SSM - Population Health 24 (2023) 101495

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

The effect of the WHO Safe Childbirth Checklist on essential delivery practices and birth outcomes: *Evidence from a pair-wise matched randomized controlled trial in Pakistan*

Jana Kuhnt [a], Ashfa Hashmi [b], Sebastian Vollmer [c,*]

[a] German Institute of Development and Sustainability (IDOS), Tulpenfel 6, 53113, Bonn, Germany
[b] GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit), University of Göttingen, Germany
[c] University of Goettingen, Center for Modern Indian Studies, Waldweg 26, 37073 Göttingen, Germany

ARTICLE INFO

Keywords:
Safe Childbirth Checklist (SSC)
Clustered pair-wise matched randomized controlled trial
Pakistan
Maternal and newborn health

ABSTRACT

We study the effect of the Safe Childbirth Checklist (SCC) – a tool developed by the WHO to improve the quality of delivery care – on a range of provider- and patient-level outcomes. We conducted a clustered pair-wise matched randomized controlled trial among 166 health providers in two districts of Pakistan. This included primary and secondary health facilities as well as non-facility based rural health workers. We do not find positive effects on health outcomes, but on the adherence to some essential delivery practices, mostly to those conducted during the patient's admission to the delivery ward. We also find increased rates of referrals to higher-level facilities.

1. Introduction

Globally, still 2.4 million babies die every year during their neonatal period (UNICEF, 2023; World Health Organization, 2022). About 40% of those deaths occur within the first 24 hours after birth and up to 75% take place within the first week of life (World Health Organization, 2022). Every year there are 832 thousand intrapartum-related stillbirths (UNICEF et al., 2020) and each day, about 810 women die from preventable complications linked to pregnancy or birth (World Health Organization, 2023). The large majority (94%) of all maternal deaths occur in low- and middle-income countries (World Health Organization, 2023) and within a rather narrow timeframe: more than 40% of maternal deaths happen during the delivery process and 45% of the maternal deaths occurring after the birth process take place within the first 24 hours after delivery (World Health Organization & UNICEF, 2010). Furthermore, there are more than one million children each year that survive birth complications, but develop complication-related illnesses, which often imply learning difficulties and other disabilities. Over 300 million women have long- and/or short-term pregnancy- or childbirth-related sicknesses that the family is unprepared for and that are often related to adverse effects on the newborn's health and survival chances (World Health Organization, 2005). Many of the deaths and morbidities are due to easily preventable causes.

Providing high-quality, evidence based delivery care presents a great opportunity to reduce the maternal and neonatal mortality and morbidity burden (Spector et al., 2013). But the gap between the knowledge and implementation of best practices is large (Cabana et al., 1999; Cochrane et al., 2007; Courtenay-Hall & Rogers, 2002; Frederiks et al., 2015; Michie et al., 2005; Sligo & Jameson, 2000). In a variety of settings, including aviation, product manufacturing and health, checklists have successfully been used to increase adherence to best practices, ensure standard operating procedures, and herewith reduce human error (e.g., Hales & Pronovost, 2006). Checklists bundle essential tasks into a practical format consisting of actionable items and, hence, help the users to remember complex or neglected tasks and reduce the possible 'information overload' (e.g., Borchard et al., 2012; Haugen et al., 2015; Workman et al., 2007). In situations characterized by high levels of cognitive load – the amount of mental activity imposed – the successful execution of certain tasks might be interrupted or impaired (e.g., Burgess, 2010). In particular, unexpected, complex or stressful events require a high cognitive capacity that can impair the short-term memory (Deck & Jahedi, 2015; Hoffman et al., 2011), reduce the quality of decision-making and, thus, increase cognitive errors (Croskerry, 2002; Hales & Pronovost, 2006; Kramer & Drews, 2017; Lichand & Mani, 2020).

