a contributory evaluation perspective. (ER, Annex S)

**Figure 1 SUN ICE Global Theory of Change: Foundational Diagram**



Source: ICE Inception Report (Mokoro 2014b)

**Figure 2 SUN ICE detailed Global Theory of Change**



49. The way the ToC was used is described as follows in the ER:

In assessing the validity of the ToC, the ICE sought to answer two major questions:

Is the ToC an accurate portrayal of the strategic logic SUN has adopted?

Is the strategy the ToC depicts a sound strategy to follow?

In answering the first question, ICE sought the views of stakeholders both directly and indirectly. The views of those most closely involved in the design of the SUN movement and its subsequent operation were sought directly, inasmuch as this group of insiders were closely involved in the preparation and review of the Inception Report, in which the ToC was a central feature. We did not present the ToC diagrams as such to subsequent interviewees (it would not have been a good use of the limited time available for each interview), but our questions, informed by the ToC and the evaluation matrix, were designed, among other things, to check if stakeholders understood the logic of the SUN movement in the same way. None of the main elements in the theory of change presented in the ICE Inception Report were disputed by SUN stakeholders during the course of the ICE, so it may be regarded as a reasonably accurate portrayal of SUN's strategic logic.

To address the second question required the ICE to investigate whether the assumptions that the ToC embodies are reasonable and realistic, and whether the SUN movement is actually achieving the changes it depicts. The key assumptions that were identified as the ToC was formulated are listed in [Figure 1 and Figure 2]. The whole of the ICE report is an investigation of SUN's ToC and how it is working out in practice. Our purpose in [ER Annex S] is not to reiterate all the ICE findings, but to highlight, against each of the assumptions in the inferred ToC, the key issues the assumption raises, and the extent to which the assumption appears to be valid or not. This annex thus also constitutes a comprehensive response to EQ4.8, which asked *Which contextual factors (anticipated or unanticipated) have positively or negatively affected the achievement of SUN objectives?* and expected the ICE "to consider which of the contextual assumptions highlighted in the hypothesised theory of change have proved most significant, positively or negatively". (ER, Annex S, ¶5–7)

50. Annex S was compiled in response to a QAA criticism that the evaluation's ToC underpinnings were insufficiently visible in the initial draft of the IR. It included a review of each ToC assumption, prefaced with the following caveats:

First, the assumptions are broad. With hindsight some of them might have been worded differently, and we do not spend time on nuances that are not central to the main issues. Second, they do not invite simple yes/no answers; whether an assumption holds or not is almost invariably a matter of degree. This leads to a second order question, does the assumption hold sufficiently to allow the causal logic of the ToC to function? From the point of view of a formative evaluation, it is especially important to identify assumptions that are problematic. In principle, a problematic assumption could be fatal, or it could undermine efficiency and effectiveness without necessarily being fatal; in either case it needs to be addressed in taking the programme forward (or deciding not to). Finally, in assessing whether a ToC is sound or not, it is important to distinguish between faulty logic (the ToC could not have worked as depicted) and practical flaws (e.g. the ToC could have worked but was not implemented as designed); in between there are issues of practicality and plausibility (e.g. issues of capacity, which might be seen either as a flaw in the design or an appropriate challenge in implementation). (ER, Annex S, ¶8)

51. The most pertinent findings from the ToC analysis are reproduced in Box 7 below.

**Box 7 ICE summary verdicts on ToC assumptions**

This review of its assumptions has not uncovered any fatal conceptual weaknesses in SUN's theory of change. However, the ICE has identified some serious weaknesses and bottlenecks which will need to be addressed as the SUN strategy is revised and updated. In sequential order these are:

- (Assumption 2) The senior governance of the SUN movement needs urgent attention, and this will involve substantial reconsideration of the LG's composition and *modus operandi*. The assumption of LG effectiveness has not yet undermined the ToC, but is likely to do so if the governance issues raised in this report are not seriously addressed.
- (Assumption 5) The four support networks have to be seen as work-in-progress; the envisaged system for mutual accountability among the networks, which was to be overseen by the LG, is not in place. The ICE sees the networks system as not inappropriate in principle (although the assumption that the country-level networks would precisely replicate the global structure is unnecessarily rigid) but it is still not fully implemented in practice.
- (Assumption 6) The way that SUN's principle of inclusiveness is interpreted could become a handicap but need not be allowed to do so.
- (Assumption 8) In their concern to avoid the imposition of top-down prescriptions, SUN structures have probably been insufficiently prescriptive in some respects (especially in ensuring/monitoring the quality of country plans/CRFs).
- (Assumption 13) If the desirable plans and CRFs are construed in the way the ICE has interpreted recent SMS guidance, then the ToC has a reasonable chance of working. However, this will remain a serious bottleneck unless (a) consistent and much clearer guidance is provided to SUN countries about the characteristics of good quality, functional plans and CRFs that can support the mutual accountability SUN advocates; (b) there is a system of independent feedback on the quality of country plans/CRFs (viewing them not simply as documents but in the context of the systems of consultation and accountability they support); and (c) there is progress with continued work to strengthen resource tracking.
- (Assumption 16) The SUN movement needs to address the weaknesses the ICE has identified in its systems for monitoring and evaluation.

*Source:* ER, Annex S, ¶53.

*Evaluation team reflections*

52. The ToC approach was fundamental. It was given added importance by the absence of a full ToC within SUN's own strategic documents. Focusing on the ToC was an effective way for the ET to come to a common initial understanding of what was being evaluated, and the inferred ToC directly informed the evaluation matrix (see below).

53. There are issues about how visible the ToC needs to be within an evaluation, and within an evaluation report. Given the nature of SUN, the ToC was inevitably complex, and not something interviewees could be expected to come to grips with in the course of an interview. Nevertheless, it was perfectly possible, by asking questions about components of the ToC logic, to check whether it fairly represented stakeholder understanding of how SUN is supposed to work. Annex S of the ER, providing a systematic review of the evaluation's findings against each of the main assumptions incorporated in the ToC, was useful in demonstrating that the ET had addressed each assumption; it made the ET reasoning more explicit, by "showing our working".

54. The ET noted at the outset: "Within any one evaluation different theories of change can nest to evaluate different aspects of an initiative, or evaluate the initiative at different levels." In practice, either for lack of time or because significant subsidiary ToCs did not become apparent, the ET did not systematically investigate such "nested" ToCs.

55. **Lessons:** *The ToC developed for the ICE should be useful as a starting point both in formulating SUN's future strategy and in its subsequent evaluation.*

## **Evaluation matrix**

### *What was intended*

56. As set out in the IR:

The questions to be addressed by the ICE are posed in the TOR, which includes an annex detailing overarching and supporting questions prioritised during consultations about the evaluation. The challenge for evaluators is to structure their enquiry so as to follow a clear sequence from findings to conclusions and recommendations, taking account of the availability of relevant evidence.

For this purpose, the evaluation team has prepared the evaluation matrix at Annex G. This is based on the theory of change described in Section 4.2 above and in Annex E; it employs the evaluation criteria spelled out in Annex F, and it draws on the specific questions posed in the TOR. [Table 1 below] below summarises the sequence of main evaluation questions and subquestions.

The evaluation matrix addresses both global and country-specific questions and will also serve as guidance for the country case studies. (IR ¶5.2–5.4)

### *What was implemented*

57. The evaluation matrix developed by the ET appears as Annex G of the IR and as Annex C of the ER. It organised the enquiry under the six main evaluation questions (EQs) shown in Table 1 below; the main EQs were supported by about two dozen main subquestions, but many of these were themselves complex. The matrix showed, in broad terms, which evaluation criteria were relevant to each question (or vice versa); the key evaluation criteria were identified and defined in Annex F (Glossary) of the IR.

**Table 1 Main Evaluation Questions**

<b>EQ1 Has the SUN movement addressed the right issues?</b>
<p>1.1 To what extent are the objectives of the SUN movement consistent with the needs, priorities and strategies of beneficiary countries?</p> <p>1.2 Has the SUN movement filled a gap in the international and country-level architecture for addressing nutrition?</p> <p>1.3 Did SUN strategies contribute to a stronger focus on nutrition-related gender and gender equity issues?</p> <p>1.4 Did the SUN movement's approach strike the right balance between global and country-level actions?</p>
<b>EQ2 Has the SUN movement followed a clear, consistent and commonly understood strategy?</b>
<p>2.1 Are the SUN movement's goals, priorities and strategies clear at the various levels of the movement?</p> <p>2.2 Have the SUN movement's main inputs, activities and outputs adequately reflected its goals, priorities and strategies?</p> <p>2.3 How is SUN seeking to mainstream gender-consciousness throughout its activities, both nutrition-specific and nutrition-sensitive?</p>
<b>EQ3 What have been the results of SUN's efforts?</b>
<p>3.1 To what extent has SUN contributed to changed attitudes and procedures, thereby creating an enabling environment for scaling up nutrition?</p> <p>3.2 To what extent has SUN brought about changed policies and resource commitments?</p> <p>3.3 Are these changes leading to the scaling up of nutrition?</p> <p>3.4 Are there plausible links between the outcomes to which SUN has contributed and medium to long term impacts for intended beneficiaries?</p>
<b>EQ4 What accounts for these results (or lack of results)</b>
<p><i>Governance and Management</i></p> <p>4.1 How effective have SUN's governance and management arrangements been?</p>
<p><i>Efficiency</i></p> <p>4.2 Concerning its own activities, has the SUN movement used its resources efficiently?</p> <p>4.3 Have the transaction costs of SUN been reasonable?</p> <p>4.4 Has SUN's advocacy for nutrition solutions taken enough account of efficiency considerations? (e.g. in the balance between nutrition-specific and nutrition-sensitive options)</p> <p>4.5 Has SUN achieved the right balance: between global work and attention to countries? between being inclusive (number of countries involved) and being effective in providing in-depth support to countries?</p>
<p><i>Coherence</i></p> <p>4.6 Have the SUN movement's various component activities reinforced each other (amounting to more than the sum of their parts)?</p> <p>4.7 How well have SUN's activities complemented other initiatives at global and country level?</p>
<p><i>Context</i></p> <p>4.8 What contextual factors (anticipated or unanticipated) have positively or negatively affected the achievement of SUN objectives?</p>

*Monitoring, Learning and Adaptation*

4. 9 How well has SUN learned from experience and adapted accordingly?

