Title (Draft)

Evaluating Public Health Interventions 10: Let's learn as we go, and minimize large scale public health intervention failures: the LAGO Design for optimizing complex multi-component Interventions

Authors

Donna Spiegelman[1 2], Daniel Nevo[2 3], Arhit Chakrabarti[4], Dale Barnhart[2], Bhaswati Ganguli[5], Judith J. Lok[2 6]

1. Yale School of Public Health, Connecticut, USA
2. Harvard T.H. Chan School of Public Health, Massachusetts, USA
3. Tel Aviv University, Tel Aviv-Yafo, Israel
4. Novartis Healthcare Pvt. Ltd., Hyderabad, India
5. University of Calcutta, Calcutta, India
6. Boston University, Massachusetts, USA

April 2019

Abstract

In the face of vast numbers of preventable deaths around the world and gaping disparities in their distribution, we cannot afford to run null trials of proven interventions. The gold standard in biomedicine, the individually randomized clinical trial, is ill-suited in its role as the primary tool for knowledge generation for large scale complex public health interventions of multi-component interventions. Here, we discuss the new Learn as You Go (LAGO) design. In LAGO trials, the components of the complex package are repeatedly optimized in pre-planned stages, until the intervention package is optimized both in terms of its target goal and cost. In this column, we elucidate key features of the LAGO' design, illustrated by the null BetterBirth study, a large-scale public health intervention trial aimed at reducing maternal and neonatal mortality through the use of WHO's best practices checklist.

In 2014, physician, scientist, author, and MacArthur genius award winner, Atul Gawande, who, among many accomplishments, successfully promoted the Safe Surgery Checklist, which has drastically reduced errors during surgery around the world (1), turned his focus to reducing maternal and neonatal mortality. Dr. Gawande's idea was simple: encourage frontline birth attendants to adhere to 28 World Health Organization-recommended, evidence-based essential birth practices using a checklist modeled after the Safe Surgery Checklist for which Gawande initially gained international renown. The effectiveness of this Safe Childbirth Checklist (2) was evaluated in a matched-pair, cluster-randomized trial in the impoverished, densely populated rural Northern Indian state of Uttar Pradesh (UP). At the time, data suggested that UP lagged far beyond many other places in the world in terms of both neonatal (49 per 1,000) and maternal (258 per 100,000) mortality, exemplifying one of the most egregious global health equity gaps currently facing us, with as much as a 100-fold difference in maternal mortality rates between high income and low income countries.(3)

Almost 160,000 mothers were enrolled in this large-scale, high-profile study.

Five years later, the results of the study were reported in the *New England Journal of Medicine*: null (4). How could this have happened? About this, much has been written, (5-7) and surely more will be still. The final report of the trial noted that, despite improvements in EBP adherence among intervention sites, the final levels of adherence to essential birth practices included on the checklist among intervention facilities may not have been sufficient to improve health outcomes. For example, at 2 months after randomization, birth attendants performed appropriate hand hygiene in only 35% of cases and although 79% of mother–infant pairs initiated skin-to-skin warming, only 19% of mother–infant pairs maintained that contact for 1 hour. Although the trial's Data Safety and Monitoring Board (DSMB) knew that these modest improvements had not resulted in improved survival among mothers and neonates, given the principles of randomized clinical trial practice, nothing could be done. My colleague, Dr. V[1], was on the BetterBirth's DSMB. He recently recounted to me "It was heartbreaking to sit on that DSMB, watching the study fail, and have our hands tied – there was nothing we could do."

[1] Alphabetic letters replace the actual names to preserve the blindness of the review. The published paper will provide the actual names.

Clearly, standard methods for the design and conduct of randomized clinical trials, including the BetterBirth trial, prohibit adaptation, tailoring and tweaking of the intervention once randomization is complete, even in the face of an unfolding intervention failure.