The integration of checklists into clinical practice has been shown to

* Corresponding author.
E-mail addresses: jana.kuhnt@idos-research.de (J. Kuhnt), ashfahashmi@gmail.com (A. Hashmi), svollmer@uni-goettingen.de (S. Vollmer).

https://doi.org/10.1016/j.ssmph.2023.101495
Received 21 December 2022; Received in revised form 29 June 2023; Accepted 18 August 2023
Available online 23 August 2023

reduce deaths and complications in intensive care medicine and surgery in low-, middle- as well as in high-income countries (Borchard et al., 2012; Haugen et al., 2015; World Health Organization, 2008). Following these successful examples, the World Health Organization (WHO) commissioned the development of the WHO Safe Childbirth Checklist (SCC) – a checklist to be used around the delivery process. The items on the SCC (see Figures A1 – A6 in the Appendix) address the major causes of maternal and neonatal mortality and birth complications in low- and middle-income countries. In this paper, we study whether the adoption of the SCC among public health providers in Pakistan leads to greater adherence to essential birth practices by health practitioners and improved child and maternal health outcomes.

Following a large call by the WHO to extend the evidence on the newly developed SCC, several studies were implemented in various countries (Hirschhorn et al., 2015; Kabongo et al., 2017; Kumar et al., 2016; Kumari et al., 2016; Patabendige & Senanayake, 2015). We are aware of two other large randomized controlled trials evaluating the effect of the SCC. The Better Birth Trial situated in Uttar Pradesh, India, used a pair-wise matching approach to randomly implement the checklist together with an intensive peer-to-peer coaching (Kara et al., 2017; Semrau et al., 2016, 2017). The authors find significantly increased adherence to essential childbirth practices but not effects on maternal and neonatal mortality (Semrau et al., 2017). Delaney and co-authors conduct a post-hoc analysis of data from the same trial and find reduced perinatal mortality among facilities with lower birth volume in the treatment group (Delaney et al., 2019). Kaplan and co-authors implemented the SCC with medium-intensity coaching in a matched-group cluster randomized clinical trial in Aceh, Indonesia (Kaplan et al., 2021). They find increased adherence to essential childbirth practices as well as reduced stillbirths and neonatal mortality.

We add to this literature by implementing the first SCC intervention that in addition to health facilities included health workers who were not based in facilities (community midwives and privately working lady health visitors). Especially in rural areas in many low and lower-middle income countries they are often the first entry point for patients to the formal (public) health care system. They work in areas with high maternal and neonatal mortality and morbidity rates and might, hence, be important actors to tackle this burden. Further, we implemented the SCC with less intensive monitoring compared to previous studies in order to better reflect the conditions under which this type of intervention would be rolled out by governments in resource-constrained environments.

2. Material and methods

2.1. Sample and randomization

The study was conducted in two districts, Haripur and Nowshera, of the province Khyber Pakhtunkhwa (KP) in the Northwest of Pakistan. Of a total population of 900 000 in Haripur and 1.2 million in Nowshera there were an estimated number of 30 701 pregnant women in Haripur and respectively 40 800 in Nowshera in the year 2010 (projected numbers based on data from PDHS 2006-07 and DHIS Cell-EDoH Office). The study population consisted of all public health providers in both districts that offer delivery services, including secondary- and primary-level health facilities (HFs), community midwives (CMWs), and private lady health visitors (LHVs). Eligibility criteria for inclusion in the study differed between providers: HFs were included if they had conducted at least four deliveries in the past month or on average at least four deliveries per month during the past 12 months. For political reasons, all CMWs were included. LHVs were included if they had on average at least three deliveries per month in the past 12 months. The eligible sample consisted of 166 health providers in two districts: 17 HFs, 102 CMWs, and 47 LHVs. Each provider represents a cluster, conducting multiple deliveries.

Providers in the eligible sample were randomly assigned into treatment and control group using a pair-wise matching approach. As shown by Bruhn and McKenzie (2009), pair-wise matching can be a useful randomization method in small samples especially for persistent outcomes to increase balance and power. This is supported by Imai et al. (2009) who argue that efficiency gains by applying pre-randomization matching are especially large in cluster-randomized setups, where clusters are of unequal sizes. We grouped the eligible sample into six strata based on health cadre type (HFs, CMWs, LHVs) and the districts (Nowshera and Haripur). Within these six strata, we created matched pairs using the methodology proposed by Greevy et al. (2004) and the R-code developed by Lu et al. (2011). Matching variables included the provider's total number of deliveries as well as their mortality, referral and birth complication rates in 2015. Data for these variables was collected in a baseline survey of health providers between January to February 2016. We then randomly assigned one health provider in each pair to the treatment group (60 health providers) and one to the control group (60 health providers). For providers with fully or partially lacking baseline information (18.1% of eligible sample) we used a pure randomization strategy to allocate them to either group within the strata.