**EQ5 How sustainable is the SUN movement?**

5. 1 Are the emerging results of SUN likely to be durable?

5. 2 How well is SUN contributing to systems development (helping to develop the appropriate national policy and institutional architecture to deliver nutritional outcomes sustainably in the medium to long term)?

5. 3 Is the SUN movement itself sustainable?

**EQ6 How should SUN evolve in the short, medium and longer term?**

*The sub-questions under this EQ are all formative. The precise questions will be refined in the light of emerging evidence as the evaluation proceeds. The SUN ICE will draw on its summative findings (above) to present alternative options, and will link its recommendations to principles of aid effectiveness and development effectiveness, with reference also to the experience of comparable partnerships.*

6. 1 Is SUN likely to remain relevant? if so, which aspects/components are likely to remain relevant and for how long?

6. 2 What are SUN's relevant strategic options in the short, medium and longer term?

6. 3 What are the corresponding implications for SUN's governance and management arrangements?

*Source: IR Table 2.*

58. As noted in ¶17 above the TOR included (in its Annex E) over 70 questions for the evaluation to consider. The evaluation matrix attempted to make this manageable by organising EQs in to a hierarchy with a limited number of main EQs; the ToC influenced the questions included in the matrix and the way that sub-questions were seen as providing supporting evidence to address the main EQs. The IR included an annex to map the TOR questions onto the matrix (see IR Annex H). In a number of cases the matrix highlighted limitations in the ET's expected ability to answer the question; for example:

TOR Issue/Question	Where/how dealt within evaluation matrix
[Achieving and measuring concrete outcomes] To what extent has SUN moved (and/or is moving) beyond its initial focus on structures, capacities and processes that can feed into results to a focus on achievement of outcomes and intermediate outcomes? To what extent is program coverage in nutrition actually increasing at country level?	Several of the subquestions under EQ3 address this. Note, however, that our ability to determine whether programme coverage in nutrition is actually increasing at country level, will depend on the ready available of secondary data. This is known to be, at best, patchy. We expect the Global Nutrition Report will assemble the best available global data; we will focus our efforts on the case study countries, and recognise that we may have to fall back on qualitative or impressionistic judgements.
[Trust Fund] Should the Multi-Partner Trust Fund -- for catalytic financing at country level when other financing is not available -- be continued? If so, what is the evidence and justification and should its volume or scope be expanded?	Under EQ2.2 ( <i>Have the SUN movement's main inputs, activities and outputs adequately reflected its goals, priorities and strategies?</i> ) we will broadly review the performance of the MPTF. However it is beyond our scope to undertake a full evaluation of the MPTF.

*Source: IR Annex H*

59. We discuss below how the EQs were used to support the ET approach to interviews (see ¶67ff) and to the CCSs (see ¶83ff).

60. The EQs were also used to structure the ER: it broadly follows the sequence of EQs, and footnotes make explicit which main EQs and subquestion each chapter/section of the ER addresses. The ER also provided a 20-page *Matrix of Findings. Conclusions and Recommendations* (ER Annex Q) which mapped findings on the major questions onto conclusions and recommendations.

*Evaluation team reflections*

61. The evaluation matrix could undoubtedly have been elaborated further with more detail about the evaluation criteria and data sources to be used in addressing each subquestion. However, a balance has to be struck in any complex evaluation between upfront time spent in refining evaluation instruments and the time spent on the evaluation proper. This trade-off was particularly acute in the case of the SUN ICE.

62. Similarly, the pressure to demonstrate that every question has been answered runs up against constricted resources, an unforgiving timetable and understandable demands for a concise report.

63. QAA comments during drafting of the ER implied both (a) that there should have been more effort to reduce the number of questions in the evaluation matrix and (b) that the ER should have been more systematic in demonstrating the extent to which each subquestion had been answered. On the latter point, the differing perspectives of the QAA and the ET are set out in Box 8 below.

## Box 8 Role of EQs and Evaluation Matrix: QAA comments and ET response

### QAA comments on EQs

.... the evaluation team faced two major challenges: defining the boundaries of the SUN ICE, and conducting the evaluation in a short timeframe with a limited budget. The evaluation team embraced these challenges and defined an ambitious approach in the Inception Report. This elaborated a set of evaluation questions based on a preliminary theory of change with underlying assumptions that was constructed, among other things, from the SUN Movement's four strategic objectives and its approach to monitoring and evaluation and accountability. The evaluation matrix of the Inception Report contained 6 main evaluation questions and 26<sup>15</sup> subquestions.

Addressing all these questions in sufficient depth would have been a major challenge for any evaluation team, even given unlimited time. As it was, the evaluation team had to make trade-offs between questions to address and with different degrees of depth.

The final evaluation report does not provide a matrix which enables a direct comparison with the plans set out in the Inception Report, or explicitly state which evaluation subquestions were addressed in what depth. However, the impression of the QAA is that while overall answers were provided to the six main evaluation questions, many of the subquestions feeding into these were not addressed in much depth.

Source: Compton et al 2015

### ET perspective on EQs and Evaluation Matrix

I read the [QAA comments on ER draft v0] comments (perhaps wrongly?) as implying that a filled-in evaluation matrix with a detailed answer to each EQ is a required element of a good evaluation report, and that each and every EQ and sub-EQ has to have a full answer or a good excuse. We never expected to be able to satisfy either criterion. As I see it, the reason sub-EQs (and sub-sub-EQs) are grouped as they are is because they help towards answering the main EQ; more often than not, the evaluator has to make a judgement on the main EQ without having full answers to all the sub-EQs. (A similar issue has come up in recent review of MOPAN methodology, where it was found that making the questions too granular made it impossible to fit them to particular organisations, but you could nevertheless ask comparable, and useful, questions across organisations at a higher level of aggregation.) As for the matrix (filled-in) being an essential part of the report, I can't find that injunction in my favourite evaluation guide, the GRPP Sourcebook [IEG 2007], and I can think of many examples of good reports unencumbered by such a matrix.

Source: Supplementary Note 1 of ET response to QAA comments on first draft of the ER

64. As regards the number and scope of the questions, the QAA have commented as follows:

The QAA did raise the issue of the broad scope of the evaluation questions, the high expectations, and short timeframe at the inception phase, saying that "*Serious consideration should be given to eliminating some subquestions ... and reducing expectations of others ... and 'managing down' stakeholder expectations at this inception stage, given the very tight timeframe and resourcing for this evaluation.*" The evaluation team agreed and did indeed make efforts to reduce the scope of the work. However with hindsight, both the evaluation

---

<sup>15</sup> There were also numerous sub-sub-questions .

team and the QAA could have done even more at the inception phase toward getting agreement on a reduced scope of work that would have given the team more time for analysis and consultation, and to address acknowledged key issues such as gender in more depth. Once the Inception Report was agreed, the evaluation felt like a roller-coaster, with the team working flat out to collect a lot of data in a short period of time. It was not possible to go back and try to renegotiate the scope. (Compton et al 2015)

65. The ET agrees that the covering the scope of work defined in the TOR was a fundamental difficulty, given the limited time and funding allocated. However it is rather simplistic to imply that reducing the number of evaluation subquestions would have been a useful way forward. As shown in Section 3 above, it is difficult to anticipate which lines of enquiry will be prove to be most important, or most difficult (e.g. in terms of data availability). At the time the IR was being prepared it was already too late to address the fundamental issue that the ICE was launched too late (see Section 5 below). The QAA are quite right that "many of the subquestions were not addressed in much depth", but it does not follow that eliminating questions in advance is a useful approach, or that even more reflection in the ER on which questions have been answered in more depth than others would have been a good use of ET time.

66. **Lessons:** (a) drafters of terms of reference should not allow the number of evaluation questions to proliferate; (b) however, the scope of an evaluation's work is determined more by its overall purpose than by the number of subsidiary questions posed; (c) the evaluation matrix should be recognised as a pragmatic tool: it is unlikely that any complex evaluation will collect complete answers to every sub-question – the aim is to collect sufficient evidence to make credible judgments on the major issues.

## Interviews and participant observation

### *What was intended*

67. The importance of interview evidence was stressed in the TOR, the IR and the ER. Thus:

Comprehensive evaluations of global partnerships are recognised to be very challenging, because of the complexities of what is being evaluated, the varied perspectives and expectations of different stakeholders, and, usually, the lack of firm evidence of final impacts. Moreover, comprehensive evaluations are not an abstract diagnostic but themselves part of the partnership's process of reflection and strategising. As such, thorough consultation with stakeholders is crucial. Accordingly, our main primary data collection was through interviews. (ER ¶1.5–1.6)

68. The IR incorporated an extensive stakeholder analysis (IR Section 3.3 and Annex D) which was used to ensure that interviewees were drawn from across the spectrum of SUN stakeholders. The approach to interviews was set out as follows in the IR:

Interviews will be the main form of primary data collection. We will draw on Mokoro's experience to use them in a way that maximises their analytical power and the possibilities of

triangulation. Interview notes will be systematically written up, consolidated and shared among team members on the internal team website (see Annex J). If allowed by the interviewee, interviews may be recorded so as to facilitate accurate note-keeping and allow full transcripts of especially significant interviews. To respect interviewee confidentiality, the interview notes and any recordings will be accessible only to team members. The consolidated file of interview notes will facilitate searches on key thematic terms, country names, and so forth.

Identification of interviewees will draw on the stakeholder analysis (Annex D). Most interviews will be by telephone or e-conferencing. Where the quality of responses is not compromised we will interview groups, recognising that interviews are part of a consultative process, not just evidence-extraction. Where the opportunity arises we will observe SUN meetings and events.

For each main category of interviewee, we are developing semi-structured interview guidelines for consistency, and to ensure that interviews focus on areas that can add most value. We will undertake relevant documentary analysis ahead of the interviews, and will use the evaluation matrix (especially the key questions in Table 2 above) to inform our interview checklists.

