

Evaluating the design and implementation fidelity of an adapted Plan-Do-Study-Act approach to improve health system performance in a Nigerian state

*Ejemai Amaize Eboreime^{1,2}, John Olajide Olawepo^{3,4}, Aduragbemi Banke-Thomas⁵, Rohit

Ramaswamy⁶

*Corresponding Author: eboreime@ualberta.ca

Authors' Affiliations:

1. Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada
2. Department of Planning, Research and Statistics, National Primary Healthcare Development Agency, Abuja, Nigeria
3. Department of Environmental and Occupational Health, School of Public Health, University of Nevada Las Vegas, USA
4. Department of Health Sciences, Bouve College of Health Sciences, Northeastern University, Boston, USA
5. Department of Health Policy, London School of Economics and Political Science, London, UK
6. Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill NC, USA

Abstract

Background: The Plan-Do-Study-Act (PDSA) cycle is fundamental to many quality improvement (QI) models. For the approach to be effective in the real-world, variants must align with standard elements of the PDSA. This study evaluates the alignment between theory, design and implementation fidelity of a PDSA variant adapted for Nigeria's health system performance improvement.

Methods: An iterative consensus building approach was used to develop a scorecard evaluating new conceptual indices of design and implementation fidelity of QI interventions (design and implementation index, defects and gaps) based on Taylor's theoretical framework.

Results: Design (adaptation) scores were optimal across all standard features indicating that design was well adapted to the typical PDSA. Conversely, implementation fidelity scores were only optimal with two standard features: prediction-based test of change and the use of data over time. The

other features, use of multiple iterative cycles and documentation had implementation gaps of 17% and 50% respectively.

Conclusion: This study demonstrates how both adaptation and implementation fidelity are important for success of QI interventions. It also presents an approach for evaluating other QI models using Taylor’s PDSA assessment framework as a guide, which might serve to strengthen the theory behind future QI models and provide guidance on their appropriate use.

Keywords: Quality Improvement; Implementation Science, PDSA, Health System, Nigeria, program design

Introduction

The Plan-Do-Study-Act (PDSA) cycle (and its concept of iterative tests of change) is fundamental to both the fields of Quality Improvement (QI) and Implementation Science. Whereas Implementation Science is primarily concerned with promoting the “integration of evidence-based interventions into health care policy and practice”, QI science concentrates on “systematically and rigorously exploring ‘what works’ to improve quality in healthcare” (Health Foundation, 2011). Both disciplines place an emphasis on improving care using a systems-level approach. Both also overlap by considering concepts such as context, the organizational and system actors (providers, staff, patients, and administrators), processes and outcomes (Williams, 2019).

Many popular QI models such as the Model for Improvement, Lean and Six Sigma (Schroeder *et al.*, 2008; Norman *et al.*, 2013; Taylor *et al.*, 2013; Reed and Card, 2016) incorporate some version of PDSA in their methodologies to test improvement solutions on a small scale before scaling up. In the field of Implementation Science, PDSA is similarly used to adapt implementation strategies to local contexts (Fixsen *et al.*, 2005; Powell *et al.*, 2015). Whereas Implementation Science considers the

PDSA as a method or strategy, the approach is viewed as an intervention in the science of improvement (Koczwara *et al.*, 2018). A recently proposed model integrating QI and implementation science, called the Model for Improvement and Implementation (Ramaswamy, Johnson and Hirschhorn, 2018) explicitly situates PDSA as a process at the intersection of both fields, necessary for both context specific intervention development (QI) and adaptation of IS strategies.

Moreover, PDSA cycles align well with models for decision-making in complex systems with non-linear processes that are unpredictable and that display emergent behaviour (WHO, 2009). In such systems, a “probe, sense, respond” approach is recommended, where decision makers initiate small experiments to test how the system behaves before engaging in large scale analysis and intervention (Snowden and Boone, 2007). As organizations such the World Health Organization and the World Bank recommend QI to address larger scale problems such as health systems strengthening to improve global quality of care (Kieny *et al.*, 2018), the use of PDSA is likely to continue to expand.

The use of PDSA to find local solutions to solve complex improvement and implementation problems offers great promise (Muczyk, 2004; Sales *et al.*, 2006), and this approach can be considered to be a best practice in developing quality improvement interventions. However, even though PDSA provides a rigorous experimental approach to improvement, it has not been always been used effectively to bring about outcomes (Reed and Card, 2016). The reason for this is there is no standard protocol or formula for using PDSAs. This in itself is not a shortcoming – by definition, since PDSA cycles are intended to drive learning based on generating hypotheses about how to change a system, a PDSA approach should be deeply embedded in the systems in which implementation and improvement work takes place and its particular application should depend on the context of the system as several researchers have pointed out (Øvretveit, 2011; Leis and Shojanian, 2017; Ramaswamy *et al.*, 2018). However, this assumes that those engaged in improvement or implementation adhere to the primary principles of the PDSA such as developing change hypothesis, designing multiple tests of increasing complexity, collecting and analysing data and making

appropriate updates to the change hypothesis. As demonstrated by Taylor and colleagues (Taylor *et al.*, 2013) who applied an evaluation framework to assess gaps in the use of PDSA cycles to improve health care, this is not often the case. Despite the theoretical promise of PDSA, it often fails in execution.