Several providers dropped out during the study period as they did not attend the training or did not conduct any deliveries. Endline data was collected among all remaining providers in April to May 2017. We further excluded from the analysis 15 CMWs who did not conduct any deliveries in the six months prior to endline data collection and the largest HF which did not have an adequate match, leading to a final analytical sample of 12 HFs, 32 LHVs, and 72 CMWs. Table A1 in the Appendix depicts the sample selection process.

We checked the balance of the final analytical sample and did not find systematic differences between providers in the control and treatment groups (see Table 1). Balance of the eligible sample as well as for subsamples is shown in Table A2 to Table A4 in the Appendix.

We conducted power calculations prior to the intervention for several indicators to judge whether power would be sufficient to detect statistically significant effects. The study region was determined by the implementation partner's operational area, limiting the study population to the providers in this area. Due to the limited availability of baseline and administrative data for outcomes of interest, we relied on endline data of the control group to calculate minimal detectable effects (MDE). We assumed a significance level of 5%, a total of 6,000 deliveries, and used the standard deviations and inter-cluster correlations from the endline data. At a treatment uptake of 90%, we would be able to detect a change of birth complications of 4.2 percentage points and changes in referrals of newborns of 2.7 percentage points. The sample size was found to be insufficiently large to capture changes in maternal or neonatal mortality rates.

2.2. Intervention

A standardized basic four-day BEmONC (Basic Emergency Obstetric and Newborn Care) training covering all aspects of the SCC items was offered from March to November 2016 to all health providers. Only 7% of providers refused to attend the training and were excluded from the study. A comparative assessment prior and after the BEmONC training showed increased short-term skill gains of all participants. After the training, 93% of providers in the treatment group and 87% of providers in the control group scored more than 80% in the knowledge assessment, as compared to 3% and 2% prior to the training, respectively (Zafar, 2017).

Following the four-day BEmONC training, the checklist was introduced to the providers of the treatment group within a standardized one-day training. The aim of this one-day training was primarily to ensure sufficient understanding of the checklist items and to establish sufficient motivation for the checklist to ensure its use in practice. After the training, the participants received copies of the checklist and a checklist poster to hang up in their delivery rooms.

Table 1
Balance table on structural variables across pooled analytical sample.

Variable	Pooled Sample			Control		Treatment		p-value difference in means *t*-test
	N	Mean	SD	Mean	SD	Mean	SD	
Cadre type (base: CMW)	116	1.483		1.464		1.500		0.777
LHV		0.276		0.321		0.233		
HF		0.103		0.071		0.133		
District	116	1.578		1.625		1.533	0.503	0.322
Rural	116	0.871		0.857		0.883		0.679
# of delivery staff	12	4.167	1.749	4.500	1.732	4.000	1.852	0.647
# of doctors	12	2.000	3.303	4.500	5.196	0.750	0.463	0.157
Position leading delivery services	12	2.250	0.754	2.250	0.957	2.250	0.707	1.000
Open 24 h	12	0.583		0.750		0.500		0.432
BEmONC	116	0.017		0.018		0.017		0.961
CEmONC	116	0.009		0.000		0.017		0.320
International Funding	10	0.800		0.750		0.833		0.785
Access emergency transport	110	0.627		0.685		0.571		0.221
Electricity 24/7	102	0.559		0.640		0.481		0.107
Clean, running water	108	0.926		0.962		0.893		0.169
Experience of staff (yrs)	116	9.491	7.920	10.036	8.393	8.983	7.487	0.479
Educational level of staff (1–6 – "Primary" – "Postgraduate")	116	4.241	0.992	4.268	1.053	4.217	0.940	0.783
Access to facility drugstore	31	0.677		0.667		0.692		0.885
Provider was born in area	104	0.596		0.615		0.577		0.693
Deliveries at patient home	104	0.798		0.750		0.846		0.226
At patient home: Running water (1–4 – "Never" – "Always")	91	1.308	0.571	1.326	0.634	1.289	0.506	0.758
At patient home: Electricity (1–4 – "Never" – "Always")	93	2.172	0.842	2.217	0.867	2.128	0.824	0.610