In approaching stakeholders during the evaluation and in all interactions, the evaluation team will abide by UNEG norms and standards. The evaluation team will exercise sensitivity both in scheduling interviews and in the conduct of interviews. (IR ¶6.7–6.120)

### *What was implemented*

69. In total, 107 global level and 167 country level interviews were carried out over the course of the evaluation (as listed in ER Annex B, Table B2). All core team members conducted interviews. As far as possible team members participated in the interviews most relevant to their areas of focus; the ET consulted in advance about interview priorities and shared suggestions about key topics to raise and triangulate with particular interviewees. All interviews were promptly written up along the lines of the standard format set out in the IR Annex J and shared via a team-only dropbox. A significant number of the most important interviews were fully transcribed (where interviewees gave advance permission for recording). CCS as well as global interviews were added to a Word compendium, which eventually ran to more than 1,000 pages and 350,000 words. Simple searches were used to compile systematic extracts on such topics as gender, governance, efficiency, M&E, particular networks and countries, and so forth.

70. Interviews were supported by direct observation. ET members joined SUN network conference calls, a Lead Group meeting, and various SUN network meetings. Several members of the ET attended the SUN Global Gathering in Rome in mid-November. Such observation, along with the CCSs – see below, was very important in giving team members a rounded picture of how SUN operates in practice.

### *Evaluation team reflections*

71. The interviews were essential to the ET's understanding of SUN. Inevitably, informants' time is limited and the team sought to make the most of each interview by doing relevant reading in advance (not only drawing on the shared library, but also reviewing notes of previous interviews, often undertaken by different members

of the team). As noted in Box 9 below, we did not develop formal interview protocols beyond the guidance provided by the evaluation matrix.<sup>16</sup>

**Box 9    Use of interviews: QAA comments and ET response**

**QAA comments on first draft of the ER:**

Interview and observation protocols are not included, but reference is made to inception report. To note that the (accepted) QAA observations on the inception report included: “We anticipate that specific indicators, methods and instruments (eg interview guidelines) will be refined and improved as the evaluation proceeds”. It would be useful to comment briefly here (and more extensively in the later note on methods) whether the plans in the inception report were followed, what was added or refined, what worked well and less well.

There is no description of how interview data were analysed (also cannot locate this in inception report).

**Response to QAA comment on first draft**

**Supplementary Note 2: Interview Notes and Protocols (relates mainly but not only to your comments at 3.5)**

We wrote up and shared notes on our interviews, and for a good number of key ones, we made full transcripts. We had interview note compendiums for country cases, and we also had an overall compendium – a Word file with the full notes on all our global and country interviews. This was available to everyone on the team to do their own searches and we also systematically extracted what people said about various topics, including gender and governance issues. The compendium runs to more than 1,000 pages and over 350,000 words. We did not use any specialised software to analyse this data. As far as demonstrating triangulation is concerned, often the key triangulation was between different interviewees. In most cases we were not free to put names to quotes, and even if we did we can be accused of selective quotation. Nevertheless we think we have a pretty good oral history of the SUN movement to draw on. On some previous evaluations, QA have inspected our interview notes to judge their quality and assess our use of them, and we would be happy to arrange the same for you. (The same applies to the text comments from the survey – not such a large file.)

We did not develop much in the way of formal interview guidelines or other protocols. We all had the evaluation matrix as a prompt. For any specific interviewee time is limited, and you rarely know in advance what will be the most fruitful way to steer the conversation; you certainly can't ask every question you might like to, and you may learn more anyway from what people choose to raise than from firing questions at them (so, no we didn't make gender a question in every interview). ... I'm not sure how useful it would have been arming ourselves with more rigid protocols or analytical frameworks, but, compromises again, we haven't had much time to wonder about it.

---

<sup>16</sup> However, many interviewees were also given the opportunity to respond to the e-survey, as discussed later.

72. Overall the ET believes that the compilation and use of interview evidence was a very strong feature of the evaluation. We regret that the QAA could not find time to assess this material directly.<sup>17</sup>

73. **Lessons:** *Interviews are key to a comprehensive evaluation, and it is important to ensure that interview evidence is systematically recorded and shared across the evaluation team. The quality of interview evidence should be assessed directly by quality assurance advisers.*

## Literature review and secondary data analysis

### *What was intended*

74. The approach to gathering of secondary data was set out in the IR:

6. 2 Given the limited time and resources available for this evaluation, the TOR rightly stress the need to maximise the use of existing data and to draw as much as possible on existing information. No primary quantitative data collection besides perceptions data amassed through the survey (see Section 6.6 below), is envisaged. Rather, the evaluation will make full use of the large volume of information that exists regarding nutrition and the SUN movement.

6. 3 During the inception phase, basic data regarding SUN countries was assembled, with the primary purpose of aiding country case study selection (see Annex L). In addition, a data mapping was undertaken to assist team members in identifying what information was readily available, from which sources and for what years and where this could be located in the evaluation library (see Annex I). Going forward, it is expected that the evaluation will draw primarily on the data sources summarised below, for the purposes of the global analysis and in the preparation of country dossiers.

6. 4 For information directly related to performance of the SUN movement, it is expected that the primary country-level data source will be Country Fiches, which are available for 2010 through 2013. These provide data on progress against the four SUN processes (and since 2013, their related sub-markers – see SMS 2013a), as well as a country's overall state of preparedness. The SMS usefully compiles these into an annual compendium, which is used as a key input in the SUN progress reports. For the most part, these documents track a consistent set of indicators each year, enabling longitudinal analysis. Where this isn't the case, the SMS may be able to fill priority data gaps. Certain analyses, including self-assessment by in-country networks against the sub-markers under the four SUN processes, have been introduced only recently, and 2013/14 will serve as a benchmark.

6. 5 Country-level documentation will be complemented by global-level SUN reports. Of particular use to the evaluation will be Annual Progress Reports, which are currently available for 2011 through 2013. The 2014 report is due to be presented to the Lead Group in

---

<sup>17</sup> The QAA commented:

To triangulate and analyse a large amount of interview data require a systematic and iterative approach for successive testing and validation to avoid undue influence from operational pressures. Overall, we found that the evaluation team did manage to present a credible analysis of the evidence assembled to respond to the evaluation questions. Because of the tight timeline, however, we had to assume that their selection of quotations from the interviews and survey were representative views of the stakeholders consulted. (Compton et al 2015)

September 2014. The team will also utilise the 2012 SUN M&E Framework Baseline report, which provides baseline information (for 2012) in terms of impact (that is, nutrition indicators), outcomes (behavioural characteristics of the constituent parts that make up the SUN Movement) and outputs (services provided by the SUN governance and management structures) –see SMS 2012r. The SUN 2014 Annual Progress Report is expected to provide the first comprehensive update against this baseline.

6. 6 In terms of data on country nutrition status and trends in performance, it is beyond the scope of this evaluation to assess the extent to which SUN has led to improved nutrition, and too early in the lifecycle of the movement to credibly make such a judgement. However, the evaluation will seek to assess whether SUN has been able to move countries in the direction of plausible pathways for improvement (question 3. 4 in Table 2 above). In determining this, data concerning nutrition outcomes, intervention coverage, determinants, policies and legislative environment, as well as the availability of resources (domestic and externally financed) for nutrition, will be critical. Again, the evaluation team will seek to capitalise on existing studies, which cover “what works” in nutrition (for instance, the Lancet series), how nutrition is governed (including the Hunger and Nutrition Commitment Index), costing of nutrition plans (as emerges from the work undertaken by the MQSUN network), and country nutrition profiles (such as those produced by the World Bank, UNICEF, and other partners). Fortunately, our ability to gather and analyse this data in a systematic and comprehensive manner stands to be greatly facilitated by the forthcoming publication of the first Global Nutrition Report, which is due to be launched at the second International Conference on Nutrition in November 2014. The report will bring together comprehensive nutrition data for over 70 indicators from multiple sources, filling in gaps and constructing an estimated 190 country profiles. The lead author has agreed to share the specification of the profiles and a list of indicators with the evaluation team, and to share an early draft of the report as soon as it is available. (IR ¶6.2–6.6)

### *What was implemented*

75. The evaluation was supported by an extensive literature review, as indicated by the size of the bibliography appended to the ER. Most of the documents and articles listed in the bibliography were directly available to team members via the team's electronic library.

76. The strategy of relying as much as possible on data compilations by the SUN SMS and by the GNR team was followed; collaboration from both teams was unstinting and very much appreciated. However, in both cases it meant that key data sets were not available until late in the evaluation process. This had implications for the evaluation process and report drafting which are discussed in Section 5 below.

77. An unanticipated problem, was that ICE analysis revealed serious weaknesses in SUN's M&E system, which were documented in the ER Section 5.4 and Annex M, and led to the conclusion that:

SUN's M&E framework is insufficiently rigorous and too subjective; it is therefore not reliable as a guide to individual country progress over time, nor for comparing progress across different countries. (ER ¶7.11b)

*Evaluation team reflections*

78. There was no viable alternative to relying on secondary data compilations by the SMS and GNR team, but their timing added to the immense pressures at the report drafting stage (see Section 5 below).

79. It will be very unfortunate if SUN stakeholders fail to take seriously enough the ICE findings about the M&E system. The ET have not been directly involved in the post ICE discussion of SUN strategy, but there are indications of some complacency in the face of these findings – as indicated in the QAA comments in Box 10 below.

**Box 10 QAA reflections on ICE M&E findings**

We found section 5.4 on monitoring, learning, and evaluation to be a strong section based on solid evidence, as well as recommendation 9 on strengthening monitoring, evaluation, and learning. ...

We have also noted the skepticism of some stakeholders with respect to SUN ICE conclusion 4 and recommendation 9 on monitoring and evaluation, and the apparent lack of consensus on how to proceed in this area. There clearly exist tensions and a need to strike a balance between national-level monitoring of nutrition indicators, project-level M&E requirements of donors and other development partners, and a desire to track global progress while facilitating comparisons among countries. Given how slowly nutrition indicators change over time, we find the evaluation's emphasis on monitoring output-level results and country-level policy and institutional outcomes to be appropriate.

We would emphasize that monitoring is important not only to assess progress in implementing individual activities but also to facilitate the next cumulative assessment of the SUN Movement's performance in achieving its objectives – presumably towards the end of the 2016–2020 strategy period. Monitoring is also important to enhance policy dialogue by identifying issues that require policy and institutional responses beyond the scope of individual nutrition interventions.

Source: Compton et al 2015, section 3.6.