That being said, the PDSA approach still remains the best theoretical model for designing quality improvement interventions, and therefore it is appropriate to test other models designed to improve quality, against the PDSA approach as a standard. While any quality improvement approach needs to be implemented well, there is an additional barrier to success if the theory underlying the approach is untested or flawed. At the same time, as mentioned above, PDSA is not an inviolable formula, and therefore needs to be adapted for use in various settings. In this paper, we evaluate a quality improvement model called Diagnose-Intervene-Verify-Adjust (DIVA), adopted in Nigeria and other low and middle-income countries (LMICs) to improve planning processes in health systems.

DIVA was developed, introduced, tested and implemented in several LMICs by United Nations Children's Fund (UNICEF) (UNICEF, 2012). The model seeks to improve processes such as routine planning cycles and the use of data by local stakeholders and policy/decision makers as suggested by the Lancet Global Health Commission on High Quality Health Systems (HQSS) (UNICEF, 2012; O'Connell and Sharkey, 2013). Typically, operational planning occurs annually. DIVA strengthens the annual planning cycles by introducing quarterly iterations for monitoring and improvement of the annual plan. Further, outputs from DIVA (such as proposed implementation strategies for health interventions) feed into subsequent operational planning cycles. DIVA also builds the capacity of district/local government health teams, communities and other local stakeholders to identify bottlenecks and barriers to the planning process. The model provides tools to address these barriers, take action and be accountable for service delivery within their jurisdiction towards improving effective coverage, quality and impact of health interventions (UNICEF, 2012; O'Connell and Sharkey, 2013).

Operationally, DIVA is a four-step model in which ‘Diagnose’ identifies bottlenecks to health interventions; ‘Intervene’ develops and implements action plans to resolve identified constraints; while ‘Verify/Adjust’ tracks the results of the action plans and recommends real-time data driven adjustments to the plan (Figure 1). DIVA uniquely doubles as a Quality Improvement method as well as an implementation strategy for evidence informed planning of interventions in complex health systems. The model was developed in response to the observation that district health planning in many LMICs is commonly fragmented across various actors and vertical programmes, and often not informed by contextual evidence. The theory of change behind the model is that evidence-informed bottleneck analysis and integrated intervention planning using DIVA, will improve efficiency of resource allocation and utilization as well as health system performance indices (UNICEF, 2012, 2016).

Research, published elsewhere (Eboreime, Nxumalo, *et al.*, 2019) showed that the DIVA model can potentially improve health system performance, but that effective implementation is critical. Other evaluations in Nigeria, Ghana and Uganda have also demonstrated that the tools and methods of DIVA are effective in evidence-informed planning of health interventions (O’Connell and Sharkey, 2013; Baker, Okuga, *et al.*, 2015; Baker, Peterson, *et al.*, 2015; Henriksson *et al.*, 2017; Eboreime, Nxumalo, *et al.*, 2018) but outcomes have exhibited significant variability (Eboreime, Eyles, *et al.*, 2018; Eboreime, Nxumalo, *et al.*, 2018). To provide guidance to countries seeking to enhance the effectiveness of DIVA, it would be useful to assess both the theoretical underpinning and the implementation fidelity of DIVA. This will help practitioners and policy/decision makers evaluate whether the failure of DIVA to achieve outcomes results from a failure in its design (i.e. the underlying premise of DIVA is flawed) or whether DIVA is a context-appropriate adaptation of the principles of PDSA, and failures in its performance can be attributed to poor implementation fidelity. If the latter is the case, strategies to improve the implementation of DIVA can be informed by the substantial body of knowledge that already exists about how to conduct PDSAs well. Specifically, in this paper, we seek to answer the following research questions:

1. To what extent can DIVA be considered to be an adaptation of PDSA?
2. To what extent does the implementation of DIVA in Kaduna state, Nigeria exhibit fidelity to its design?

Materials and Methods

Study context

Nigeria operates a decentralized three-tier health system governance structure. The federal level is responsible for overarching policy thrusts including programme design, the state ministries of health are largely responsible for subnational coordination and translation, while the local governments are mostly the implementers (FMOH, 2016; E. A. Eboreime *et al.*, 2017). DIVA was introduced nationwide in 2012, most states experienced implementation failure with only Kaduna sustaining implementation as at the time of this study (Eboreime, Eyles, *et al.*, 2018; Eboreime, Nxumalo, *et al.*, 2018).

Theoretical Framework

To answer the first research question, we used the theoretical framework developed by Taylor and colleagues (3) mentioned earlier to assess the quality of implementation of PDSA cycles in 47 studies in healthcare settings. This framework uses five features to test the theory and the practice of the approach. These features are the use of Iterative cycles; prediction-based test of change, small-scale testing; use of data over time, and documentation of processes and outcomes for learning (Table 1). Given the use of the DIVA implementation for adjusting the annual plan on a quarterly basis, one of these characteristics (small scale testing) was not relevant for DIVA, because the steps of the DIVA process were applied to the same plan every quarter. We adapted this framework first to evaluate whether the design guidelines for DIVA include these key features. We used our findings from this evaluation to develop the assessment questions to evaluate the fidelity of implementation.

Table 1: Taylor's theoretical framework based on key features of the plan-do-study-act (PDSA) cycle (Source: (Taylor *et al.*, 2013))

Feature of PDSA	Description of feature
-----------------	------------------------

Iterative cycles	To achieve an iterative approach, multiple PDSA cycles must occur. Lessons learned from one cycle link and inform cycles that follow. Depending on the knowledge gained from a PDSA cycle, the following cycle may seek to modify, expand, adopt or abandon a change that was tested.
Prediction-based test of change	A prediction of the outcome of a change is developed in the 'plan' stage of a cycle. This change is then tested and examined by comparison of results with the prediction.
Small-scale testing	As certainty of success of a test of change is not guaranteed, PDSAs start small in scale and build in scale as confidence grows. This allows the change to be adapted according to feedback, minimises risk and facilitates rapid change and learning.
Use of data over time	Data over time increases understanding regarding the variation inherent in a complex healthcare system. Use of data over time is necessary to understand the impact of a change on the process or outcome of interest.
Documentation	Documentation is crucial to support local learning and transferability of learning to other settings.

