The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 DECEMBER 14, 2017 VOL. 377 NO. 24

Outcomes of a Coaching-Based WHO Safe Childbirth Checklist Program in India

Katherine E.A. Semrau, Ph.D., M.P.H., Lisa R. Hirschhorn, M.D., M.P.H., Megan Marx Delaney, R.N., M.S.N., M.P.H., Vinay P. Singh, P.G.D.G.B.M., B.Tech., Rajiv Saurastri, M.A., Narender Sharma, Ph.D., M.B.A., Danielle E. Tuller, M.H.S., Rebecca Firestone, Sc.D., M.P.H., Stuart Lipsitz, Sc.D., Neelam Dhingra-Kumar, M.D., M.B., B.S., Bhalachandra S. Kodkany, M.D., D.G.O., M.B., B.S., Vishwajeet Kumar, M.D., M.P.H., and Atul A. Gawande, M.D., M.P.H., for the BetterBirth Trial Group*

ABSTRACT

BACKGROUND

The prevalence of facility-based childbirth in low-resource settings has increased dramatically during the past two decades, yet gaps in the quality of care persist and mortality remains high. The World Health Organization (WHO) Safe Childbirth Checklist, a quality-improvement tool, promotes systematic adherence to practices that have been associated with improved childbirth outcomes.

METHODS

We conducted a matched-pair, cluster-randomized, controlled trial in 60 pairs of facilities across 24 districts of Uttar Pradesh, India, testing the effect of the BetterBirth program, an 8-month coaching-based implementation of the Safe Childbirth Checklist, on a composite outcome of perinatal death, maternal death, or maternal severe complications within 7 days after delivery. Outcomes — assessed 8 to 42 days after delivery — were compared between the intervention group and the control group with adjustment for clustering and matching. We also compared birth attendants' adherence to 18 essential birth practices in 15 matched pairs of facilities at 2 and 12 months after the initiation of the intervention.

RESULTS

Of 161,107 eligible women, we enrolled 157,689 (97.9%) and determined 7-day outcomes for 157,145 (99.7%) mother–newborn dyads. Among 4888 observed births, birth attendants' mean practice adherence was significantly higher in the intervention group than in the control group (72.8% vs. 41.7% at 2 months; 61.7% vs. 43.9% at 12 months; P<0.001 for both comparisons). However, there was no significant difference between the trial groups either in the composite primary outcome (15.1% in the intervention group and 15.3% in the control group; relative risk, 0.99; 95% confidence interval, 0.83 to 1.18; P=0.90) or in secondary maternal or perinatal adverse outcomes.

CONCLUSIONS

Birth attendants' adherence to essential birth practices was higher in facilities that used the coaching-based WHO Safe Childbirth Checklist program than in those that did not, but maternal and perinatal mortality and maternal morbidity did not differ significantly between the two groups. (Funded by the Bill and Melinda Gates Foundation; Clinical Trials number, NCT02148952.)

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Semrau at Ariadne Labs, 401 Park Dr., 3rd Flr. E., Boston, MA 02215, or at ksemrau@ariadnelabs.org.

*A complete list of investigators in the BetterBirth Trial Group is provided in the Supplementary Appendix, available with the full text of this article at NEJM.org.

Drs. Kodkany, Kumar, and Gawande contributed equally to this article.

This is the *New England Journal of Medicine* version of record, which includes all *Journal* editing and enhancements. The Author Final Manuscript, which is the author's version after external peer review and before publication in the *Journal*, is available under a CC BY license at PMC5672590.

N Engl J Med 2017;377:2313-24.
DOI: 10.1056/NEJMoa1701075

Globally, maternal mortality ranges from 3 to 1360 per 100,000 births, neonatal mortality from 0.95 to 40.6 per 1000 live births, and the rate of stillbirths from 1.2 to 56.3 per 1000 births, with low-income and middle-income countries having rates an order of magnitude higher than those in high-income regions.[1,2] Although there have been reductions in mortality in recent decades, there is substantial room for improvement.[1,3-5] Despite a dramatic shift from home to facility-based births, birth attendants often do not adhere to practices known to reduce mortality, and mortality has not decreased as expected.[6]

Research has shown that programs with the sole purpose of strengthening birth attendants' training or improving supply availability are insufficient to meaningfully improve patient care or outcomes.[7] Conversely, interventions incorporating job aids, such as checklists or case sheets, and direct, in-person support have proved effective in improving clinical practices[8,9] as well as outcomes.[10-12] To bridge the gap between evidence and practice, the World Health Organization (WHO) created the Safe Childbirth Checklist, a practical tool to assist birth attendants in planning for and performing a more complete bundle of 28 essential birth practices.[13,14] These practices are related to the most common causes of avoidable death for women and newborns.[15]

Studies have previously shown that, when well implemented at a small scale, the WHO Safe Childbirth Checklist improves facility-based birth attendants' adherence to evidence-based care.[16-18] We performed a large cluster-randomized trial of coaching-based implementation of the checklist (the BetterBirth program) in Uttar Pradesh, India.[19] We intended our intervention to support providers at multiple levels of the health system in using the checklist appropriately, to identify gaps in the quality of care at facilities, and to activate resources (e.g., skills training and supply provision) within the existing health care system to address these gaps. We hypothesized that this intervention, implemented at the facility (cluster) level, would result in a reduction in a composite outcome of stillbirth, early neonatal death, maternal death, or maternal severe complications during days 0 to 7.

Methods

Trial Design

We conducted a matched-pair, cluster-randomized, controlled trial in government health facilities that received either the BetterBirth program, a coaching-based implementation of the WHO Safe Childbirth Checklist (60 facilities), or the existing standard of care (60 facilities). We have described the methods of the BetterBirth Trial,[19] the BetterBirth intervention,[20,21] and our data quality-assurance system[22] elsewhere. The trial protocol and statistical analysis plan are available with the full text of this article at NEJM.org. The trial sponsor (the Bill and Melinda Gates Foundation) reviewed the trial design and sample-size calculations but was not involved in data collection, management, analysis, or interpretation; the writing of the manuscript; or the decision to submit the manuscript for publication.