80. The ER cited the 2014 SUN Progress Report as follows:

The progress report notes stronger performance of countries that have been in the movement longer: “the longer countries are within the SUN Movement, the more likely they are to establish evidence-based nutrition policies and legislation that supports effective implementation.” Specifically, it sees the following process as being “particularly evident” in countries that have been in the Movement for more than two years:

- Once a new national nutrition policy has been endorsed, there is a rapid adjustment of programmes in line with the new policies and associated legislation.
- As relevant policies and legislation are implemented, governments and other nutrition actors move quickly to ensure the widespread dissemination of policies to those who are required to implement them.
- They also tend to follow up with a focus on aspects of policy that need further attention and laws that need to be updated. (ER¶4.25)

81. In view of the serious weaknesses in SUN's monitoring systems uncovered by the ICE, the ER noted: "The ICE is unable to confirm/otherwise the relationship between policy implementation and length of membership that is claimed in the SUN progress report... " (ER ¶4.28). Perhaps, in order not to alienate those involved, we put the point too gently. We would have been equally justified in pointing out that, on the evidence the ICE reviewed, such a process is not "particularly evident" at all, and the SUN progress report could be interpreted as somewhere between wishful thinking and whistling in the dark.

82. **Lessons:** (a) *Reliance on secondary data made sense, and the data collected for the GNR were extremely valuable. However the ICE was locked into the timetable for publication of the GNR and of SUN's annual report, which further compressed the time available for data analysis and report writing.* (b) *SUN's M&E system, in the form reviewed by the ICE, does not reliably track performance either within or across countries. Both the performance of SUN 2.0 and its eventual evaluability will be seriously prejudiced if the ICE recommendations on M&E are not effectively taken forward.*

## Country case studies

### *What was intended*

83. The TOR highlighted importance of country case studies:

The principal focus of the evaluation will be on the SUN countries and on the added value of the Movement over and above what countries can achieve on their own. (TOR ¶19)

It is essential for all aspects of the evaluation that the evaluation team visit SUN countries, in addition to working through other forms of enquiry such as questionnaires and telephone interviews. It is through country-level assessments that the most important findings, lessons, conclusions and recommendations of the evaluation will be derived. (TOR ¶28)

84. Eight country case studies were planned (the maximum feasible number within the ICE time and financial budgets). Bangladesh, Burkina Faso, Ethiopia, Guatemala, Indonesia, Mozambique and Senegal were selected using the criteria shown in Box 11 below. Case studies could not be representative in any statistical sense, and the IR noted that:

The utility of the country studies will depend on their ability to explore "what works" (Woolcock 2013). A theory-based approach is the key to judicious extrapolation of case study findings, and the evaluation matrix ... has been designed to serve as a basis for country-level as well as global analysis of how SUN operates. The case studies will explore how well the SUN support is tailored to the specific nutritional constraints and priorities in each country, and the influence of different contexts on SUN's performance.

### **Box 11 Selection criteria for case study countries**

<b>criteria</b>	<b>Rationale</b>
Choose countries which joined SUN not later than 2012	Likely to learn more from these than from more recent joiners. The verification survey (see Annex M) will test case study findings in countries which joined the movement more recently.
Include countries with largest vulnerable populations (hence Indonesia, Bangladesh and Ethiopia)	To link analysis to numbers of potential beneficiaries. This selection to be balanced by smaller countries in the rest of the sample.
A balance of countries of different income status	To ensure coverage which reflects the different level of economic development in SUN countries (as per the TOR).
Geographical variety	Ensures regional coverage in response to the TOR requirement for countries to be selected from across different areas.
Fragility status	To ensure SUN is evaluated in both fragile and stable-state contexts
State of nutrition preparedness	To ensure that countries across the range of preparedness levels are assessed in order to understand how SUN adds value in countries at different ends of the preparedness scale.
Whether REACH is active in the country or not	To evaluate countries where REACH is active and not active to understand how REACH interacts with SUN.
Whether SUN business network is active	To include SBN activity among the detailed studies
Whether SUN MPTF is active.	To include MPTF activity among the detailed studies
Draw on existing team expertise and country experience	The time available for studies is compressed, and better value will be achieved if it is possible to deploy consultants already familiar with the country or region concerned. Similarly, making full use of the core members of the evaluation team will maximise cross-learning and consistency in the approach to the studies.

Source: IR, Annex K, Table 5.

85. As already noted, the ICE drew as much as possible on core ET members to undertake the case studies. In each case a one-week visit by two evaluators was envisaged.<sup>18</sup>

86. Case studies were scheduled for September and October to allow time for adequate preparation and to ensure that country-level enquiries were informed by the document review and analysis that would already have taken place. It was envisaged that later visits would be influenced by experiences from earlier ones, but there was not scope, given the timetable, for formal piloting of the case studies.

87. Because of the constraints on the ICE, separate formal country case study reports were not a specified or offered deliverable. The IR (¶6.13) set out key stages shown in the left-hand column of Table 2 below.

#### *What was implemented*

88. Implementation of case studies closely followed the approach described in the IR – see the right-hand column of Table 2 below.

---

<sup>18</sup> The ICE research coordinator was a third member of the Tanzania country team. The Evaluation Manager joined the Ethiopia country visit as an observer.

**Table 2 Implementation of Country Case Studies**

<b>As planned (IR ¶6.13)</b>	<b>As implemented</b>
<p>For each case study country a dossier will be prepared in advance of the mission. The dossier will draw on standard SUN and non-SUN data available (cf. Annex I) as well as collecting relevant country-specific reports, plans and policy documents. The dossier will include a stakeholder mapping.</p>	<p>Done as planned. As well as a briefing document including stakeholder mapping, a country-specific chronology, standard available data and a select bibliography, an e-library was assembled for each case study country, along with relevant extracts from interview notes.</p>
<p>Ahead of the week-long visit the CCS leader will prepare a brief note identifying the key issues to explore and stakeholders to contact. The programme will be prepared in close collaboration with the Focal Point. In addition, efforts will be made to publicize the country missions in advance, particularly within the relevant in-country networks, to facilitate broad engagement.</p>	<p>Done, and discussed between the leader of the CCS and the ICE team leader in each case, although time pressures left little time for intra-team discussion of these issues notes ahead of each country visit.</p>
<p>Each mission will start with a briefing, and an in-country feedback session on the main emerging findings will be organized at the conclusion of each country visit.</p>	<p>Done; the early visits provided valuable guidance to subsequent CCSs on the most effective format for these introductory and valedictory presentations.</p>
<p>A brief (5–10 page) aide memoire will be produced after the return from each mission to capture the main findings. A standard format will be developed for this aide memoire, based on the EQs, to facilitate comparability between findings. In-country interviews will be added to the evaluation's interview compendium.</p>	<p>After some trial and error, the eventual internal products (for discussion and use within the evaluation team) were:</p> <ul style="list-style-type: none"> <li>• country-specific interview notes;</li> <li>• the valedictory powerpoint presentations as shared with in-country stakeholders;</li> <li>• a narrative report using standard headings as a working document (not written for publication);</li> <li>• an Executive Summary; these were drafted to publishable standards and appeared in the ER (see ER Annex O); they ranged from 2,000 to 3,500 words in length.</li> </ul> <p>The draft Executive Summaries were shared with country stakeholders and amended in response to their comments.</p>

<p>A standard checklist, based on the evaluation matrix, will be developed to guide interviews, and we will develop an evaluation grid to serve as a basis for recording country findings in a common format that facilitates cross-country comparisons according to topic.</p>	<p>A sub-set of EQs and sub-EQs were used as a framework for the narrative report and also for the table of country-wise responses to evaluation questions that appeared as Annex P in the ER.</p>
---	--

### *Evaluation team reflections*

89. The selection criteria for the CCSs worked well and yielded an appropriate diversity of countries, all with long enough engagement in SUN to make for a meaningful review. The strategy for making the CCSs feasible within the ICE constraints was broadly successful, although, even without the requirement to produce full formal reports, the demands on CCS team members, relative to their time allocations, were very onerous.

90. The case studies were crucial in informing the ET's understanding of how SUN and its component networks operated in practice. They reflected the range of contexts in which SUN operated and demonstrated that simple generalisations about SUN's relevance and effects do not apply across all contexts (see the ER's overview of case study findings (ER Section 4.3. ¶4.9–4.14); this noted contrasts between ostensibly similar pairs of cases (Ethiopia and Guatemala; Bangladesh and Indonesia; Tanzania and Mozambique; Burkina Faso and Senegal) and concluded:

This report draws throughout on detailed findings from the case studies. Meanwhile this brief overview underlines how varied the case studies were, with quite different patterns of findings emerging even from pairs of countries that might have been expected to be more similar. (ER ¶4.14)

91. Nonetheless the case study approach was severely handicapped by the time and budgetary constraints on the ICE. This limited the depth of the studies, and also the extent of interaction between case study teams and the country stakeholders; although in-country feedback sessions were useful and stakeholders had an opportunity to comment on the CCS executive summaries, they could legitimately feel that more extensive country studies would have been useful.

92. The QAA comments on the CCSs are tantalising because they imply that the CCSs could have been done better but without specifying how. Thus:

It is appropriate to use mixed methods to assemble evidence in a comprehensive evaluation of an initiative like SUN, and then to triangulate the evidence derived from the various sources to draw findings and conclusions. Thus, document reviews and stakeholder interviews are standard features of all such evaluations. **For additional instruments such as electronic surveys and country visits, which can be costly and time-consuming, it is important to pay particular attention to their design and implementation for them to make a cost-effective contribution to the findings and conclusions.**

[...] while recognizing the considerable time pressure, we think that some modifications to the design and implementation of the survey and the country case studies could have

enhanced their contributions even more to the evidence base of the evaluation. (Compton et al 2015. emphasis added)

93. The Evaluation Manager, who took the opportunity to join and observe a country visit, and who followed in detail the ET's planning and implementation of the CCSs, reflected as follows:

Mokoro conducted eight country case studies. Countries were selected on the basis of diverse criteria also ensuring that each had joined SUN no later than 2012. Each visit was of one week's duration. As the evaluation focus was very much the SUN countries, the country visits proved to be a vital, if not the most important, aspect of evidence gathering. It is likely that in a less time-constrained review, more case studies would have been favoured and with longer durations. Given the multi-sectorial nature of nutrition and consequently the diversity of stakeholders to be engaged, country visits of greater than a week would prove valuable.