The culture, values and methodologies of randomized clinical trials, although extremely useful for knowledge generation in drug and device development, are not necessarily optimal when applied rigidly to large scale public health evaluations, whether randomized, often by large clusters as in BetterBirth, or not. Large scale public health interventions typically do not involve evaluation of a single simple intervention, such as multivitamins vs. placebo;(8, 9) rather, they tend to evaluate the impact of a complex, multi-component intervention, whether compared to control clusters as in effectiveness research, or when two or more complex intervention strategies are compared as in implementation science (10, 11).

In the BetterBirth trial, the complex intervention consisted of use of the Safe Childbirth Checklist (SCC) as well as an implementation package that involved, among other features, a 2-3 day motivational and educational launch and 43 full days of peer-to-peer coaching visits delivered to birth attendants at the intervention facilities.(12) In addition, each intervention facility chose at least one staff member to serve as a Childbirth Quality Coordinator, a local champion for use of the checklist and continued coaching. Had the investigators been able to adapt, tweak or tailor the intervention as the trial progressed, they likely could have adapted the implementation package to address the barriers and challenges that prevented the successful implementation of this complex intervention.

The need for a Learn as You Go Design (LAGO) that could allow researchers to optimize the development of a complex, multi-component interventions, was first pointed out to me by physician and implementation scientist, Dr. X[1], one of the original BetterBirth investigators. It took quite some time, and the engagement and support of an expert mathematical statistician, Dr. Y[1], and my former post-doctoral fellow, Dr. Z[1], to arrive at a solution that solved this

problem in a mathematically rigorous fashion.(13) In LAGO, researchers initially propose a multi-component intervention package based on their subject-matter knowledge. This initial intervention is implemented in stage one, then, based on a pre-planned analysis, data from this first stage are evaluated and used to suggest the optimal intervention for the second stage. This process continues across subsequent stages, with data from all previous stages being used to inform the selection of the intervention for the subsequent stage. Ideally, an optimal intervention, defined as an intervention that attains some pre-specified goal subject to cost or other constraints, is identified and evaluated in the last stage. As in a standard randomized controlled trial, a valid test of the overall complex intervention framework can be conducted, with the usual Type I error rate. Additionally, researchers can obtain estimates and 95% confidence intervals for the effect of each individual intervention component and a 95% confidence region for the optimal intervention package. The Box summarizes concisely the required inputs and outputs of the LAGO design.

Illustrative Example – The BetterBirth study

Let's make this more concrete and consider what could have been done during the BetterBirth trial had it followed a LAGO design. The following analysis of the BetterBirth trial is adapted from a pre-published manuscript that developed and illustrated the LAGO design.(13) While this retrospective analysis does not necessarily reflect the optimal LAGO, it can be used to illustrate how data can help us "learn as we go". Here, we present a hypothetical optimization of the intervention with respect to a single essential birth practice, oxytocin administration immediately after delivery, which is as recommended by the WHO(14) to prevent postpartum hemorrhage, a major cause of maternal mortality. We will consider a simplified two-component implementation package and seek to identify the optimal duration of the launch (in days) and number of coaching visits needed to achieve oxytocin administration for at least 85% of the births. To identify the most cost-effective intervention package, we will exploit variation in the launch duration and number of visits that occurred across three development stages: Before the main trial, the BetterBirth team conducted two pilot studies to refine the intervention

package, and we will consider these two pilot studies and the randomized controlled trial to be the first, second, and third stages of the overall study. In stage 1, the launch duration was 3 days, and in stages 2-3, it was 2 days. Compared to stage 1, the intensity of coaching visits was increased in stage 2, and further increased in stage 3. For our analysis, pre-post data on EBP adherence was available from 2 centers in the first stage and 4 in the second stage, while the third stage included 15 centers in the control arm and 15 centers in the intervention arm.