Data collection

Data for the overall implementation research of which this work is a component were obtained from content analyses of 39 policy documents, 15 in-depth interviews and an embedded process observation between 2012 and 2016 to identify the design and implementation processes of DIVA in Nigeria. The documents were identified from the researchers' knowledge (one author was an embedded researcher) and interviews of four national level subject matter experts who contributed to the initial design and implementation processes. The documents include policy guidelines, a DIVA implementation guide and implementation reports obtained at both the design phase (national) and implementation phases (subnational). This sub-study uses data related to the implementation of DIVA in Kaduna state because this is the only state still sustaining the implementation of DIVA since its design and deployment in 2012 (Eboreime, Eyles, *et al.*, 2018; Eboreime, Nxumalo, *et al.*, 2018).

Analytic parameters

We developed the following indicators to perform our evaluation:

1. *Adaptation score/index*: This is a measure of the extent to which the steps of DIVA is consistent with the applicable elements of PDSA. It assesses the conceptual similarity between the standard improvement model and its local adaptation.
2. *Implementation score/index*: This is a measure of the extent to which the adapted model was implemented in the real world. The implementation index may be considered to be a measure of adherence to how DIVA was intended to be used, which is a dimension of implementation fidelity (James Bell Associates, 2009).

Figure 2 is an illustration of these concepts.

Taylor’s framework used 13 questions to assess the implementation fidelity of PDSAs, which are paraphrased in Table 2. To apply this framework to the design of DIVA, an adapted guide was created. The design evaluation helped to clarify the key features of the DIVA model and the extent to which these features were aligned with the key steps of a PDSA approach. Based on the results, an implementation fidelity evaluation questionnaire was developed. This questionnaire is shown in Table 3.

Table 2: Taylor’s PDSA assessment questions

Feature of PDSA	Assessment questions
Iterative cycles	<ul style="list-style-type: none"> • Were multiple cycles used? • Were multiple cycles linked to one another (i.e., does the ‘act’ stage of one cycle inform the ‘plan’ stage of the cycle that follows)? • When isolated cycles were used were future actions postulated in the ‘act’ stage?
Prediction-based test of change	<ul style="list-style-type: none"> • Was a change tested? • Was an explicit prediction articulated?
Small-scale testing	<ul style="list-style-type: none"> • Sample size per cycle • Temporal duration of cycles • Number of changes tested per cycle
Use of data over time	<ul style="list-style-type: none"> • Did sequential cycles increase scale of testing? • Was data collected over time? • Were statistics used to test the effect of changes and/or understand variation?

Documentation

- How thoroughly was the application of the PDSA method detailed in the reports?
 - Was each stage of the PDSA cycles documented?
-

Table 3: Data analysis tool

Feature of PDSA	Analysis criteria		Response scale for each question (Scores: 0/1/2)
	Design evaluation	Implementation evaluation	
Iterative cycles	<ul style="list-style-type: none"> Does the DIVA model incorporate the principle of multiple cycles? Are multiple cycles linked to one another (i.e., does the 'adjust' stage of one cycle inform the 'diagnose' stage of the cycle that follows)? 	<ul style="list-style-type: none"> Were multiple cycles used? Were multiple cycles linked to one another (i.e., does the 'adjust' stage of one cycle inform the 'diagnose' stage of the cycle that follows)? 	No/partly/yes
Prediction-based test of change	<ul style="list-style-type: none"> Does DIVA depend on well-defined theory of change? Does the use of the model require explicit predictions that are tested in the "intervene" step? 	<ul style="list-style-type: none"> Was a theory of change articulated prior to the implementation? Was an explicit prediction made in each step? 	No/partly/yes
Small-scale testing	<ul style="list-style-type: none"> Sample size per cycle Temporal duration of cycles Number of changes tested per cycle 	<ul style="list-style-type: none"> Not applicable 	Decrease/ same/ increase
Use of data over time	<ul style="list-style-type: none"> Does the DIVA model require the collection and display of data over time? 	<ul style="list-style-type: none"> Was data collected over time? 	No/partly/yes
Documentation	<ul style="list-style-type: none"> Does the design require detailed reporting of the application of the DIVA model in each quarter? Does the design require detailed reporting of each step of the DIVA process in each quarter? 	<ul style="list-style-type: none"> Was the application of the DIVA method detailed in the reports? Was each stage of the DIVA cycle(s) documented? 	No/partly/yes

1 Data analysis

2 To assess the level of alignment between the data and standard features, we created a scorecard.
3 The scorecard assigned quantitative scores that measured the extent to which the reported or
4 documented design or implementation activity aligned with the theoretical framework. An iterative
5 consensus building approach was used to discuss and finalize scores. Fowles' Delphi process guided
6 the analysis (Van Wyk, 1980). This process mirrors the approach used in evaluating the deployment
7 and installation processes of DIVA in the other studies (Eboeime, Eyles, *et al.*, 2018; Eboeime,
8 Nxumalo, *et al.*, 2018).

9 A team of three PDSA/ DIVA subject matter experts was constituted and a process facilitator was
10 appointed. The first round of questionnaires was developed using the questions from Taylor's
11 framework. These questions were tested and adapted to the study context.