Trial Setting and Participants

The most populous state in India (population of 204 million, 77% rural),[23] Uttar Pradesh is a high-priority region for national and international public health organizations owing to its persistently high neonatal mortality (32 per 1000 live births) and maternal mortality (258 per 100,000 births).[24,25] The government of Uttar Pradesh permitted the trial to proceed in 38 districts, in which we identified 320 eligible facilities. We considered a facility to be eligible if it was designated as a primary health center, community health center, or first referral unit; had at least 1000 deliveries annually; had at least three birth attendants with training of at least the level of an auxiliary nurse midwife; had no other concurrent quality-improvement or research programs; and had district and facility leadership willing to participate. The final trial sample included 120 facilities across 24 districts (see the Supplementary Appendix, available at NEJM.org).

We matched facilities (i.e., clusters) on the basis of the following criteria before randomization: geographic zone, functional classification (primary health center, community health center, or first referral unit), distance to a district hospital, annual birth volume, and number of birth attendants. We randomly assigned facili-

ties to trial groups within each matched pair. After matching and randomization, we collected baseline data on practice adherence in 10 sites to confirm successful matching.

Women who were registered for labor and delivery — excluding women who delivered outside the facility, women who were referred from another facility, or women who were admitted for abortion — were eligible for the trial. At each intervention facility and its matched control site, we began enrolling patients 2 months after the initiation of the intervention. Enrollment continued until the target sample size of the site was reached or until 24 weeks after the completion of the intervention, whichever occurred first, with a 12-week minimum follow-up.

INTERVENTION

We implemented the BetterBirth program in accordance with the Engage–Launch–Support model (Fig. S1 in the Supplementary Appendix) that was piloted at nontrial sites in Karnataka and Uttar Pradesh, India.[18,21,26] Coaches (nurses) and coach team leaders (physicians or public health professionals), all of whom were unaffiliated with the facilities and comprehensively trained to apply a standard behavior-change framework, conducted site visits during the 8-month Support phase.[20,21] We expected coaches to conduct 43 daylong visits to each facility, beginning twice weekly and tapering to monthly visits. Coach team leaders accompanied coaches on alternating visits (23 total visits). Each 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.

Coaches motivated birth attendants to use the checklist and to identify, understand, and resolve barriers to providing quality care.[20,21] Coach team leaders supported facility leadership in fostering team communication and addressing gaps in care at facility and district levels by accessing resources through the established health care system. Cloud-based data collection enabled rapid feedback on the progress of a facility. We provided no clinical-skills training, financial support, or clinical supplies (except paper copies of the checklist).

DATA COLLECTION AND OUTCOMES

The primary outcome was a composite outcome of events occurring within the first 7 days after delivery, incorporating stillbirth; early neonatal death; maternal death; or self-reported maternal severe complications, including seizures, loss of consciousness for more than 1 hour, fever with foul-smelling vaginal discharge, hemorrhage, or stroke. We selected measures of complications on the basis of definitions from the WHO guidance on maternal severe complications, using questions that had previously been validated for the reporting of complications by patients.[27-30] We calculated a prespecified, additional composite outcome consisting of maternal and perinatal death only.

Secondary maternal outcomes by 7 days after delivery included maternal death, maternal complications, interfacility transfer (referral), cesarean section, hysterectomy, blood transfusion, and return to the facility for a health problem. Secondary newborn outcomes included stillbirth, early neonatal death, and interfacility transfer. We assessed all outcomes from facility register information and by contacting women or close family members by telephone between 8 and 22 days after delivery. If we received no response by 22 days after delivery, a field worker conducted a home visit and attempted to follow up until 42 days after delivery.

In addition, we selected a convenience sample of 15 matched pairs of facilities in which trained nurse–data collectors directly observed birth attendants providing care during a 12-hour (daytime) shift at 2 months after the initiation of the intervention (during the highest intensity of coaching) and 12 months after the initiation of the intervention (4 months after the cessation of coaching). These independent observers measured practice adherence, including supply availability (Table S3 in the Supplementary Appendix). Intervention staff and independent observers were not present at the same facility simultaneously.

Owing to the nature of the intervention, we were unable to prevent any facility staff, most trial staff, or any investigators from being aware of the identity of intervention and control facilities. Call-center staff, who collected the majority of outcome data, were unaware of facility assignments.

ETHICS COMPLIANCE

The leadership at each facility provided facility-level consent for participation and permission for trial staff to collect deidentified data on every eligible woman from facility registers. Before

patient discharge, we obtained verbal informed consent and contact information from each woman or her surrogate for follow-up. Data collectors reconfirmed verbal consent at the start of the follow-up call or visit. In directly observed births, women or their surrogates provided written consent for observation.

At trial initiation, birth attendants and facility staff verbally agreed to participate. Before an independent observer collected data, the birth attendant verbally reconfirmed agreement. Electronic data were deidentified and stored in a Health Insurance Portability and Accountability Act–compliant database to ensure participant privacy.

The trial protocol was approved by ethics review boards at Community Empowerment Lab, Jawaharlal Nehru Medical College, the Harvard T.H. Chan School of Public Health, Population Services International, the WHO, and the Indian Council of Medical Research. The protocol was reviewed and reapproved on an annual basis. A data and safety monitoring board met every 6 months after the initiation of enrollment; one interim analysis was conducted when 30% of the data were collected (see the Supplementary Appendix).

SAMPLE SIZE

We hypothesized a priori that the rate of the primary composite outcome would be 15% lower in the intervention group than in the control group. We estimated the intracluster (within-facility) correlation to be 0.01 and the matching effect to reduce the standard error by 45%, basing these parameters on previous studies.[31] We aimed to enroll 171,964 women (85,982 per group) to detect a 15% relative difference (60 events per 1000 births in the control group vs. 51 events per 1000 births in the intervention group) with 80% power and an alpha level of 0.05. On the basis of limited data available from Uttar Pradesh, the baseline rate of the primary composite outcome may have been as high as 120 events per 1000 births. The baseline rate that was used in calculations was purposively set lower than the estimated rate owing to limited information as well as the inclusion of home-based birth events in the available data, which may have elevated mortality.