In Table 1 (taken from (13)), we estimated the relative risks for each package component, with higher values reflecting greater adherence to oxytocin administration after delivery. Data used for the analysis at each stage includes all data obtained up to that stage. The component-wise odds ratios stabilized over the successive 3 stages, and the optimal interval package was finalized by stage 2, with no subsequent further changes. Both package components had a significant impact on uptake of this recommended birth practice ($p < 0.001$). Using this information, we can identify the most cost-effective intervention package that will achieve oxytocin administration of at least 85%. Assuming unit costs of $800 per day of launch and $170 per coaching visit, and assuming (for pragmatic reasons) that the launch duration can be between 1 and 5 days long while between 1 and 40 coaching visits are needed, the recommended cost-effective intervention package for achieving the 85% oxytocin administration for a facility with an average birth volume of 175 births per month was a launch duration of 3 days and 1 coaching visit. Using this intervention, the 95% confidence interval for the proportion of births receiving oxytocin administration is (79%-93%). The estimated total cost of the estimated optimal intervention package was $2570/facility (interquartile range $2462-$6797). Our analysis also produces a 95% confidence set of alternative combinations of launch duration and coaching visits that could also have achieved 85% oxytocin administration. Alternative interventions packages in the confidence set were (1.5, 40), (2, ≥27), (2.5, <20), (3, <5), where the first number refers to the launch duration (days) and the second the number of couching visits. This example uses the actual study data based upon somewhat ad hoc choices made by the BetterBirth investigators, who did not have access to the LAGO methodology, at the end of each stage. Future interventionists who prospectively use the LAGO design would

experience the added benefit of being about to deploy an optimized intervention in each phase.

Towards "personalized" public health evaluation with LAGO

You may have noticed that in the example above, we estimated the optimal intervention package for a particular facility type, one with the average monthly birth volume (174 births per month) across the BetterBirth facilities. The optimal intervention package could have been estimated for facilities with any other birth volume; in fact, the optimal intervention package is estimated at 1 coaching visit and 3.5 days launch duration for facilities at the 25^{th} percentile of birth volume (130 births per month). Because the optimal intervention package rule is one of the outputs of the LAGO design, a customized optimal intervention package can be given for facilities based upon the observed characteristics used to estimate the rule, whether or not the actual values of the characteristics were observed in the study used to estimate the rule. The recommended intervention package from a previous LAGO design can be used as the starting value for a new LAGO design in perhaps a new setting or implementation at a larger scale. Of course, implementation at a larger scale may involve extrapolating outside the range of the observed data and can be risky, but in real-life public health settings it might nevertheless be best to use prior empirical data , in addition to common sense, expert advice, and contextual considerations, to guide intervention scale-up.

Variations of LAGO designs

There are several versions of the LAGO design. In a controlled LAGO design, there is an initial randomization of clusters or individuals to a group that does not receive the intervention. Typically, in public health research, this will be the 'standard of care' group. In the BetterBirth example above, stage 3 had this design. Alternatively, or in addition, there may be concurrent variation in the intervention package component intensities. This variation may be planned, and, ideally, randomized, or unplanned. If it is unplanned, the LAGO design takes on a quasi-