*Lesson learned:* Country case studies should continue to be prominent elements of any future evaluations but with potentially more countries covered and in greater depth. Case studies could also feature as a management tool during the five-year span of SUN 2.0 as it provides an excellent reality check on what is being achieved on the ground. (de Mel 2015, section 4.6)

94. **Lessons:** *The country case studies were crucial for a country-focused movement. The ICE had to adopt second-best approaches to fit the case studies into the time and resources available. The next ICE should incorporate in-depth country case studies, conducted over months rather than weeks, and arranged to precede the main phase of the evaluation, with published CCS reports feeding into the main phase of the evaluation. They should be linked to a more systematic comparison of SUN and non-SUN countries.*

## E-survey

### *What was intended*

95. The survey was intended as a synthesis and verification tool, to be conducted in the later stages of the evaluation to test the wider relevance and comprehensiveness of the preliminary findings emerging from the country case studies and other evaluation instruments. Recognising that the SUN Movement is a very diverse and far reaching collection of stakeholders, it was also a means of reaching more people than would otherwise have been possible to interview directly. It was additionally expected that the survey would be used to “road test” the feasibility and political acceptability of preliminary recommendations.

### *What was implemented*

96. The survey was implemented in November as planned, after the country case studies were complete. A number of the country case study reports by that point had not yet been written up, and so the e-survey wasn't able to draw on the reports directly, but its drafting was a collaborative effort with inputs from all team members, who reflected their experiences from the case studies. Furthermore, the majority of the stakeholder interviews had been completed by that stage, and the Interim Progress

Report had also been written, and so these influenced the survey questions significantly. The QAA also provided useful feedback for strengthening the survey questions.<sup>19</sup>

97. The survey served to solicit opinions on the following key topics:

1. *Performance of the SUN movement to date*, covering issues including understanding of the movement, multi-stakeholder engagement and coordination, multi-sector plans for nutrition, the role of gender in the movement, and governance.
2. *Future of the SUN movement*, including objectives and activities, as well as membership and timeframe of the movement going forward.

98. At the stage that the survey was being drafted (late October), it proved too early to expect recommendations to have emerged even in draft form,<sup>20</sup> and thus we were not able to use the survey to “road test” recommendations as envisaged. However as a step towards this, the survey solicited views on the movement's future, which included a broad range of issues which we knew, even at that stage, could potentially evolve into recommendations.

99. The survey was administered using SurveyGizmo, a web-based programme selected for its design functionality. It was sent to 711 contacts, which had been provided by the SMS, SUN global network coordinators, and team members, which covered all the main stakeholder groups, although not equally. It was available in English, French and Spanish, and guaranteed anonymity to solicit frank responses. It was open for two weeks (as opposed to the planned one week) to allow us to use the Global Gathering as a last opportunity to encourage responses. In the end, it received 186 responses, representing a respectable 26% response rate as well as a high absolute number of respondents.

100. The analysis took longer than initially envisaged, but also proved more fruitful than the evaluation team had expected. Quantitative answers were disaggregated by stakeholder groups and by level of respondent (Global/regional, and country), which revealed some insightful variations in responses. These results and analysis were presented in full in Annex G of the report, and some were reproduced in the main body of the report where they related to specific evaluation questions.

101. The comment box responses also served as a rich source of additional perception data, and were taken into account by the team when assessing the

---

<sup>19</sup> The QAA's "Statement" of May 1 states: "The QAA made a number of suggestions that (in our view) would have made the survey more inviting and accessible. For example, rather than asking questions about the SUN Movement as a whole, we suggested asking recipients about their own experiences and interactions with the SUN Movement, and then aggregating their responses to get an overall picture." (Compton et al 2015) The evaluation team appreciated this advice, and believe this is the approach that was adopted. See the survey instrument appended to the ER Annex G (bearing in mind that skip logic was used to bypass questions not relevant to a particular respondent's experience).

<sup>20</sup> The country case study reports, Options Discussion Paper, and the Global Gathering were all critical to the formulation of recommendations, and these came after the survey was sent out.

findings and writing the overall report. A representative selection of the comments were (anonymously) reproduced in Annex G also, to illustrate findings emerging from the quantitative results.

*Evaluation team reflections*

102. Overall, the e-survey proved a valuable evaluation instrument. The following factors were key to this:

- (a) The support of the SMS in particular in furnishing the team with contacts was vital. In such a disparate and large group of people as the SUN Movement, the evaluation team would have struggled to put together a comprehensive invitation list otherwise.
- (b) The software used, SurveyGizmo, served its purpose well. Its functionality in skip logic, and respondent tracking and reminders, was particularly well received, and although we did face some initial design problems (around the visibility of the language bar) these were ironed out quickly. The relatively good response rate is in part a reflection of the accessibility of the system.
- (c) The survey employed open-text comment boxes liberally, and we reaped high quality responses from them. The comments were rich in additional information and supportive evidence, and were used widely by the evaluation team in producing the final report.

103. There were some weaknesses with the survey, which may serve as potentially useful lessons for other evaluations using such an instrument:

- (a) The survey respondents were not evenly spread across the stakeholder groups. Two categories in particular were under-represented: the private sector and governments of non-SUN countries. The fact that the analysis included systematic disaggregation of the results to look for any differences in mean responses, meant that for those underrepresented groups the sample size got quite small, which limits what you can legitimately say from the results. More intensive efforts up front to expand the contact lists for these groups would have been worthwhile.
- (b) After the first draft of the analysis had been completed, we had some back and forth discussion with the QAA about some of the statistical methods used (see Box 12) and our methods of selecting illustrative quotes from the qualitative responses. In the end they were sufficiently assured to allow us to proceed with the methods employed, but the discussion would have been more useful and timely at inception phase rather than late into the evaluation, after the survey analysis was already complete.
- (c) Whilst the comments from survey respondents were rich and very useful, open text is challenging to analyse, and this contributed to the survey analysis taking longer than had initially been envisaged. This, teamed with the fact that we wanted the survey to be fairly late in proceedings in order to solicit

opinions on emerging findings from other sources, meant the survey analysis spilled over into some of the time initially envisaged for synthesis and reporting. This is one element of a broader issues relating to the unrealistic timeline set for the ICE.

**Box 12 Statistical Methods used in the e-Survey Analysis**

In the comments provided on the draft e-survey annex, the QAA raised concerns about the evaluation team's use of parametric statistics – like the mean – for ordinal data such as Likert scales (which is where response options are ordered, such as strongly agree / agree / disagree / strongly disagree). They were justified in doing so, as the potential limitations of the approach are widely documented in academic publications; one such paper labelled it one of the “deadly sins of statistical analysis” (Kuzon et al 1996). The line of argument is that with ordinal data, the response categories have a rank order but the intervals between values cannot be presumed equal; as such the data cannot be assumed to be normally distributed, and where you don't have a normal distribution, using parametric methods risks arriving at incorrect conclusions.

However, as acknowledged by the QAA panel, there is by no means a consensus on this point, and a substantial literature has amassed which berates the academic community's review culture as one which encourages “overvaluation of criticism for its own sake [and] inappropriate statistical dogmatism” (Bacchetti 2002). A counter-argument has been made which argues that the criticism fails to take account the robustness of parametric tests – that is, the extent to which the test will give the right answer even when assumptions are violated – which essentially means the chance of an erroneous conclusion is very limited at best. See for instance Norman 2010, whose paper shows that four- or five-point or longer Likert scales can indeed safely be regarded as continuous data for many purposes, making parametric statistics perfectly appropriate. Fayers 2011 shows that employing such methods can make type II errors more likely (that is, failing to detect an effect that is present), but only when there are *major* asymmetries in the underlying distribution, and furthermore, type I errors (mis-detecting an effect when there isn't one) are unlikely.

After some consideration, the evaluation team decided to maintain its use of parametric statistics in the ICE survey analysis (alongside other methods), in part because they are widely used and therefore likely to be understood by the audience. Indeed, the Stewardship Study (Isenman et al 2011), which the ICE survey sought to emulate, used them liberally. However in view of these potential limitations, and given the ICE's prevailing commitment to triangulation, no conclusions were drawn on the basis of survey results alone.

104. **Lessons:** *The e-survey was a valuable component of the ICE methodology. Conducting it towards the end of the investigative process enabled it to focus on the most relevant emerging issues. Liberal use of text boxes for open-ended responses provided rich material but added to the time required for analysis and write-up.*

## 5. Reflections on the evaluation process

### The effects of timetable and budget pressures

105. The full QAA-approved ICE report was delivered on 19 January 2015 (see ¶5–6 above), instead of on 31 December 2014. A planned review of the draft by members of the VSG was omitted from the process. There was time pressure throughout, and even more so in the final stages, when turnaround times for both the ET and the QAA were extremely tight, as shown in Table 3 below.

**Table 3 Schedule for ICE Report Finalisation**

Activity	Envisaged in Inception Report	Actual submission dates
Submission of v1 Final Report to EM and QAA	5 December 2014	w/o conclusions and recommendations – 16 December 2014 full report – 19 December 2014*
Provision of comments from EM and QAA panel on v1 Final Report	10 December 2014	EM comments – 23 December 2014 QAA comments – 23 & 24 December 2014
Submission of v2 Final Report to VSG	15 December 2014	[this step omitted; instead v2 report was submitted for further review by QAA with v3 then going to the VSG]
Provision of comments from VSG on v2 of the Final Report	by 19 December 2014	
Incorporation of final comments (with QAA to check revisions)	19-31 December 2014	v2 report to QAA – 8 January 2015 matrix responding to QAA comments – 10 January 2015 QAA comments on v2 – 13 January 2015
Submission of v3 Final Report	31 December 2014	(including Executive Summary) – 19 January 2015

\* The bulk of the annexes had been provided earlier, and the QAA commented on many of them by 11 December, but the consolidated report, with annexes incorporated in the document, was provided on 20 December.

106. As we discuss below, the timetable pressures had implications for the quality of the report preparation process and the stresses experienced by the evaluators and others closely involved with the ICE.

107. The fundamental problem, neatly described by the Evaluation Manager (Box 13 below), is that the ICE was commissioned too late and under-funded. Actually, the ICE was foreseen in the 2012 Revised Roadmap but initially envisaged

for 2015,<sup>21</sup> so the compression of the timetable was a combination of a drawn-out commissioning process and the belated realisation that ICE findings were needed at the beginning of 2015.