In assessing practice adherence, we assumed an intracluster (within-facility) correlation of 0.01 and a design effect of matching of 0.75. With 15 matched pairs, we had more than 80% power at an alpha level of 0.05 to detect an absolute difference of 8.5 percentage points in the rate of any birth practice between the trial groups.

STATISTICAL ANALYSIS

Using an intention-to-treat approach, we compared outcomes between trial groups using a Rao–Scott chi-square test, accounting for the matched-pair, cluster design.[32] The main outcome was the dichotomous composite outcome that was present if any of the three main outcomes occurred (maternal death, stillbirth or early neonatal death, or maternal severe complications). This variable was then used to estimate a composite relative risk.[33,34] An additional secondary composite outcome included maternal or perinatal death only. In secondary analyses, each of the main outcomes was compared across groups; a Rao–Scott test with 3 degrees of freedom was used to assess the overall causal effect. No adjustment for multiplicity of testing was made.

For the subgroup of directly observed births, we calculated adherence frequencies for each prespecified measured practice at 2 months and 12 months after the initiation of the intervention. Furthermore, we calculated an aggregate adherence score by summing the total number of 18 checklist practices that should be performed regardless of the health status of the mother–newborn dyad (see the Supplementary Appendix). We generated the mean number of practices (presented as a fraction of 18) performed in each trial group and compared the differences at each time point, using a Rao–Scott chi-square test.[32] For comparison of individual practices, we used a bias-corrected logistic-regression approach that can handle zero cells and complete separation of points within strata and clusters.[35] We conducted all statistical analyses using SAS software, version 9.4 (SAS Institute).

RESULTS

FACILITY AND PATIENT CHARACTERISTICS

All 120 matched and randomly assigned facilities initiated the trial. During data collection, 2 facilities closed for renovations, halting enrollment prematurely for those facilities and their matched pairs. Of the 163,939 women who were

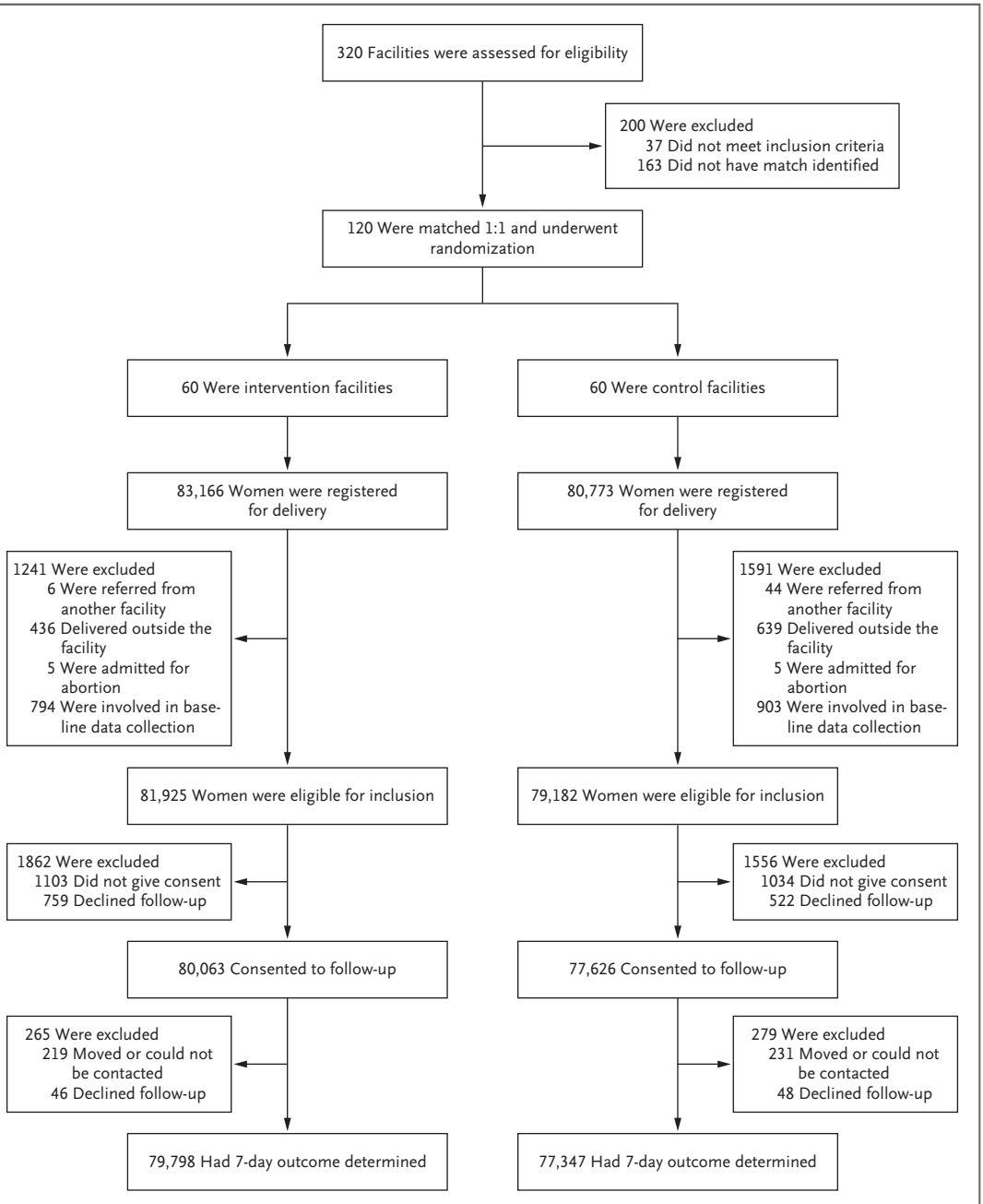

Figure 1. Randomization and Follow-up.