experimental flavor, since before-after comparisons within clusters provide full control for time-invariant confounding. For time-varying confounding, causal inference methods can be used in the stage-wise and final analyses to allow for rigorous inference. Further variations can be contemplated. For example, consider the NIMH-sponsored **E**nabling translation of **S**cience to **S**ervice to **EN**hance Depression Car**E (**ESSENCE) training trial (15) led by Vikram Patel of Harvard and Sangath – India, to which a number of the authors of this column are also contributing. To fill the overwhelming need for mental health professionals in India, a country of 1.5 billion people and less than 10,000 mental health professionals, non-specialist health workers, such as community health workers, can successfully develop proficiency in depression screening and its treatment through a brief psychological intervention, the Healthy Activity Program, after a short face-to-face training (16). Unfortunately, even with this level of task shifting, it is prohibitive in terms of time and money to offer face-to-face training to all of India's community health workers. Thus, we are now rolling out a non-inferiority trial to compare the effectiveness and cost-effectiveness of face-to-face training to a digital alternative. The digital alternative can be viewed as a multi-component intervention package, since components of the curriculum can be tweaked in many ways: reading level, amount of video content, frequency and number of quizzes, digital content and graphics to promote engagement, required intensity of the training period, required duration of the training period, etc. With such a large number of components to alter, careful experimental design methodologies, such as the fractional factorial design, must be utilized to ensure that there is sufficient unconfounding variation to estimate the effects of each individual component. We will likely utilize a design in which, in each successive stage, an additional component is varied, until variation has been observed on all components. Then, in the final stage we will use LAGO to estimate a recommended intervention, and its likely effect on the outcomes of interest. The investigative team's use of LAGO in this setting reflects their commitment to ensure that this trial does not fail.

Discussion

As evidenced by the vast disparities in the rates of many major diseases between high to low and middle income countries, many of the world's most pressing health problems, including tuberculosis, HIV/AIDS, cardiometabolic disorders, maternal and under-5 mortality to mental health, and cervical, colorectal and lung cancer, can be virtually eliminated using well-known, effective interventions. Without any new scientific discoveries, implementation and prevention science promises to address the gap between what we know about the prevention of these diseases and what we do in practice. What is needed is cost-effective interventions, perhaps in the form of comprehensive complex packages that target one or more diseases simultaneously, to address them in contextually sustainable forms. We who follow the scientific method as the primary, if not sole, form of knowledge generation neglect to recognize that the engineering paradigm, which indeed involves tweaking, tailoring, and adapting of an underlying concept until an acceptable or even superlative model is attained, has brought us most of what is central to modern societies: the car and plane; the telephone, television, radio and computer; electricity and light after dark, roofs that don't leak when it rains, central heating, much of modern agriculture and nutritional energy abundance; movies, videos, and mass-distributed music. By using the rigorously developed LAGO design, we can leapfrog over the ill-fitting paradigm of the individually randomized clinical trial from which biomedicine has achieved so much, primarily in terms of late stage treatment, and may be able to avoid further null results for trial studying the implementation of proven public health interventions.

Box

LAGO Inputs

1. Maximum sample size or minimum power for the overall test of the impact of the overall intervention framework
2. Number of components, and for each one
    a. Minimum and maximum feasible value
    b. Best guess as to the effect per unit increase
    c. Unit cost
3. Mean and standard deviation (or cumulative incidence) of the outcome
4. Maximum number of stages, and maximum and minimum sample size per stage
5. Number of clusters, if any (not required); ICC if clustered
6. Study goal – e.g. 85% of births administering oxytocin
7. Cluster- or patient characteristics for which the intervention is to be 'personally' optimized, and effect of each per unit increase, if any (not required)

LAGO Outputs

1. P-value for the test of the overall intervention framework effect. This p-value will not depend on the outcome model.
2. Point and interval estimates for each package component, based on the outcome model.
3. Point and interval estimate of the optimal intervention, by baseline characteristics (optional), with
    a. Confidence interval for estimated goal probability
    b. Point and interval estimate of cost
4. Personalized estimation rule for optimal intervention for future facilities or patients (optional)