**Box 13 Evaluation Manager comments on time and budget**

4.2 Available time and budget

The Lead Group made its initial decision to seek a comprehensive evaluation in September 2013. There was then a protracted process to draft and approve Terms of Reference for the evaluation and a call for proposals culminating in the appointment of Mokoro only in June 2014. This allowed barely six months to undertake the evaluation, which was also constrained by a modest budget. As even the authors of the Terms of Reference recognized, this timing was not ideal. The evaluators themselves acknowledge that available time and budget placed restrictions on consultation follow up and influenced the number of country case studies. These time pressures eventually led to a need to extend the final reporting deadline. This was disruptive for all parties to the evaluation and for the initiation of the visioning.

*Lesson learned:* Given the complexity of SUN and the ICE evaluation process, the available time and budget was not optimal. Consistent with paragraph 4.1, early planning and securing of an adequate budget will be important for any future review during SUN 2.0.

*Source:* de Mel 2015

108. The QAA also judged that:

Unrealistic expectations for what could be achieved in the time frame had some adverse effects on quality of process, despite impressive efforts from the evaluation team. In the view of the QAA, the evaluation Terms of Reference (ToR) was too demanding given the time and resources allocated. The rushed nature of the process adversely affected the depth of the analysis of the extensive and rich material assembled by the evaluation team. The restricted timeline also meant that broad stakeholder consultations were not undertaken at the draft report stage, as is normal practice. (Compton et al 2015 – emphasis in original.)

109. The QAA imply that the way to address this problem would have been to limit the scope of the TOR, either during their original drafting or during the inception phase. This would not have altered the fact that the timetable for the evaluation, once it commenced, was organically linked to the timetable of SUN events. Of most significance, the 2014 Global Gathering was set for mid-November, closely followed by the ICN2. This influenced the scheduling of Lead Group meetings and the timetable for preparation of the SUN Annual Report and the GNR – both key inputs

---

<sup>21</sup> "An independent evaluation of the Movement's impact will be undertaken in 2015 to demonstrate how it has influenced SUN countries and other stakeholders to scale up nutrition and assess how financing has been used. The results of this evaluation will determine the Movement's future direction and mode of working." (SMS 2012b)

to ICE, as discussed in earlier sections. Findings of a survey undertaken by the donor network also became available around this time. These were all key inputs to the ICE, as was participation in the Global Gathering itself. It was therefore unavoidable that the main phase of drafting and completion of the ICE ER became compressed into the period between mid-November and mid-December. It was during this period that two of the most important pieces of ICE analysis also came together, namely the assessment of SUN's M&E system and the ET's comparison of SUN practice against previous experiences with multi-sector nutrition planning. This was also when it became possible to integrate e-survey findings with other material.

110. By now, however, all ET members had exceeded their original allocation of days. Ideally there should have been a full meeting of the ET (similar to the inception workshop) to bookend the evaluation, but in practice there was neither the budget nor room in the calendar to organise this, and the team had to rely on limited teleconferences and e-mail exchanges. The final stages of drafting – integrating material provided by all team members– increasingly fell on the team leader, the research team and one or two senior evaluators providing additional support. This limited the opportunity for the whole team to take ownership of the final report. The period between the first and second drafts of the ER (which spanned the Christmas and New Year holiday period) was dominated by efforts to respond to QAA comments on the first draft (see ¶117ff).

111. The risks of limited consultation were foreseen in the IR as follows:

However, the most serious risk that arises is to the consultative process that lies at the heart of this type of evaluation. We will be seeking many interviews over the European vacation season, and country visits will perforce be short. There will be limited time for stakeholders to comment on our interim report, and our final report will be delivered during another holiday period. We look forward to discussing with the VSG how best to mitigate these constraints, and to ensure that our report has broad stakeholder acceptance as a valid and useful foundation for the visioning exercise. (IR ¶8.3)

112. As discussed in Section 4 above, the ET considers that its interviews with SUN stakeholders were adequately comprehensive and thorough, but consultations with stakeholders around the ET's emerging findings were very limited – linked mainly to discussion of the Options Paper (see ¶6 above) at the Global Gathering. In the event there were two unanticipated consequences: (a) the stresses imposed on the ET by the lack of sufficient time for internal digestion and discussion of the emerging draft ER, and (b) knock-on effects on the quality of stakeholder consideration of the report after its delivery, which we discuss next.

### **Post-delivery inputs**

113. As we have noted, the evaluation schedule did not allow for the usual discussion by stakeholders of the draft report before it was finalised; in the event, even efforts to give VSG members an opportunity to comment on the draft had to be abandoned, and the only comments taken into account between the first draft and

the draft delivered in January 2015 were those from the QAA and (separately) the Evaluation Manager.

114. Once the draft ER was delivered to the VSG, the QAA review was deemed to have addressed the methodological quality of the report, and stakeholders were asked only for substantive factual corrections to the report. There were surprisingly few of these, so that the differences between the January and May editions of the ER were very minor, with no changes to conclusions or recommendations.

115. ICE team members were commissioned to give conference call briefings to the country networks on the ER during January 2015, but were not otherwise directly involved in the post-delivery discussion of the report. Apart from specific requests for amendments to the report, comments on it were not systematically shared with the ET, and the ET was not asked to respond to such comments. We have seen three sets of comments: (a) comments by the SMS (SMS 2015a); (b) a "consolidated network response" (SUN Networks 2015); (c) a "synthesis of responses" compiled by the SMS (SMS 2015b).

116. In retrospect, it is regrettable that the ET were not given the opportunity to respond systematically to all the comments on the ICE. A cursory review of the SMS comments, for example, shows (alongside much appreciation of the ICE work and many of ICE's conclusions) some very tendentious interpretations of the ICE analysis of the SUN M&E system: to take just one example, it is certainly not the case that "the outcome mapping methodology was not fully appreciated by the evaluators" (SMS 2015a p5). It will be unfortunate (as noted at ¶79 above) if SUN fails to take seriously the ICE critique of its M&E system and its approach to assessing the quality of country-level planning. More generally, one of the benefits of commissioning an independent evaluation should be to challenge understandable "groupthink" within partnerships: an unanticipated effect of the compressed ICE timetable, and the way subsequent report discussion was managed, was to lose some of the challenge function that is played by independent evaluators.

### **Management, independence and quality assurance**

117. There were some very strong features in the way the ICE was contracted and managed and in the quality assurance system adopted:

- a) The contracting of the evaluation by BMGF was rapid and efficient. BMGF fast-tracked the contracting process once Mokoro had been identified as the preferred bidder, the contractual relationship between BMGF and Mokoro was smooth, and transaction costs were minimised. BMGF did not use its role as the contracting agency to seek any special influence on the ICE.
- b) BMGF's detachment was assisted by the deployment of an Evaluation Manager (on the recommendation of the QAA). The EM's communications with the ET were excellent; he joined the ET's initial visit to Geneva, attended the inception workshop in Oxford and joined the country visit to Ethiopia as

an observer. The EM role reinforced the organisational independence of ICE vis-à-vis BMGF and the VSG, and the EM was extremely helpful as an interlocutor and buffer between the evaluation team, the VSG and the QAA.

- c) The evaluation's independence was safeguarded through the QAA mechanism, and the QAA inputs helped to strengthen the ICE ER. This was an extremely valuable part of the evaluation design, which was incorporated at the TOR stage. Global partnerships frequently do not have a built-in independent evaluation function (such as the evaluation departments of many multilateral and bilateral organisations), and conflict of interest is a serious risk when an evaluation management committee is comprised of people with strong vested interests in the evaluation. Mokoro had experienced this in its worst form during the mid-term evaluation of the Education for All Fast Track Initiative (see Cambridge Education, Mokoro & OPM 2010b and Cambridge Education, Mokoro & OPM 2010c). It is hard to overestimate the value of the QAA in guaranteeing genuine independence for the SUN ICE.

118. At the same time there were missed opportunities in the way the QAA chose to operate, which meant that their value added was significantly less than it could have been. This was especially frustrating because of the high calibre of the QAA panel members.

119. Mokoro's proposal made clear that we welcomed the QAA involvement, and indicated that we would seek to confer with them at an early stage in the evaluation. However, the QAA adopted an arms-length approach, which was explained to us as follows:

Finally, with regard to the invitation to meet, the QAA appreciate the request but must “reluctantly refuse” as such a meeting would create a potential conflict of interest for them in their role as independent quality assessors (post-hoc). They want to avoid the chance that they might influence (or be perceived to influence) the direction of the inception phase. This is one of the reasons that the QAA proposed recruitment of an independent evaluation manager, who can brief your team and also fill other important roles (e-mail from Ellen Piwoz (BMGF) on 13 June 2014).

120. The first task undertaken by the QAA was to score the various bids for the evaluation, and obviously this required complete detachment from the bidders. However, once an evaluation team was chosen, it is not at all obvious where the supposed conflict of interest lies: the ET's perspective is that the ET, the QAA and the EM all had a common interest in ensuring the best possible evaluation was delivered in the face of acknowledged constraints. On many aspects it would have been mutually beneficial for the QAA and ET to have advance discussions about how aspects of the evaluation could best be implemented (e-survey methodology is one example; how to reflecting a Theory of Change approach in the evaluation report is another). It was frustrating and inefficient to be subjected to post-hoc criticism that could have been avoided by ex ante discussion.

121. The QAA's own reflections are interesting: in the message cited at ¶119 above, not influencing the inception phase is a desideratum; but in retrospect they wish they had exerted more influence:

However with hindsight, both the evaluation team and the QAA could have done even more at the inception phase toward getting agreement on a reduced scope of work that would have given the team more time for analysis and consultation, and to address acknowledged key issues such as gender in more depth. (Compton et al 2015)

122. The QAA TOR were probably too narrowly framed in terms of reviewing the designated outputs of the evaluation:

The QAA will review the inception report, the interim report and the final report with regard to their adequacy, methodological rigor, application of good practice in comprehensive evaluations, soundness of evidence and independence. At each of these stages, they will provide brief advisory notes to the VSG. (QAA TOR ¶11)

123. There is no good reason why QAA should be limited to the evidence of the evaluation reports themselves in judging the quality of the evaluation. The QAA judgments would have been even more credible if they had (confidentially) accessed the interview notes compiled by the ET, and if they (like the EM) had taken the opportunity to accompany and observe one of the field visits. The arms-length approach thus limited the value of their reports to the client, as well as adding to stress and frustration on the part of the ET.