After the matching and randomization of facilities, baseline data on adherence to birth practices were collected in 10 sites to confirm successful matching. Women who were involved in the baseline data collection (794 at intervention facilities and 903 at control facilities) were not included in the trial sample size.

registered for labor and delivery (83,166 in the intervention group and 80,773 in the control group), 98.3% (161,107) were eligible for trial inclusion. Of the eligible women, 97.9% (157,689) provided consent (Fig. 1). We collected 7-day outcomes for all but 544 (0.3%) of the consenting women. There were no significant differences between intervention and control groups in facility, maternal, or newborn characteristics (Table 1).

The BetterBirth program was successfully im-

plemented in all 60 intervention facilities, with high fidelity to the expected number of coaching visits (average, 42.1 visits of 43 expected), interactions with facility leadership (average, 14.8 interactions of 11 expected), and facility data-sharing meetings (average, 8.6 meetings of 11 expected).

ADHERENCE TO BIRTH PRACTICES

After 2 months of twice-weekly coaching, birth attendants in intervention facilities (1259 observations) performed, on average, 72.8% of the 18 measured practices, whereas birth attendants in control facilities (1304 observations) performed 41.7% of the practices (P<0.001) (Table 2). Birth attendants performed the majority of specific practices, such as blood-pressure and temperature assessment, proper hand hygiene, and early newborn care, at significantly higher rates in the intervention group than in the control group. Supply availability was similar in the two trial groups. In intervention sites, the checklist was used at admission in 56.8% of observed births and within the first hour after delivery in 74.3% of observed births.

Although adherence in intervention facilities remained significantly higher than in control facilities throughout the trial, adherence in the intervention group had decreased to 61.7% of practices per childbirth at 12 months, or 4 months after coaching ceased. For example, administration of oxytocin soon after delivery decreased by nearly one third (from 79.5% to 53.9%) between 2 and 12 months. Similarly, checklist use had declined in intervention sites at 12 months to 17.4% of cases at admission and 35.1% within 1 hour after delivery. In control sites, birth attendants' average adherence to the 18 measured practices remained low at both 2 and 12 months (41.7% and 43.9%, respectively).

MORTALITY AND MORBIDITY

We found no significant difference between intervention and control facilities in our primary outcome (15.1% in the intervention group and 15.3% in the control group; relative risk, 0.99; 95% confidence interval, 0.83 to 1.18; P=0.90) or in any secondary outcomes (Table 3). Event rates varied widely across facilities, by as much as a factor of 10 (Fig. S2 in the Supplementary Appendix). We found no significant differences between the trial groups in the rates of follow-up care required for women or newborns, hysterectomy, blood transfusion, or interfacility transfer (referral) for women or newborns. In stratified analyses, we observed no significant differences between the groups according to the phase of the intervention (intensive coaching, tapered coaching, and 4 months after the intervention), time of delivery, or in-facility mortality (data not shown).

DISCUSSION

Previous studies have suggested that implementation of the WHO Safe Childbirth Checklist and similarly constructed tools can improve quality of care, but these studies have lacked rigorous data evaluating both adherence to essential birth practices and morbidity and mortality.[8,9,16-18] In this large matched-pair, cluster-randomized, controlled trial in Uttar Pradesh, India, we found that the BetterBirth program — a coaching-based implementation of the WHO Safe Childbirth Checklist — had no significant effect on our primary composite outcome with respect to maternal and perinatal health (nor on any secondary health outcomes), despite significantly higher rates of birth attendants' adherence to essential practices in intervention facilities than in control facilities.

The majority of maternal and neonatal deaths occur around the time of birth and within the first 7 days after delivery[3,36]; thus, interventions to improve early outcomes are desperately needed. The theory of change in the BetterBirth program — that improving the quality of childbirth-related care provided in facilities would translate into improved patient outcomes — reflects basic assumptions underlying current childbirth work in global health. We found that the largely rural population living in this resource-limited setting had a perinatal mortality (47 per 1000) and a maternal morbidity (11.6%) that were much higher than anticipated, and there was wide variation of event rates across facilities.[25] Quality of care in control sites, as measured through birth attendants' adherence to practices, was far lower than previously recognized.[16,25,37-41] Overall, birth attendants in nonintervention facilities performed approximately 40% of measured essential practices in a typical birth, such as appropriate hand hygiene (used in <1% of deliveries) or administration of oxytocin within the first

Table 1. Facility and Participant Characteristics.*

Characteristic	Intervention	Control
Facility characteristics†		
No. of facilities	60	60
Mean annual delivery load (95% CI)	1599 (1486–1712)	1683 (1552–1814)
Functional classification — no. (%)		
Primary health center	23 (38)	23 (38)
Community health center	27 (45)	29 (48)
First referral unit	10 (17)	8 (13)
Mean distance to district hospital (95% CI) — km	29.5 (25.9–33.1)	30.3 (27.2–33.4)
Mean skilled birth attendants per facility (95% CI)	4.4 (4.1–4.7)	4.4 (4.1–4.7)
Maternal characteristics		
No. of women	81,925	79,182
Cluster size‡		
Mean (95% CI)	1365 (1256–1474)	1320 (1229–1411)
Range	515–2697	646–2198
Mean age (95% CI) — yr	25.6 (25.5–25.8)	25.7 (25.5–25.9)
Mean previous pregnancies (95% CI)	2.4 (2.3–2.4)	2.3 (2.3–2.4)
Minutes between admission and delivery		
Mean (95% CI)	200 (184–216)	206 (191–220)
Median (IQR)	105 (33–260)	110 (35–265)
Providers attending delivery — no. (%)		
Doctor	11,115 (13.6)	11,599 (14.6)
Nurse	66,687 (81.4)	64,117 (81.0)
Auxiliary nurse midwife	15,311 (18.7)	14,549 (18.4)
Others	2,633 (3.2)	6,560 (8.3)
No. of offspring — no. (%)		
Singleton	80,402 (98.1)	77,582 (98.0)
Sets of twins	515 (0.6)	533 (0.7)
Sets of triplets	5 (<0.1)	5 (<0.1)
Unknown§	1,003 (1.2)	1,062 (1.3)
Newborn characteristics		
No. of newborns	81,447	78,663
Sex — no. (%)		
Male	40,558 (49.8)	39,063 (49.7)
Female	36,976 (45.4)	36,266 (46.1)
Unknown§	3,913 (4.8)	3,334 (4.2)
Low birth weight — no. (%)¶	22,316 (27.4)	22,728 (28.9)
Preterm birth — no. (%)‖	15,941 (19.6)	17,703 (22.5)

* No significant differences were observed between trial groups ($P<0.05$). CI denotes confidence interval, and IQR interquartile range. Numbers may not sum to 100 because of rounding.
† These characteristics were used for facility matching before randomization.
‡ Cluster size is the total number of pregnant women enrolled per site (cluster).
§ The mother was referred to another facility before the data were captured.
¶ Low birth weight was defined as 2500 g or less.
‖ Preterm birth was defined as fewer than 37 weeks of gestation.