12 Each question in the design and implementation was evaluated on a 3-point ordinal scale (No = 0,
13 Partly = 1, Yes = 2). 'No' and 'Yes' were attributed to absolute non-compliance and absolute
14 compliance respectively, whereas a 'partly' was assigned to partial compliance. The three evaluators
15 with expertise and knowledge of the PDSA and DIVA independently reviewed the qualitative data
16 and assigned a score to each question. Inter-rater variances were resolved using a consensus
17 building discussions. The rules were that disparities between each rater's scores on any question
18 must not exceed one point. Thus, any rating disparity greater than one point was discussed and re-
19 evaluated in the next round. The underlying assumption was that a one-point difference may
20 represent random variation which could be harmonized by calculating a mean. However, a two-point
21 disparity may be indicative of a significant interpreter difference which may need more clarification.
22 The mean was used as the final score for each question. Each question carried an equal weight per
23 standard feature. The mean score for each feature was computed and composite scores for each
24 feature were created by dividing by a factor of 2 expressed in percentage values ranging from 0%
25 (Zero implementation) to 100% (full implementation/full design alignment).

26 Results

27 Table 4 presents the scores for design and implementation fidelity of DIVA on each evaluation
 28 criterion. Our design evaluation indicated that all the applicable characteristics of PDSA were part of
 29 the DIVA design, and therefore DIVA can be considered an adaptation of the PDSA model. We could
 30 therefore use the same questions we used to evaluate the design for the assessment of the
 31 implementation fidelity. The design of DIVA was assessed as being 100% compliant with three out of
 32 five PDSA features (Iterative cycles, Prediction-based test of change, Small scale tests of change, Use
 33 of data over time). However, documentation was assessed to be 92% compliant and small-scale tests
 34 of change were omitted from the DIVA design. On the other hand, Table 5 also shows the gaps on
 35 each respective feature. Both tables highlight where design failure and/or implementation failure
 36 may exist, which have implications on the effectiveness of the intervention.

37

38 *Table 4: Scores on analyses criteria*

PDSA Feature	Analyses criteria	Response score (mean) for each question	
		<i>Design</i> (Adaptation Score)	<i>Implementation</i> (Fidelity score)
Iterative cycles	Multiple cycles	2.00	2.00
	Cycles linked to one another	2.00	1.33
	Composite score (%)	100	83
Prediction-based test of change	Test of change	2	2
	Explicit prediction articulated	2	2
	Composite score (%)	100	100
Small scale tests of change	Sample size per cycle	0	Not applicable
	Temporal duration of cycles		
	Number of changes tested per cycle		
	Composite score (%)	0	Not applicable
Use of data over time	Data collected over time	2	2
	Composite score (%)	100	100
Documentation	Application of the PDSA method detailed in the reports	2	1
	Documentation of each stage of the PDSA cycles	1.67	1

39	<hr/> <i>Composite score (%)</i> <hr/>	92	50 <hr/>
----	--	-----------	-----------------

Table 5: Summary of findings and assessments

Feature of PDSA	DIVA Design	Design gap (%)	DIVA implementation	Implementation gap (%)	Assessment
Iterative cycles	DIVA is intended to be implemented using multiple iterative cycles which are interlinked such that the 'Verify/Adjust' phases of one cycle feed into the 'Diagnose' phase of the next.	0	DIVA is implemented using multiple (quarterly) cycles which are linked. However, in practice, we observed that while in many cycles the 'Verify/Adjust' phases of one cycle feed into the 'Diagnose' phase of the next, there were occasions where such interlinks were interrupted. This is often due to contextual influences such as donor interests, leadership and administrative deficiencies (Eboireime, Nxumalo, <i>et al.</i> , 2018).	17	Optimal theory-design consonance with mild implementation gap
Prediction-based test of change	The model tests changes in six supply and demand side determinants of health system performance. It includes tools and measures for explicitly predicting change in these determinants (NPHCDA, 2013).	0	During the 'Diagnose/Intervene' phases, action plans to address health system bottlenecks are made which include quantified coverage improvement targets to be met following implementation.	0	Optimal theory-design-implementation consonance
Small scale tests of change	DIVA was designed to be integrated into and implemented across the existing health system. Consequently, it was not practicable to build in small tests of change as that will imply a parallel health system.	100	Given that the intervention was introduced and implemented at scale ab initio, the sample size remains the same in every DIVA cycle. Thus, implementation aligns with the adapted design but not with the PDSA theory.	Not applicable	Design defect, implementation fidelity not applicable
Use of data over time	The design of DIVA is to build on existing data systems. Thus, routine data from the Health Management Information System is used over time to analyse change. Given the capacity of implementers (mostly district health managers and other stakeholders from the local community), DIVA guidelines recommend use of simple descriptive statistical evaluations rather than complex analytic processes such as statistical control charts	0	Data was collected over time and analysed using time series charts and bar graphs. The implementation wholly complies with design.	0	Zero design defect with optimal implementation fidelity
Documentation	The implementation guide contains processes and tools for documenting activities, plans and lessons learned from implementation for each stage of the implementation.	8	There was detailed documentation of the 'Diagnose/Intervene' phases. This was less so with the 'Verify/Adjust' phases which were observed to be the weakest link in the implementation chain.	50	Good theory-design consonance with poor implementation fidelity

1 Conversely, implementation fidelity scores were only optimal with prediction-based test of change
2 and the use of data over time. Multiple cycles were implemented but not appropriately linked to
3 each other. Further the application of the PDSA methods, including the four stages, were not well
4 documented in the reports. This may imply that any observed poor performance of DIVA may have
5 been caused by implementation failure.