124. On each main deliverable, the QAA prepared a quality checklist and then commented on the deliverable against the criteria of the checklist. The ET provided a response matrix in each case, to accompany a revised draft. This process was extremely pressured because of delays in drafting and an approaching holiday season (see Table 3 above). Nevertheless, as documented in Box 14 below, we were surprised by the QAA approach to its comments on the first draft ER.

## Box 14 Comments on first draft ER – interchange between QAA and ET

**Overall QAA comment on first draft** - received in full on 20 Dec 2014:

We are using the Quality Checklist below but have not passed judgment on the quality of this first draft or specifically highlighted its positive points – this will be rectified in the use of the checklist on the next draft. The matrix below focuses on areas for improvement. We recognize and appreciate the tremendous amount of work that the evaluation team has put in, the extensive coverage of the report and the pressure imposed by the extremely tight timelines. Nevertheless, we would not be doing our QAA job properly if we did not point out areas that we believe need strengthening to comply with international evaluation standards. Key attributes which commonly distinguish a (theory based) ‘evaluation’ from a ‘review’ include a systematic approach to testing the Theory of Change and its assumptions, the use of explicit performance standards and analytical frameworks, a clear evidence trail and systematic scrutiny of possible alternative explanations. This first draft requires strengthening in several of these areas, as noted below.

(Note: Please excuse us if we wrongly said something is missing from the report when it’s not. Timelines have been tight for us as well.)

### Response from ICE team leader

Thank you for your comments. Many were very helpful; taking account of them has made the report stronger, and I am genuinely grateful for that. At the same time, while I do recognise the time pressures on you, I was taken aback by this explanation of your approach. Not so much because you commented only on how bad the report was and not how good, but because in deferring a judgment on the report’s quality you sidestepped what is for me, the most important criterion in your own TOR, of which ¶11 reads:

*The QAA will review [the reports] with regard to their **adequacy**, methodological rigor, application of good practice in comprehensive evaluations, soundness of evidence and independence.* [emphasis added]

As evaluation team leader I have to try to deliver the best report I can with limited resources and on a timescale that the authors of our TOR regarded as much too short. Inevitably this means compromises on all the criteria on which you are asked to judge the report, except independence. Adequacy is the criterion under which I would expect you to assess whether we achieved a reasonable balance in what we managed to do in the time available, and whether and to what extent the VSG and others can regard the result as a credible report (so it’s probably the most important criterion for them too). More particularly, I was hoping your comments on adequacy would be specific enough to help me judge (in the context of more compromises) which were the most important things I should spend scarce remaining time on to make the report as useful as possible to the client.

Anyway, you will see below that we did something about most of your points, but that there are two areas – evaluation matrix and interview evidence –where I suspect we have disappointed you. ....

125. It should not be regarded as good practice for QAA to provide an undifferentiated shopping list of potential improvements to a report without also providing an overall assessment of the adequacy of the report and indicating in which specific respects it is judged to fall short of an acceptable threshold.

126. **Lessons on process:** (a) *The compressed timetable for the ICE had foreseeable consequences in limiting the evaluation team’s consultations with stakeholders around its conclusions and recommendations. An unforeseen consequence was to lower the quality of the post-report debate, because the evaluation team was not given the opportunity to defend the report’s conclusions*

*and recommendations. (b) The arrangements for contracting and independent quality assurance were crucial in safeguarding the independence of the evaluation. (c) The quality assurance arrangements also contributed to strengthening the evaluation, but would have contributed more if the QAA had relied less on ex post review of drafts and engaged more fully with the evaluation team throughout the process.*

## **6. Recommendations**

127. In the judgement of external assessors, the ICE delivered a credible and useful report in spite of acknowledged constraints of timing and resources. Nevertheless, as this NOAM has documented, these constraints had adverse effects on the process, both during the conduct of the ICE itself and in the subsequent stakeholder discussion of its findings.

128. We assume that the SUN's next phase (SUN 2.0) will itself require a thorough evaluation prior to any decision about further continuation of SUN.<sup>22</sup> Given SUN's iterative evolution, the impractical timetable for ICE1 may be regarded as merely unfortunate; SUN's leadership would have no excuse for a similarly compressed timetable for ICE2.

129. It is right to point out (as the QAA have done) that more care may be needed at TOR stage to ensure a proper fit between the three key dimensions of any ICE: the scope of its TOR, the time allowed for conducting it, and the budget. But it is dangerous to put all the emphasis on the TOR dimension, because it runs the risk of eating up more time on discussion of TOR and further compressing the time available for the evaluation itself. The primary challenge for the SUN leadership is to commit firmly to a realistic timetable for ICE2, with commensurate commitment of funding, and to ensure that inevitable debates about the TOR are not allowed to frustrate this timetable.

130. Some of the design criteria should be:

- a) The contract for ICE2 should be let not later than 2 years before the end of the SUN 2.0 period. If necessary there should be scope for modification of the overall TOR and the TOR for component studies by negotiation after the contract is let.
- b) The evaluation plan for SUN 2.0 should provide for component studies to be completed ahead of the main overall evaluation. These would certainly include full country case studies (to involve deeper studies than were feasible for ICE1, complemented by systematic consideration of the lessons for SUN that may be

---

<sup>22</sup> The ER recommended:

The SUN 2.0 strategy should have a five-year horizon, with provision to complete a thorough independent review by the end of the fourth year so as to guide any subsequent phase.

drawn from non-SUN countries), plus other thematic studies (e.g. of nutrition architecture and lessons from other partnerships, gender, DBM) that would provide a stronger platform for the main evaluation.

- c) It should strengthen the use of QAA by (a) engaging them earlier (while taking care that their involvement does not serve to elongate the pre-contracting process); (b) making it clear that the QAA are expected to engage fully with the evaluation team (after it has been competitively selected) and not rely solely on post-hoc review of draft reports in assessing the adequacy of an evaluation.

## Bibliography

"Location" in the listing below refers to folder and document numbers in the evaluation team's electronic library.

short ref	full ref	location
Bacchetti 2002	<i>Peer review of statistics in medical research: the other problem.</i> British Medical Journal, 234, 1271–1273. P. Bacchetti 2002	
Balogun 2011	<i>Synthesis Note of Comprehensive Evaluations for International Institutions.</i> Paul Balogun, prepared on behalf of the Global Environment Facility Evaluations Office (GEFEO), 16 August 2011.	4.2-2
Berg 1987	<i>Nutrition planning is alive and well, thank you.</i> In <i>Food Policy</i> 12 (4): 365– 67. A Berg, 1987.	
Bezanson & Isenman 2012	<i>Governance of New Global Partnerships: Challenges, Weaknesses and Lessons.</i> Keith A. Bezanson and Paul Isenman. CGD Policy Paper 014. Center for Global Development, October 2012.	7.1-3
Bezanson et al. 2014	<i>Terms of Reference for the Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement.</i> Keith Bezanson, Lola Gostelow and Paul Isenman, March 2014.	
Cambridge Education, Mokoro & OPM 2010a	<i>Final Synthesis Report: Volumes 1–5. Mid-Term Evaluation of the EFA Fast Track Initiative.</i> Cambridge Education, Mokoro and OPM, February 2010.	t7.1.3-1-5
Cambridge Education, Mokoro & OPM 2010b	<i>Mid-Term Evaluation of the EFA Fast Track Initiative, Final Synthesis Report Volume 4 – Appendix V: Note on Approach and Methods.</i> Cambridge Education, Mokoro and OPM, July 2010.	
Cambridge Education, Mokoro & OPM 2010c	<i>Mid-Term Evaluation of the EFA Fast Track Initiative, Final Synthesis Report Volume 4 – Appendix V: Note on Approach and Methods – Disclaimer and Response.</i> Cambridge Education, Mokoro and OPM, July 2010.	
Carter 2012	<i>Theory-based evaluation approach.</i> GSDRC Helpdesk Research Report. Becky Carter. Birmingham, UK: Governance and Social Development Resource Centre, University of Birmingham, December 2012.	4.6-2
CGIAR 2013	<i>CGIAR Standards for Independent External Evaluation. Annexes. October 2013 draft.</i> CGIAR, October 2013	t2.4-1
Clark & Anderson 2004	<i>Theories of Change and Logical Models: telling them apart.</i> Presentation at American Evaluation Association. H. Clark and A. Anderson. Atlanta, Georgia, November 2004	
Compton et al 2015	<i>Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement, 2014-15: Statement by the Evaluation Quality Assurance Advisors,</i> Julia Compton, Chris Gerrard, and Eva Lithman May 1, 2015.	

short ref	full ref	location
Copenhagen Consensus 2012	<i>Copenhagen Consensus 2012: Expert Panel Findings</i> . Summary of findings of expert panel asked to set priorities among a series of proposals for confronting ten of the world's most important challenges.	5-5
Della Porta & Diani 2006	<i>Social Movements: an introduction</i> . Blackwell publishing. Second edition. 2006	7-1
de Mel 2015	<i>On the process of the Evaluation and the Lessons Learned: Evaluation Manager Report</i> (draft 29.01.15), Ruwan de Mel	
Fayers 2011	<i>Alphas, betas and skewy distributions: two ways of getting the wrong answer</i> . <i>Advances in Health Sciences Education: Theory and Practice</i> . P Fayers 2011, 16:291–296.	
Field 1987	<i>Multisectoral nutrition planning: a post-mortem</i> , in <i>Food Policy</i> 12 (1): 15–28. John Osgood Field, Butterworth & Co Ltd, 1987.	5-18
Garrett & Natalicchio 2011	<i>Conclusion: Approaches to Working Multisectorally in Nutrition</i> in Garrett and Natalicchio (eds.), <i>Working Multisectorally in Nutrition – Principles, Practices and Case Studies</i> . J Garrett and M Natalicchio. IFPRI, 2011	
Garrett et al. 2011	<i>Rationale and Historical Perspectives</i> , in Garrett and Natalicchio (eds.) <i>Working Multisectorally in Nutrition – Principles, Practices, and Case Studies</i> , pp 20–47. J Garrett, L Bassett and J Levinson, IFPRI, 2011.	
GEF EO 2012	<i>Learning lessons from comprehensive evaluations of international institutions</i> . Briefing note. GEF Evaluation Office, 15 November 2012.	4.2-1
Hammer 1997	<i>Economic Analysis for Health Projects</i> in 'The World Bank Research Observer' Volume 12 (1), February 1997.	4.3-2
IASO 2014	<i>The prevention of obesity and NCDs: challenges and opportunities for governments</i> . IASO Policy Briefing. IASO, January 2014	5.4-1
IDD & Associates 2007	<i>Joint Evaluation of General Budget Support 1994–2004: Note on Approach and Methods</i> . IDD and Associates. Glasgow: DFID, February 2007.	t7.1.1.1-2
IEG 2007	<i>Sourcebook for Evaluating Global and Regional Partnership Programs: Indicative Principles and Standards</i> . Independent Evaluation Group. Washington, DC: World Bank, 2007.	4.2-0
IEG 2011	<i>The World Bank's Involvement in Global and Regional Partnership Programs. An Independent Assessment</i> . Independent Evaluation Group. Washington, DC: World Bank, 2011.	7.1-4 t2.1.1-4
IFPRI 2014	<i>Global nutrition report 2014: actions and accountability to accelerate the world's progress on nutrition</i> . Washington DC.	
Isenman 2012a	<i>Improving assessments of effectiveness of multilateral organizations</i> . SADEV Report 2012:3. Paul Isenman. SADEV, August 2012.	4.2-3
Isenman 2012b	<i>Improving the Quality and Impact of Comprehensive Evaluations of Multilateral Organizations</i> , Paul Isenman. Comprehensive Evaluation Platform for Knowledge Exchange (CEPKE), December 2012	4.2-4