Table 2. Adherence of Birth Attendants to Childbirth Practices and the WHO Safe Childbirth Checklist in 4888 Deliveries across 30 Facilities.

Variable	2 Months				12 Months			
	Intervention	Control	Relative Risk	P Value	Intervention	Control	Relative Risk	P Value
Mean no. of 18 essential birth practices performed (95% CI)	13.1 (12.1–14.0)	7.5 (6.9–8.1)		<0.001	11.1 (10.4–11.8)	7.9 (7.4–8.4)		<0.001
Essential birth practices								
At admission								
No. of women observed	1048	1090			1007	1009		
Birth companion present — no. (%)*	1048 (100)	1087 (99.7)	—†	0.18	1004 (99.7)	1006 (99.7)	1.0 (1.0–1.0)	1.00
Maternal blood pressure taken — no. (%)	606 (57.8)	64 (5.9)	9.8 (3.4–28.3)	<0.001	219 (21.7)	11 (1.1)	19.1 (7.9–46.5)	<0.001
Maternal temperature taken — no. (%)	569 (54.3)	4 (0.4)	132 (26.9–645)	<0.001	153 (15.2)	0	—†	<0.001
Partography started — no. (%)*	10 (1.0)	1 (0.1)	7.3 (1.5–35.6)	0.01	7 (0.7)	0	—†	<0.001
Checklist use — no. (%)	595 (56.8)	1 (0.1)	413 (65.8– 2589)	<0.001	175 (17.4)	0	—†	<0.001
Just before pushing								
No. of women observed	1039	1130			1018	1042		
Hand hygiene — no. (%)*	367 (35.3)	7 (0.6)	53.3 (13.1–217)	<0.001	126 (12.4)	6 (0.6)	19.9 (6.6–60.4)	<0.001
No oxytocin given before delivery — no. (%)	704 (67.8)	333 (29.5)	2.3 (1.3–4.2)	0.007	527 (51.8)	263 (25.2)	2.1 (1.0–4.1)	0.04
Clean towel available — no. (%)*	874 (84.1)	166 (14.7)	5.7 (2.7–12.1)	<0.001	721 (70.8)	314 (30.1)	2.4 (1.3–4.3)	0.006
Clean scissors or blade available — no. (%)*	843 (81.1)	847 (75.0)	1.1 (0.7–1.7)	0.72	1004 (98.6)	1008 (96.7)	1.0 (1.0–1.1)	0.34
Cord tie available — no. (%)*	1035 (99.6)	1122 (99.3)	1.0 (1.0–1.0)	0.48	1009 (99.1)	1039 (99.7)	1.0 (1.0–1.0)	0.25
Mucus extractor available — no. (%)*	990 (95.3)	1057 (93.5)	1.0 (0.9–1.1)	0.73	999 (98.1)	1000 (96.0)	1.0 (1.0–1.1)	0.50
Neonatal bag and mask available — no. (%)*	991 (95.4)	1109 (98.1)	1.0 (0.9–1.1)	0.56	1017 (99.9)	1036 (99.4)	1.0 (1.0–1.0)	0.32
Pads available — no. (%)*	965 (92.9)	502 (44.4)	2.1 (1.3–3.5)	0.005	768 (75.4)	559 (53.6)	1.4 (0.9–2.1)	0.11
Checklist use — no. (%)	255 (24.5)	1 (0.1)	185 (19.7–1738)	<0.001	55 (5.4)	1 (0.1)	37.9 (7.3–196)	<0.001
Within 1 min after delivery								
No. of women observed	1038	1129			1019	1041		
Oxytocin administered — no. (%)*	825 (79.5)	231 (20.5)	3.9 (2.1–7.2)	<0.001	549 (53.9)	154 (14.8)	3.6 (1.8–7.5)	<0.001
Neonatal bag used — no. (%)	47 (4.5)	82 (7.3)	0.6 (0.3–1.4)	0.25	30 (2.9)	49 (4.7)	0.6 (0.3–1.3)	0.19
Birth companion present — no. (%)*	1037 (99.9)	1125 (99.6)	1.0 (1.0–1.0)	0.25	1016 (99.7)	1035 (99.4)	1.0 (1.0–1.0)	0.50
Within 1 hr after delivery								
No. of women observed	1006	1102			1000	1014		
Newborn weight taken — no. (%)*	906 (90.1)	884 (80.2)	1.1 (0.9–1.4)	0.25	947 (94.7)	843 (83.1)	1.1 (1.0–1.3)	0.08
Newborn temperature taken — no. (%)*	433 (43.0)	1 (0.1)	317 (50.4–1989)	<0.001	225 (22.5)	2 (0.2)	91.5 (11.0–758)	<0.001