6 Discussion

7 This study interrogates the conceptual underpinnings of DIVA and its implementation fidelity in a
8 subnational health system in Nigeria. The implementation science literature recognizes that
9 interventions are successful only, when adapted to the local contexts (McKleroy *et al.*, 2006;
10 Galbraith *et al.*, 2009). However, the process of modifying an intervention must not compromise its
11 core features (theory and internal logic), which are most likely to produce the main expected
12 outcomes. The adaptation therefore must preserve the mechanism through which the desired
13 change in outcomes is achieved, while at the same time assuring that the interventions is a good fit
14 to the problem being addressed (McKleroy *et al.*, 2006; Pérez *et al.*, 2015; Sundell *et al.*, 2016).
15 Design gaps occur when the core components of the intervention are compromised during the
16 adaptation. Implementation gaps occur when implementers fail to adhere to the execution of the
17 design as specified by guidelines. This gap in implementation fidelity commonly results from
18 unstable implementation environment such as evolving policies and discretionary action of
19 implementers, among others (Eboreime, Eyles, *et al.*, 2018; Eboreime, Nxumalo, *et al.*, 2018).

20 Our principal finding in this study is that DIVA remains true to the principles of the PDSA approach,
21 while excluding some components that are not germane to the objective of strengthening the
22 planning process. One such component is the PDSA hallmark of conducting small tests of change.
23 This is mainly because typical PDSAs are used within an organizational context such as to improve
24 services within a clinic. DIVA, however, was designed for higher (systems) level improvement such as
25 complex district health systems which coordinate governance and services across several
26 communities and facilities located within defined geographical boundaries (UNICEF, 2012; Reed and

27 Card, 2016; Eboreime, Idika, *et al.*, 2019; Eboreime, Nxumalo, *et al.*, 2019). Typically, in
28 organizational contexts, PDSAs start small in scale and expand scale as confidence grows. However,
29 DIVA, in Nigeria, was designed to review healthcare processes that had already been implemented
30 and changes in each cycle were intended to address reallocation of resources or priorities rather
31 than making changes to the processes themselves. In fact, small tests of change may disrupt existing
32 routine systems, compromising the goal of integrated health system strengthening with implications
33 for sustainability of the intervention (UNICEF, 2012; Eboreime *et al.*, 2017; Eboreime, Nxumalo, *et*
34 *al.*, 2018). This being said, there may be potential value in applying small tests of change should DIVA
35 be implemented in a newly designed health system or as part of holistic reforms of an existing
36 system. That way lessons learned from these tests can inform strategies and interventions that could
37 sustainably improve health system performance.

38 Another variation in DIVA compared to PDSA is the method of analysing data over time. While
39 temporal trends are reviewed the method of analysis is descriptive, primarily using before and after
40 bar charts. This prevents a more rigorous analysis of process variation. The adaptation was
41 necessary because DIVA is designed to be an easy-to-use tool for decision making by health care
42 managers/policymakers who are not trained in rules for interpreting run charts.

43 We also found that the implementation of DIVA varied partly from the design. Complete
44 implementation fidelity (adherence to design) was observed in two of the five features of PDSA
45 (prediction-based tests of change and the use of data over time). Implementation gaps were
46 observed in two features: iterative cycles and documentation. The remaining PDSA feature, small
47 scale tests of change, was not applicable to assess as it was omitted from the DIVA design, as
48 discussed above. Typical PDSA cycles are iterative, with lessons learned from one cycle feeding into
49 the 'Plan' phase of the cycle that follows. While this is also the design of DIVA, in practice,
50 interruptions to these cycle linkages were observed.

51 In an another article (Eboreime, Nxumalo, *et al.*, 2018), using Kaplan’s Model for Understanding
52 Success in Quality (MUSIQ) (Kaplan *et al.*, 2012), the authors described how contextual factors such
53 as weak political will, administrative inefficiencies and donor interests partly compromised the
54 implementation of DIVA, particularly the ‘Verify/Adjust’ phases. These deficiencies had implications
55 on linkages between cycles on some occasions. The detailed documentation observed with the
56 ‘Diagnose/Intervene’ phases, was not replicated with the ‘Verify/Adjust’ phases, thus reinforcing the
57 untoward effect of these contextual factors with implications on effectiveness.

58 As noted in previous studies (Eboreime *et al.*, 2017; Eboreime, Nxumalo, *et al.*, 2018, 2019), DIVA is
59 potentially effective as it was able to improve isolated performance indicators of vertical
60 programmes such as malaria management and measles vaccination, but did not significantly
61 improve integrated health system performance due to poor implementation of the model. This
62 study provides further insights into possible reasons for suboptimal effectiveness of DIVA. First, a
63 smaller 17% gap in implementing iterative cycles, then the larger 50% gap in documentation. The
64 poor attention given to documentation could have compromised the quality of other processes,
65 given that lessons learned from one cycle may not completely feed into subsequent cycles. This
66 phenomenon may potentially create some form of cyclical redundancies, compromising the
67 effectiveness of the intervention. Thus, to improve the effectiveness of DIVA in this context, more
68 attention must be paid to documentation of processes, outcomes and lessons learned. Beyond the
69 context in which this research was carried out, this study demonstrates that the standard principles
70 of quality improvement interventions should inform both the design and implementation of the
71 interventions in order to ensure effective outcomes.