short ref	full ref	location
Isenman et al 2011	<i>Stewardship of the SUN Movement: Taking SUN to the Next Level.</i> Paul Isenman, Keith Bezanson and Lola Gostelow, 30 September 2011.	0.3-4
Jones 2011	<i>A guide to monitoring and evaluating policy influence, Background Note.</i> Harry Jones. Overseas Development Institute, 2011.	4.4-1
Kuzon et al 1996	<i>The seven deadly sins of statistical analysis.</i> Annals of Plastic Surgery, 37, 265–272. Kuzon, W. M., Urbanchek, M. G., & McCabe, S. (1996)	
Mokoro 2008a	<i>Putting Aid On Budget: Synthesis Report.</i> A Study for the Collaborative Africa Budget Reform Initiative (CABRI) and the Strategic Partnership with Africa (SPA). Stephen Lister, Mokoro Ltd. CABRI, April 2008.	t7.1.1.4-1
Mokoro 2008b	<i>Good Practice Note: Using Country Budget Systems.</i> A Study for the Collaborative Africa Budget Reform Initiative (CABRI) and the Strategic Partnership with Africa (SPA). Mokoro Ltd. CABRI, June 2008.	
Mokoro 2011	<i>WFP's School Feeding Policy: A Policy Evaluation.</i> Stephen Lister, Stephen Anderson, Mirella Mokbel Genequand, Anthea Gordon, Judith Sandford, Stephen Turner. Mokoro Ltd. Report number: OE/2012/002. Rome: WFP Office of Evaluation, 30 November 2011.	t7.1.1.8-1
Mokoro 2012	<i>Study on better reflecting aid flows in country budgets to improve aid transparency and public financial management.</i> Alta Fölscher, Rebecca Carter, Samuel Moon, Gareth Graham and Frédéric Jeanjean, Oxford: Mokoro, 27 August 2012.	t7.1.3-3
Mokoro 2014a	<i>An evaluation of UNICEF's upstream work in basic education and gender equality (2003–2012). Evaluation Synthesis Report: Penultimate Draft.</i> Mokoro Ltd. UNICEF, March 2014.	t7.1.1.6-1
Mokoro 2014b	<i>Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement: Inception Report.</i> Oxford: Mokoro Ltd, 01 August 2014.	
Mokoro 2014c	<i>Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement: Interim Progress Report.</i> Oxford: Mokoro Ltd, 03 October 2014.	
Mokoro 2014d	<i>Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement: Discussion paper: future options for the SUN movement.</i> Oxford: Mokoro Ltd, 04 November 2014.	
Mokoro 2014e	<i>Independent Comprehensive Evaluation of the Scaling Up Nutrition Movement: Final Report – Main Report and Annexes.</i> Oxford: Mokoro Ltd, 15 January 2015.	
MQSUN 2014	<i>MQSUN Briefing.</i> PowerPoint presentation. Undated, possibly 2014.	8-9
N4G 2013a	<i>Nutrition for Growth Commitments: Executive Summary.</i> No date. [2013]	6.6-1, 8-11
Nisbett et al 2014a	<i>Why Worry About the Politics of Childhood Undernutrition?</i> Nicholas Nisbett, Stuart Gillespie, Lawrence Haddad and Jody Harris, in <i>World Development</i> , Vol. 64, pp420–433, 2014.	5-29
Norman 2010	<i>Likert scales, levels of measurement and the “laws” of statistics.</i> Advances in Health Sciences Education, 15, 625–632.2, 2010	

<b>short ref</b>	<b>full ref</b>	<b>location</b>
ODI & Mokoro 2010a	<i>Sector Budget Support in Practice: Study Methodology.</i> ODI and Mokoro, March 2009.	t7.1.1.2-2
ODI & Mokoro 2010b	<i>Sector Budget Support in Practice: Synthesis Report.</i> Tim Williamson and Catherine Dom, ODI and Mokoro, February 2010	t7.1.1.2-1
REACH 2013	<i>A multi-sectoral perspective to nutrition: From analysis to action.</i> Presentation to SUN Annual Gathering, New York, September 2013	6.3-3
Renard & Lister 2013	<i>Technical Note on Efficiency Analysis.</i> Mokoro: Robrecht Renard and Stephen Lister. WFP OEI, 14 June 2013.	t7.1.2-1
Renard 2013	<i>WFP Efficiency Analysis Guidance Materials – Definitions, Concepts and Methodology.</i> Mokoro: Robrecht Renard, 30 March 2013	t7.1.2-2
Shrimpton & Rokx 2012	<i>The Double Burden of Malnutrition – A Review of Global Evidence,</i> Roger Shrimpton and Claudia Rokx, World Bank, HNP Discussion Paper, November 212	5.4-3
SMS 2012a	<i>Scaling Up Nutrition (SUN) Movement Strategy [2012-2015].</i> September 2012.	0.0-3
SMS 2012b	<i>SUN Movement: Revised Road Map.</i> Secretariat of the Scaling Up Nutrition Movement. September 2012.	0.0-2
SMS 2013a	<i>SUN Movement Monitoring &amp; Evaluation Framework. Final Draft.</i> SUN Movement Secretariat, 10 April 2013.	0.6-1
SMS 2014a	<i>SUN Movement Secretariat's contractual obligations to donors on evaluations/reviews.</i> SUN Movement Secretariat, March 2014.	0.3.2-8
SMS 2015a	<i>SUN Movement Secretariat Response to the Independent Comprehensive Evaluation (ICE),</i> SMS (apparently 27 March 2015)	
SMS 2015b	<i>Synthesis of responses to the SUN Movement's Independent Comprehensive Evaluation,</i> SUN Movement Secretariat March 2015	
SUN 2010a	<i>Scaling Up Nutrition: A Framework for Action.</i> 2010 (Reprint April 2011.)	0.0-1
SUN Networks 2015	<i>SUN Movement - Compilation of the Networks response to the ICE Final Report</i> [apparently 10 March 2015]	
SUN Road Map Task Team 2010a	<i>A Road Map for Scaling-Up Nutrition (SUN).</i> First Edition. Scaling Up Nutrition Road Map Task Team, September 2010.	0.0-5
The Lancet 2008	<i>The Lancet's series on Maternal and Child Undernutrition: Executive Summary.</i> The Lancet, 2008.	5.3-1
The Lancet 2013	<i>Maternal and Child Nutrition.</i> Executive Summary of <i>The Lancet</i> Maternal and Child Nutrition Series. The Lancet, 2013.	5.2-2
Vogel 2012	<i>Review of the use of 'Theory of Change' in international development.</i> Isabel Vogel. Review Report for DFID, April 2012.	4.6-1
Woolcock 2013	<i>Using case studies to explore the external validity of 'complex' development interventions.</i> WIDER Working Paper No. 2013/096. M. Woolcock. United Nations University, UNU-WIDER, 2013.	

## **Abbreviations**

BMGF	The Bill and Melinda Gates Foundation
CCS	Country Case Study
CE	Comprehensive Evaluation
CEPKE	Comprehensive Evaluation Platform for Knowledge Exchange
CGIAR	Consultative Group on International Agricultural Research
CHAI	Clinton Health Access Initiative
CRF	Common Results Framework
CSN	Civil Society Network
CSO	Civil Society Organisation
DAC	Development Assistance Committee (of the OECD)
DBM	Double Burden of Malnutrition
EC	European Commission
EFA	Education For All
EM	Evaluation Manager
EQ	Evaluation Question
ER	Evaluation Report
ET	Evaluation Team
EU	European Union
FTI	Fast Track Initiative (Education for All)
GEF	Global Environment Facility
GNR	Global Nutrition Review
GOI	Government of Indonesia
GPE	Global Partnership for Education (previously FTI)
GPR	Global Program Report
ICE	Independent Comprehensive Evaluation
ICN	International Conference on Nutrition
IEG	Independent Evaluation Group (of the World Bank)
IPR	Interim Progress Report
IR	Inception Report
LG	Lead Group
M&E	Monitoring and Evaluation
MOU	Memorandum of Understanding
MPTF	Multi-Partner Trust Fund
N4G	Nutrition for Growth
NGO	Non-Governmental Organisation
NOAM	Note on Approach and Methods
NPAN	National Plan of Action for Nutrition
OECD	Organisation for Economic Co-operation and Development
PA	Personal Assistant
QAA	Quality Assurance Advisor

QS	Quality Support
SBN	SUN Business Network
SCN	(UN) Standing Committee on Nutrition
SMS	The Sun Movement Secretariat
SUN	The Scaling Up Nutrition movement
SWAp	Sector Wide Approach
TL	Team Leader
ToC	Theory of Change
TOR	Terms of Reference
UN	United Nations
UNDP	United Nations Development Program
UNEG	United Nations Evaluation Group
VSG	Visioning Sub-Group
USD	US Dollar