Skin-to-skin care initiated at birth — no. (%)*‡	794 (78.9)	119 (10.8)	7.3 (2.4–22.0)	<0.001	685 (68.5)	84 (8.3)	8.2 (2.5–27.5)	<0.001
Skin-to-skin care maintained for 1 hr — no. (%)*‡	194 (19.3)	5 (0.5)	38.7 (7.7–194)	<0.001	51 (5.1)	0	—†	0.01
Initiation of breast-feeding — no. (%)*	700 (69.6)	39 (3.5)	19.4 (11.4–33.2)	<0.001	369 (36.9)	47 (4.6)	7.9 (3.7–16.7)	<0.001
Checklist use — no. (%)	747 (74.3)	0	—†	<0.001	351 (35.1)	0	—†	<0.001
At anytime								
No. of women observed	1259	1304			1127	1198		
Maternal temperature taken — no. (%)*	792 (62.9)	4 (0.3)	182 (33.5–993)	<0.001	343 (30.4)	0	—†	<0.001
Maternal blood pressure taken — no. (%)*	854 (67.8)	89 (6.8)	9.9 (3.8–25.8)	<0.001	425 (37.7)	35 (2.9)	12.7 (4.7–34.3)	<0.001
Mother given magnesium sulfate — no. (%)	1 (0.1)	3 (0.2)	0.4 (0.1–2.1)	0.30	1 (0.1)	1 (0.1)	1.1 (0.2–6.6)	0.95

* The practice or supply was included as part of the composite 18-item summary score.
† The relative risk could not calculated owing to 100% or 0% values in one group.
‡ Skin-to-skin care refers to the practice of placing a dried, naked newborn directly on the mother's chest or abdomen and covering both with a blanket.

minute after delivery to reduce hemorrhage (used in <25%).

We found that a coaching-based implementation of the checklist could produce broad-based improvement in the quality of care of facility-based childbirth, a necessary step in effecting improvement in health outcomes. In intervention facilities, birth attendants substantially increased their performance of measured practices. At 2 months, rates of performing recommended practices were significantly higher in intervention sites than in control sites, where some practices were rarely performed. During the coaching intervention, staff at intervention facilities also had correspondingly higher rates of checklist use. However, overall levels of adherence and checklist use diminished after coaching ceased, and the rates of some practices never differed significantly between the intervention and control groups. It is possible that checklist use was not sustained owing to lack of checklist stock, staff belief that they knew the items on the checklist, lack of enthusiasm, or other reasons. We did not collect specific data on reasons for unsustained checklist use. The lack of effect on health outcomes despite improvements in performance of recommended practices challenges assumptions that better practice adherence would directly result in decreased mortality.

One potential interpretation of our findings is that increasing adherence to these practices is not a worthwhile goal, because these practices did not lead to improved outcomes. We strongly believe this conclusion to be false. Each of the practices incorporated in the checklist (and therefore in the BetterBirth program) has its own evidence base, including effectiveness with regard to improving maternal outcomes, improving neonatal outcomes, or both.[13,15] Evidence of the benefits of proper hand hygiene for laboring women and their children, for example, dates back to the 1800s.[42]

Several other factors may have affected the results of the trial. Levels of adherence to essential birth practices in the intervention sites may have been insufficient to affect outcomes: 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. The rate of

Table 3. Primary and Secondary Outcomes of Women and Their Newborns Enrolled in the Trial.*

Outcome	Intervention	Control	Relative Risk (95% CI)†	P Value
Primary composite measure: perinatal death, maternal death, or maternal severe complications within 7 days — %‡	15.1	15.3	0.99 (0.83–1.18)	0.90
Secondary composite measure: perinatal death or maternal death within 7 days — %	4.9	4.7	1.03 (0.89–1.20)	0.67
Perinatal death within 7 days — no./total no. (%)	3839/79,790 (4.8)	3606/77,338 (4.7)	1.03 (0.89–1.20)	0.68
Stillbirth	1513/80,061 (1.9)	1559/77,454 (2.0)	0.94 (0.76–1.16)	0.56
Early neonatal death	2409/78,360 (3.1)	2119/75,851 (2.8)	1.10 (0.95–1.27)	0.19
Maternal death within 7 days — no./total no. (%)	78/79,797 (0.1)	71/77,346 (0.1)	1.11 (0.74–1.53)	0.73
Any maternal severe complication within 7 days — no./total no. (%)	9086/79,342 (11.5)	9037/76,907 (11.8)	0.97 (0.79–1.20)	0.81
Seizures	67/79,706 (0.1)	70/77,257 (0.1)	0.89 (0.57–1.52)	0.76
Loss of consciousness for >1 hr	505/79,649 (0.6)	493/77,196 (0.6)	0.98 (0.70–1.41)	0.97
High fever with foul-smelling vaginal discharge	4073/79,459 (5.1)	3871/77,018 (5.0)	1.02 (0.76–1.38)	0.90
Hemorrhage	5745/79,648 (7.2)	5875/77,198 (7.6)	0.95 (0.77–1.17)	0.61
Stroke	9/79,703 (<0.1)	12/77,257 (<0.1)	0.50 (0.34–1.58)	0.41
Clinical management, reported by facilities — no./total no. (%)				
Cesarean section	1469/80,922 (1.8)	1330/78,120 (1.7)	1.07 (0.72–1.58)	0.74
Maternal referral, before or after delivery	5381/81,925 (6.6)	4779/79,182 (6.0)	1.09 (0.81–1.47)	0.57
Newborn referral	1455/81,925 (1.8)	1186/79,182 (1.5)	1.19 (0.78–1.80)	0.41
Clinical management, reported by patients — no./total no. (%)				
Hysterectomy within 7 days	19/79,705 (<0.1)	18/77,252 (<0.1)	1.00 (0.45–2.13)	0.95
Blood transfusion within 7 days	640/79,697 (0.8)	625/77,254 (0.8)	0.99 (0.69–1.43)	0.97
Woman returning to facility for a health problem within 7 days	2014/79,655 (2.5)	2141/77,220 (2.8)	0.91 (0.76–1.10)	0.32
≥1 Newborn returning to facility for a health problem within 7 days	4474/77,419 (5.8)	4722/75,117 (6.3)	0.92 (0.78–1.08)	0.30

* The term "within 7 days" refers to the first 7 days after delivery. Maternal severe complications included any of the following: seizures, loss of consciousness for more than 1 hour, high fever with foul-smelling vaginal discharge, hemorrhage, or stroke.
† Relative risks are for the intervention group as compared with the control group and are adjusted for clustering and matching. The denominators shift owing to missing data on outcomes.
‡ In secondary analyses, a Rao–Scott test with 3 degrees of freedom resulted in a P value of 0.88.

use of magnesium sulfate was no higher in the intervention facilities than in the control facilities. Persistent gaps in technical skills, management of complications, the quality and quantity of supplies and medicines, access to supportive management, and systems-level accountability — mostly unmeasured — could also have had a substantial effect on the ability to improve health outcomes. Factors that were not targeted by the BetterBirth program may have also limited its effects, including women's underlying health and nutrition status, the quality of prenatal and postnatal care, and the quality of referral care for those with more complex needs.