72 The design of this study may subject the findings to some interpreter bias and subjectivity. Despite
73 these apparent biases, we view the measures and approaches used here as practical proxies for
74 evaluating the alignment of theory, design and implementation in the real-world.

75 Conclusion

76 This study demonstrates how both adaptation and implementation are important for success of QI
77 interventions (Borrelli *et al.*, 2005; Carroll *et al.*, 2007; Stirman *et al.*, 2013; El-Sadr, Philip and
78 Justman, 2014). It also presents an approach for evaluating other QI models using Taylor's PDSA
79 assessment framework as a guide, which might serve to strengthen the theory behind future QI
80 models and provide guidance on their appropriate use. This will be particularly important for
81 improvement models like DIVA, designed for multi-contextual improvements, where different
82 adaptations may be necessary in different countries and where implementation challenges may vary
83 by setting.

84

85 References

- 86 Baker, U., Okuga, M., *et al.* (2015) 'Bottlenecks in the implementation of essential screening tests in
87 antenatal care: Syphilis, HIV, and anemia testing in rural Tanzania and Uganda', *International Journal*
88 *of Gynecology & Obstetrics*, 130, pp. S43–S50.
- 89 Baker, U., Peterson, S., *et al.* (2015) 'Identifying implementation bottlenecks for maternal and
90 newborn health interventions in rural districts of the United Republic of Tanzania', *Bull World Health*
91 *Organ*, 93, pp. 380–389.
- 92 Borrelli, B. *et al.* (2005) 'A new tool to assess treatment fidelity and evaluation of treatment fidelity
93 across 10 years of health behavior research.', *Journal of Consulting and Clinical Psychology*, 73(5),
94 pp. 852–860. doi: 10.1037/0022-006X.73.5.852.
- 95 Carroll, C. *et al.* (2007) 'A conceptual framework for implementation fidelity', *Implementation*
96 *Science*. BioMed Central, 2(1), p. 40. doi: 10.1186/1748-5908-2-40.
- 97 Eboime, E. *et al.* (2017) 'Implementing a Model for Improving Integrated Primary Healthcare
98 Planning and Performance: An Effectiveness Evaluation of DIVA in Kaduna, Nigeria.', in *Cochrane*
99 *Database of Systematic Reviews*. doi: 10.1002/14651858.CD201702.
- 100 Eboime, E. A. *et al.* (2017) 'Evaluating the sub-national fidelity of national initiatives in
101 decentralized health systems: Integrated Primary Health Care Governance in Nigeria', *BMC Health*
102 *Services Research*. BioMed Central, 17(1), p. 227. doi: 10.1186/s12913-017-2179-2.
- 103 Eboime, E. A., Eyles, J., *et al.* (2018) 'Implementation process and quality of a primary healthcare
104 system improvement initiative in a decentralized context: A retrospective appraisal using the Quality
105 Implementation Framework', *Int J Health Plann Mgmt*.
- 106 Eboime, E. A., Nxumalo, N., *et al.* (2018) 'Strengthening decentralized primary healthcare planning
107 in Nigeria using a quality improvement model: how contexts and actors affect implementation',
108 *Health Policy and Planning*, pp. 1–14. doi: 10.1093/heapol/czy042.
- 109 Eboime, E. A., Nxumalo, N., *et al.* (2019) 'Effectiveness of the Diagnose-Intervene-Verify-Adjust

- 110 (DIVA) model for integrated primary healthcare planning and performance improvement: An
111 embedded mixed methods evaluation in Kaduna state, Nigeria', *BMJ Open*, In Press. doi:
112 <http://dx.doi.org/10.1136/bmjopen-2018-026016>.
- 113 Eboreime, E. A., Idika, O., *et al.* (2019) 'Primary healthcare planning, bottleneck analysis and
114 performance improvement: An evaluation of processes and outcomes in a Nigerian context',
115 *Evaluation and Program Planning*. Pergamon, 77, p. 101712. doi:
116 10.1016/J.EVALPROGPLAN.2019.101712.
- 117 El-Sadr, W. M., Philip, N. M. and Justman, J. (2014) 'Letting HIV Transform Academia — Embracing
118 Implementation Science', *New England Journal of Medicine*, 370(18), pp. 1679–1681. doi:
119 10.1056/NEJMp1314777.
- 120 Fixsen, D. L. *et al.* (2005) 'Implementation Research: A Synthesis of the literature'. The National
121 Implementation Research Network, University of South Florida, Louis de la Parte Florida Mental
122 Health Institute. Available at: [https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-](https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf)
123 [MonographFull-01-2005.pdf](https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf) (Accessed: 29 June 2019).
- 124 FMOH (2016) *National Health Policy*. Abuja, Nigeria: Federal Ministry of Health.
- 125 Galbraith, J. S. *et al.* (2009) 'Exploring implementation and fidelity of evidence-based behavioral
126 interventions for HIV prevention: Lessons learned from the focus on kids diffusion case study',
127 *Health Education and Behavior*. doi: 10.1177/1090198108315366.
- 128 Health Foundation (2011) *Evidence scan: Improvement science*. London.
- 129 Henriksson, D. K. *et al.* (2017) 'Bottleneck analysis at district level to illustrate gaps within the district
130 health system in Uganda', *Glob Health Action*, 10(1), p. 1327256.
- 131 James Bell Associates (2009) *Evaluation brief: Measuring Implementation Fidelity*. Arlington, VA.
132 Available at: [https://www.jbassoc.com/wp-content/uploads/2018/03/Measuring-Implementation-](https://www.jbassoc.com/wp-content/uploads/2018/03/Measuring-Implementation-Fidelity.pdf)
133 [Fidelity.pdf](https://www.jbassoc.com/wp-content/uploads/2018/03/Measuring-Implementation-Fidelity.pdf) (Accessed: 10 November 2018).
- 134 Kaplan, H. C. *et al.* (2012) 'The Model for Understanding Success in Quality (MUSIQ): building a
135 theory of context in healthcare quality improvement', *BMJ Qual Saf.*, 21(1), pp. 13–20.
- 136 Kieny, M.-P. *et al.* (2018) *Delivering quality health services: a global imperative for universal health
137 coverage*. Geneva, Switzerland: World Health Organization, OECD, and International Bank for
138 Reconstruction and Development/The World Bank. Available at: <http://apps.who.int/bookorders>.
139 (Accessed: 8 April 2019).
- 140 Koczwara, B. *et al.* (2018) 'Harnessing the Synergy Between Improvement Science and
141 Implementation Science in Cancer: A Call to Action', *Journal of Oncology Practice*. American Society
142 of Clinical Oncology (ASCO), 14(6), pp. 335–340. doi: 10.1200/jop.17.00083.
- 143 Leis, J. A. and Shojania, K. G. (2017) 'A primer on PDSA: executing plan-do-study-act cycles in
144 practice, not just in name.', *BMJ quality & safety*. BMJ Publishing Group Ltd, 26(7), pp. 572–577. doi:
145 10.1136/bmjqs-2016-006245.
- 146 McKleroy, V. S. *et al.* (2006) 'Adapting evidence-based behavioral interventions for new settings and
147 target populations', *AIDS Education and Prevention*. doi: 10.1521/aeap.2006.18.supp.59.
- 148 Muczyk, J. P. (2004) 'A Systems Approach to Organizational Effectiveness: The Alignment of Critical
149 Organizational Dimensions with Selected Business/Competitive Strategies', *Journal of Comparative
150 International Management*, 7(1). Available at:
151 <https://journals.lib.unb.ca/index.php/JCIM/article/view/454>.
- 152 Norman, C. L. *et al.* (2013) *The improvement guide: a practical approach to enhancing*