Citations

1. Haynes AB, Berry WR, Gawande AA. What do we know about the safe surgery checklist now? Ann Surg 2015;261(5):829-30.
2. World Health Organization, Ariadne Labs. WHO safe childbirth checklist implementation guide: improving the quality of facility-based delivery for mothers and newborns. In. Geneva: World Health Organization; 2015.
3. Alkema L, Chou D, Hogan D, Zhang S, Moller A-B, Gemmill A, et al. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Maternal Mortality Estimation Inter-agency Group. The Lancet 2016;387(10017):462-474.
4. Semrau KEA, Hirschhorn LR, Marx Delaney M, Singh VP, Saurastri R, Sharma N, et al. Outcomes of a Coaching-Based WHO Safe Childbirth Checklist Program in India. N Engl J Med 2017;377(24):2313-2324.
5. Pai M, Schumacher SG, Abimbola S. Surrogate endpoints in global health research: still searching for killer apps and silver bullets? BMJ Glob Health 2018;3(2):e000755.
6. Molina RL, Bobay L, Semrau KEA. Historical Perspectives: Lessons from the BetterBirth Trial: A Practical Roadmap for Complex Intervention Studies. NeoReviews 2019;20(2):e62-e66.
7. Barnhart DA, Semrau K, Zigler C, Spiegelman D. Optimizing the development and evaluation of complex interventions: Lessons learned from the BetterBirth Program and associated trial. [dissertation]. Boston, MA: Harvard T.H. Chan School of Public Health; 2019.
8. Fawzi WW, Msamanga GI, Urassa W, Hertzmark E, Petraro P, Willett WC, et al. Vitamins and perinatal outcomes among HIV-negative women in Tanzania. N Engl J Med 2007;356(14):1423-31.
9. Fawzi WW, Msamanga GI, Spiegelman D, Wei R, Kapiga S, Villamor E, et al. A randomized trial of multivitamin supplements and HIV disease progression and mortality. N Engl J Med 2004;351(1):23-32.
10. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655.
11. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015;350:h1258.
12. Kara N, Firestone R, Kalita T, Gawande AA, Kumar V, Kodkany B, et al. The BetterBirth Program: Pursuing Effective Adoption and Sustained Use of the WHO Safe Childbirth Checklist Through Coaching-Based Implementation in Uttar Pradesh, India. Global Health: Science and Practice 2017;5(2):232-243.
13. Nevo D, J. Lok J, Spiegelman D. Analysis of "Learn-As-You-Go" (LAGO) Studies. arXiv preprint arXiv:1808.06310; 2018.
14. World Health Organization. WHO recommendations: intrapartum care for a positive childbirth experience. In. Geneva: World Health Organization; 2018.
15. Sangath. The ESSENCE Trial. In: What we do: Adult mental health.
16. Patel V, Weobong B, Weiss HA, Anand A, Bhat B, Katti B, et al. The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. Lancet 2017;389(10065):176-185.

Figures

	Stage 1 $n_1 = 73$ OR (95% CI)	Stages 1-2 $(n_1 + n_2 = 1780)$ OR (95% CI)	Stages 1-3 $(n_1 + n_2 + n_3 = 6124)$ OR (95% CI)
Coaching Visits (per 3 visits)	7.95 (1.77,73.95)	1.11 (0.96,1.28)	1.08 (1.04,1.12)
Launch Duration (days)	1.41 (0.76,2.64)	2.65 (1.95,3.77)	2.79 (2.41,3.23)
Birth Volume (monthly, per 100)	0.37 (0.00,32.33)	2.11 (1.93,2.33)	1.94 (1.84,2.06)
	$\hat{x}^{opt,(2,n_1)} = (1,5)$	$\hat{x}^{opt,(3,(n_1,n_2))} = (3,1)$	$\hat{x}^{opt} = (3,1)$

OR, estimated odds ratio $\exp(\hat{\beta})$; 95% CI, 95% Confidence interval for the odds ratio. n_k, sample size of stage k, $k = 1, 2, 3$. $\hat{x}^{opt}$ is the estimated recommended intervention at the end of the study. $\hat{x}^{opt,(k,(n_1,\ldots n_{k-1}))}$ is the estimated recommended intervention at the start of stage k based upon the combined data up to stage $k-1$. In the estimated recommended interventions, the first component is the launch duration (in days) and the second component is the number of coaching visits.

Table 1 Stage-wise results for the BetterBirth Study LAGO design. Table reprinted from (13)