Despite the careful design of the trial, limitations remain. Data on observed adherence to practices were available for only a select subsample — 4888 of the 157,689 deliveries at 30 of the 120 facilities. For practical reasons, these observations were performed at nonrandomly selected sites and during daytime hours, which potentially limits generalizability to the unobserved births. The measured levels of adherence may have been misleading, if staff practiced

much differently when unobserved.[43] In addition, although measures of maternal complications reported by the patient had previously been validated by other studies,[27-30] misreporting is possible, because we did not verify reported complications with medical records or physical examination. Some trial staff were aware of the facility assignments; however, the staff members collecting the majority of outcome data were unaware of the facility assignments. Furthermore, 7-day outcome information was collected between 8 and 42 days after delivery. Thus, recall bias may have affected the reporting of maternal morbidity but is not likely to have affected the reporting of mortality. Whereas the target sample size was 171,964, the final sample size was 157,689 (91.7% of the target). However, our event rate was significantly higher than expected. In a post hoc power analysis, we had 80% power to detect a 9.08% relative difference in the rate of the primary composite outcome between the two groups. Finally, the implementation of the trial in only one region of India limits the generalizability of the findings. Whether the BetterBirth program could achieve improvement in health outcomes in a setting with, for example, higher baseline adherence to practices, a different health care system or facility-level organization, or a different patient population is not known.

We found that a coaching-based implementation of the WHO Safe Childbirth Checklist had no significant effect on adverse maternal and perinatal health outcomes, despite positive effects on essential birth practices. High-quality research on large-scale programs to improve childbirth must continue to measure both processes and outcomes of care because we currently lack a complete understanding of the complex interaction among quality of care, context, and outcomes.

Supported by the Bill and Melinda Gates Foundation.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the governments of India and Uttar Pradesh for collaboration and support to conduct this trial in public health facilities; the facility staff, women, and newborns for their participation in the trial; the members of the scientific advisory committee, who contributed crucial guidance to the development of this trial protocol (Himanshu Bhushan, Zulfiqar Bhutta, Waldemar Carlo, Vinita Das, Amod Kumar, Matthews Mathai, Packirisamy Padmanbhan, Vinod Paul, and Rajiv Tandon); and Jonathan Spector, Steve Ringer, Robyn Churchill, Gary King, Grace Galvin, Atul Kapoor, and the past and current members of the BetterBirth study teams in Boston and Uttar Pradesh.

APPENDIX

The authors' affiliations are as follows: Ariadne Labs–Brigham and Women's Hospital, Harvard T.H. Chan School of Public Health (K.E.A.S., M.M.D., D.E.T., S.L., A.A.G.), the Divisions of Global Health Equity (K.E.A.S.) and General Internal Medicine (S.L.) and the Department of Surgery (A.A.G.), Brigham and Women's Hospital, the Department of Medicine, Harvard Medical School (K.E.A.S.), and the Department of Health Policy and Management, Harvard T.H. Chan School of Public Health (A.A.G.) — all in Boston; Feinberg School of Medicine, Northwestern University, Chicago (L.R.H.); Population Services International (V.P.S., R.S., N.S.) and the Community Empowerment Lab (V.K.), Lucknow, and Jawaharlal Nehru Medical College, Belgaum (B.S.K.) — all in India; Population Services International, Washington, DC (R.F.); and the World Health Organization, Geneva (N.D.-K.).

REFERENCES

1. GBD 2015 Child Mortality Collaborators. Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388:1725-74.
2. Alkema L, Chou D, Hogan D, 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. Lancet 2016;387:462-74.
3. Kassebaum NJ, Bertozzi-Villa A, Coggeshall MS, et al. Global, regional, and national levels and causes of maternal mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014;384:980-1004.
4. WHO recommendations on postnatal care of the mother and newborn. Geneva: World Health Organization, 2013.
5. Aminu M, Unkels R, Mdegela M, Utz B, Adaji S, van den Broek N. Causes of and factors associated with stillbirth in low- and middle-income countries: a systematic literature review. BJOG 2014;121:Suppl 4:141-53.
6. Miller S, Abalos E, Chamillard M, et al. Beyond too little, too late and too much, too soon: a pathway towards evidence-based, respectful maternity care worldwide. Lancet 2016;388:2176-92.
7. Dettrick Z, Firth S, Jimenez Soto E. Do strategies to improve quality of maternal and child health care in lower and middle income countries lead to improved outcomes? A review of the evidence. PLoS One 2013;8(12):e83070.
8. Magge H, Anatole M, Cyamatare FR, et al. Mentoring and quality improvement strengthen integrated management of childhood illness implementation in rural Rwanda. Arch Dis Child 2015;100:565-70.
9. Bradley J, Jayanna K, Shaw S, et al. Improving the knowledge of labour and delivery nurses in India: a randomized controlled trial of mentoring and case sheets in primary care centres. BMC Health Serv Res 2017;17:14.
10. Goudar SS, Somannavar MS, Clark R, et al. Stillbirth and newborn mortality in India after helping babies breathe training. Pediatrics 2013;131(2):e344-e352.
11. Msemo G, Massawe A, Mmbando D, et al. Newborn mortality and fresh stillbirth rates in Tanzania after helping babies breathe training. Pediatrics 2013;131(2):e353-e360.
12. Salam RA, Mansoor T, Mallick D, Lassi ZS, Das JK, Bhutta ZA. Essential