- 153 *organizational performance*. Jossey-Bass. Available at:
154 https://books.google.ca/books/about/The_Improvement_Guide.html?id=kE4aEnZgBO8C&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false (Accessed: 7 November
155 2018).
156
- 157 O'Connell, T. S. and Sharkey, A. (2013) *Reaching Universal Health Coverage through District Health*
158 *System Strengthening: Using a Modified Tanahashi Model Sub-nationally to Attain Equitable and*
159 *Effective Coverage*. New York: UNICEF.
- 160 Øvretveit, J. (2011) 'Understanding the conditions for improvement: research to discover which
161 context influences affect improvement success.', *BMJ quality & safety*, 20 Suppl 1(Suppl 1), pp. i18-
162 23. doi: 10.1136/bmjqs.2010.045955.
- 163 Pérez, D. *et al.* (2015) 'A modified theoretical framework to assess implementation fidelity of
164 adaptive public health interventions', *Implementation Science*. BioMed Central, 11(1), p. 91. doi:
165 10.1186/s13012-016-0457-8.
- 166 Powell, B. J. *et al.* (2015) 'A refined compilation of implementation strategies: results from the
167 Expert Recommendations for Implementing Change (ERIC) project', *Implementation Science*. BioMed
168 Central, 10(1), p. 21. doi: 10.1186/s13012-015-0209-1.
- 169 Ramaswamy, R. *et al.* (2018) 'Unpacking the black box of improvement.', *International journal for*
170 *quality in health care: journal of the International Society for Quality in Health Care*. Oxford
171 University Press, 30(suppl_1), pp. 15–19. doi: 10.1093/intqhc/mzy009.
- 172 Ramaswamy, R., Johnson, J. K. and Hirschhorn, L. R. (2018) 'Intergrating implementation science
173 approaches into continuous quality improvement', in Johnson, J. K. and Sollecito, W. A. (eds)
174 *McLaughlin & Kaluzny's Continuous Quality Improvement in Health Care*. 5th edn. Burlington, MA:
175 Jones & Bartlett Learning.
- 176 Reed, J. E. and Card, A. J. (2016) 'The problem with Plan-Do-Study-Act cycles', *BMJ Qual Saf*, 25, pp.
177 147–152.
- 178 Sales, A. *et al.* (2006) 'Models, strategies, and tools', *Journal of General Internal Medicine*. Springer-
179 Verlag, 21(S2), pp. S43–S49. doi: 10.1007/s11606-006-0274-x.
- 180 Schroeder, R. G. *et al.* (2008) 'Six Sigma: Definition and underlying theory', *Journal of Operations*
181 *Management*. Elsevier, 26(4), pp. 536–554. doi: 10.1016/J.JOM.2007.06.007.
- 182 Snowden, D. J. and Boone, M. E. (2007) 'A leader's framework for decision making. A leader's
183 framework for decision making.', *Harvard business review*, 85(11), pp. 68–76, 149. Available at:
184 <http://www.ncbi.nlm.nih.gov/pubmed/18159787> (Accessed: 29 June 2019).
- 185 Stirman, S. W. *et al.* (2013) 'Development of a framework and coding system for modifications and
186 adaptations of evidence-based interventions', *Implementation Science*. BioMed Central, 8(1), p. 65.
187 doi: 10.1186/1748-5908-8-65.
- 188 Sundell, K. *et al.* (2016) 'Novel Programs, International Adoptions, or Contextual Adaptations? Meta-
189 Analytical Results From German and Swedish Intervention Research', *Journal of Clinical Child &*
190 *Adolescent Psychology*. Routledge, 45(6), pp. 784–796. doi: 10.1080/15374416.2015.1020540.
- 191 Taylor, M. J. *et al.* (2013) 'Systematic review of the application of the plan–do–study–act method to
192 improve quality in healthcare', *BMJ Quality & Safety*, 14(23), pp. 290–298.
- 193 UNICEF (2012) *The guidebook: Strengthening district management capacity for planning,*
194 *implementation and monitoring for results with equity*. New York: UNICEF.
- 195 UNICEF (2016) *The UNICEF Health Systems Strengthening Approach*. New York.

196 WHO (2009) *Systems Thinking for Health Systems Strengthening*. Geneva. doi: World Health
197 Organization.

198 Williams, D. M. (2019) *Improvement Science and Implementation Science*. Available at:
199 <http://www.dmwaustin.com/2019/07/26/improvement-science-and-implementation-science/>
200 (Accessed: 6 May 2020).

201 Van Wyk, R. J. (1980) 'Handbook of Futures Research', *Technological Forecasting and Social Change*.
202 doi: 10.1016/0040-1625(80)90077-3.

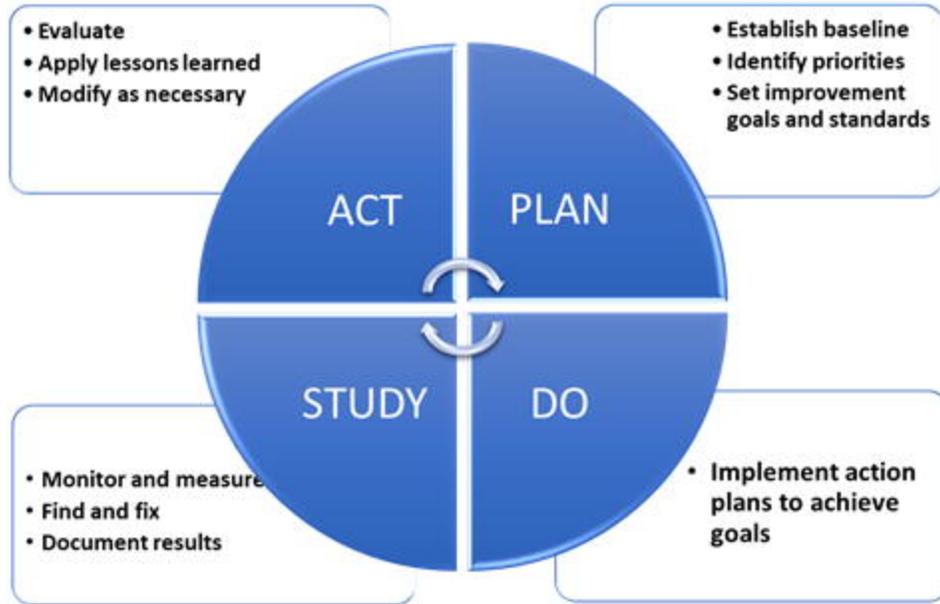
203 **Acknowledgements**

204 We acknowledge the support (including permission to use data) of Nigeria's National Primary Health
205 Care Development Agency and the Kaduna State Primary Healthcare Development Agency, Nigeria.

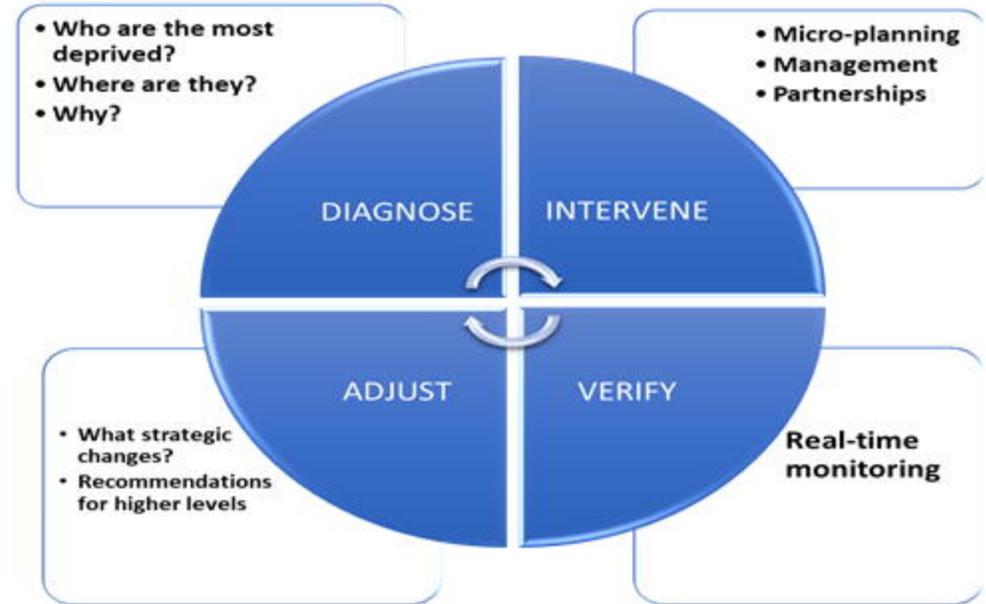
206 **Conflict of interest statement:** All authors declare no conflict of interests

207

A Typical PDSA cycle



The DIVA cycle



Key

- Standard design
- ⬡ Adaptive design
- ◇ Real-world implementation
- Zero Implementation