childbirth and postnatal interventions for improved maternal and neonatal health. Reprod Health 2014;11:Suppl 1:S3.
13. Spector JM, Lashoher A, Agrawal P, et al. Designing the WHO Safe Childbirth Checklist program to improve quality of care at childbirth. Int J Gynaecol Obstet 2013;122:164-8.
14. Quality of care: a process for making strategic choices in health systems. Geneva: World Health Organization, 2006.
15. The Safe Childbirth Checklist program background paper: International Consultation Meeting. Geneva: World Health Organization, 2009.
16. Kumar S, Yadav V, Balasubramaniam S, et al. Effectiveness of the WHO SCC on improving adherence to essential practices during childbirth, in resource constrained settings. BMC Pregnancy Childbirth 2016;16:345.
17. Patabendige M, Senanayake H. Implementation of the WHO safe childbirth checklist program at a tertiary care setting in Sri Lanka: a developing country experience. BMC Pregnancy Childbirth 2015;15:12.
18. Spector JM, Agrawal P, Kodkany B, et al. Improving quality of care for maternal and newborn health: prospective pilot study of the WHO Safe Childbirth Checklist program. PLoS One 2012;7(5):e35151.
19. Semrau KEA, Hirschhorn LR, Kodkany B, et al. Effectiveness of the WHO Safe Childbirth Checklist program in reducing severe maternal, fetal, and newborn harm in Uttar Pradesh, India: study protocol for a matched-pair, cluster-randomized controlled trial. Trials 2016;17:576.
20. Marx Delaney M, Maji P, Kalita T, et al. Improving adherence to essential birth practices using the WHO Safe Childbirth Checklist with peer coaching: experience from 60 public health facilities in Uttar Pradesh, India. Glob Health Sci Pract 2017;5:217-31.
21. Kara N, Firestone R, Kalita T, 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. Glob Health Sci Pract 2017;5:232-43.
22. Gass JD Jr, Misra A, Yadav MNS, et al. Implementation and results of an integrated data quality assurance protocol in a randomized controlled trial in Uttar Pradesh, India. Trials 2017;18:418.
23. Office of the Registrar General and Census Commissioner. Census of India 2011 provisional population totals. Paper 2. Vol. 1. New Delhi: Government of India, 2012.
24. Office of the Registrar General & Census Commissioner. Sample registration system statistical report 2014. New Delhi: Government of India, 2015.
25. Office of the Registrar General & Census Commissioner Vital Statistics Division. Annual health survey 2012-13 fact sheet Uttar Pradesh. New Delhi: Ministry of Home Affairs, Government of India, 2013.
26. Hirschhorn LR, Semrau K, Kodkany B, et al. Learning before leaping: integration of an adaptive study design process prior to initiation of BetterBirth, a large-scale randomized controlled trial in Uttar Pradesh, India. Implement Sci 2015;10:117.
27. Souza JP, Cecatti JG, Pacagnella RC, et al. Development and validation of a questionnaire to identify severe maternal morbidity in epidemiological surveys. Reprod Health 2010;7:16.
28. Ronsmans C, Achadi E, Cohen S, Zazri A. Women's recall of obstetric complications in south Kalimantan, Indonesia. Stud Fam Plann 1997;28:203-14.
29. Filippi V, Ronsmans C, Gandaho T, Graham W, Alihonou E, Santos P. Women's reports of severe (near-miss) obstetric complications in Benin. Stud Fam Plann 2000;31:309-24.
30. Evaluating the quality of care for severe pregnancy complications: the WHO near-miss approach for maternal health. Geneva: World Health Organization, 2011.
31. Taljaard M, Donner A, Villar J, et al. Intracluster correlation coefficients from the 2005 WHO Global Survey on Maternal and Perinatal Health: implications for implementation research. Paediatr Perinat Epidemiol 2008;22:117-25.
32. Rao JNK, Scott AJ. On simple adjustments to chi-square tests with sample survey data. Ann Stat 1987;15:385-97.
33. Carter RE, Lipsitz SR, Tilley BC. Quasi-likelihood estimation for relative risk regression models. Biostatistics 2005;6:39-44.
34. Lipsitz SR, Dear KB, Zhao L. Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. Biometrics 1994;50:842-6.
35. Rader KA, Lipsitz SR, Fitzmaurice GM, Harrington DP, Parzen M, Sinha D. Bias-corrected estimates for logistic regression models for complex surveys with application to the United States' Nationwide Inpatient Sample. Stat Methods Med Res 2017;26:2257-69.
36. Belizán JM, McClure EM, Goudar SS, et al. Neonatal death in low- to middle-income countries: a global network study. Am J Perinatol 2012;29:649-56.
37. Sarin E, Kole SK, Patel R, et al. Evaluation of a quality improvement intervention for obstetric and neonatal care in selected public health facilities across six states of India. BMC Pregnancy Childbirth 2017;17:134.
38. Bhattacharyya S, Issac A, Rajbangshi P, Srivastava A, Avan BI. "Neither we are satisfied nor they" — users and provider's perspective: a qualitative study of maternity care in secondary level public health facilities, Uttar Pradesh, India. BMC Health Serv Res 2015;15:421.
39. Sudhinaraset M, Treleaven E, Melo J, Singh K, Diamond-Smith N. Women's status and experiences of mistreatment during childbirth in Uttar Pradesh: a mixed methods study using cultural health capital theory. BMC Pregnancy Childbirth 2016;16:332.
40. Stanton CK, Deepak NN, Mallapur AA, et al. Direct observation of uterotonic drug use at public health facility-based deliveries in four districts in India. Int J Gynaecol Obstet 2014;127:25-30.
41. Sudhinaraset M, Beyeler N, Barge S, Diamond-Smith N. Decision-making for delivery location and quality of care among slum-dwellers: a qualitative study in Uttar Pradesh, India. BMC Pregnancy Childbirth 2016;16:148.
42. Semmelweis IF. The etiology, concept, and prophylaxis of childbed fever. Carter KC, trans., ed. Madison: University of Wisconsin Press, 1983.
43. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014;67:267-77.