This table refers to the pooled final analytical sample used in the main regressions of this paper and is based on the endline provider survey. Standard errors are clustered at the provider level. Mean refers to the mean for integer variables and the share for binary variables, SD gives the standard deviation for integer variables, N is the number of observations in the pooled sample, i.e. providers.

Table 2
ITT and CACE results – Patient health outcomes

	Maternal Complication (Dummy = 1 if occurred)	Newborn Complication (Dummy = 1 if occurred)	Birth Complication (Dummy = 1 if occurred)	Referral of Mother (Dummy = 1 if referral occurred)	Referral of Newborn (Dummy = 1 if referral occurred)	Initiation of breastfeeding (Dummy = 1 if initiated)
	(1)	(2)	(3)	(4)	(5)	(6)
UNADJUSTED						
ITT	**−0.020**	**−0.014**	**−0.029**	**−0.006**	**0.024**	**0.090***
SE	[0.016]	[0.013]	[0.020]	[0.025]	[0.017]	[0.053]
p-value	(0.190)	(0.270)	(0.162)	(0.792)	(0.153)	(0.091)
RI p-value	(0.236)	(0.279)	(0.237)	(0.879)	(0.148)	(0.151)
ADJUSTED						
ITT	**−0.014**	**−0.009**	**−0.019***	**0.002**	**0.025****	**0.082***
SE	[0.010]	[0.007]	[0.011]	[0.015]	[0.011]	[0.043]
p-value	(0.148)	(0.216)	(0.096)	(0.904)	(0.026)	(0.059)
RI p-value	(0.228)	(0.480)	(0.187)	(0.958)	(0.164)	(0.133)
CACE	**−0.016**	**−0.009**	**−0.021**	**0.002**	**0.031****	**0.087***
SE	[0.017]	[0.014]	[0.016]	[0.029]	[0.015]	[0.046]
p-value	(0.353)	(0.514)	(0.182)	(0.947)	(0.038)	(0.058)
Mean dep. variable	0.0240	0.0175	0.0335	0.0253	0.0060	0.8452
N	3609	3609	3609	3467	1424	1130

***p < 0.01, **p < 0.05, *p < 0.1.
Data is from provider records. Point estimates of logit regressions are displayed as average marginal effects. Robust standard errors (SE) are clustered at the provider-level. Adjusted regressions include strata dummies and a dummy indicating if a facility is open 24/7. All regressions are based upon the analytical sample. ITT gives the Intent-to-Treat effects and CACE the Complier-Average-Causal effects, where the 'number of filled checklist' is used as complier indication (IV Regression). RI p-value is calculated using randomization inference. Mean dependent variable is based upon the values of the control group over the entire study period. N refers to the number of deliveries.

The introduction of the SCC through the training was accompanied by a light monitoring approach. All providers, in treatment and control groups, received monthly visits following a standardized procedure. During these visits, data on deliveries and checklists (in the treatment group) were collected, data was checked for completeness and accuracy, and new copies of the checklist were provided to providers in the treatment group. In HFs in the treatment group, additional internal monitoring was put in place by assigning a Checklist Quality Coordinator (CQC), a designated member of staff. The CQC's tasks were to coordinate use of the SCC in the HFs, to provide support to staff members, and to train new staff members on the checklist. We refrained from intensive coaching or support systems to test the SCC as low-cost and easy-to-implement tool.

Due to the design of the intervention, providers were only blinded to

their assignment until after completion of the BEmONC training. A timeline of the study can be found in Fig. 1.

2.3. Data sources

Provider records. Following the trainings, we collected a rich dataset of patient-level delivery data based on records filled by health provider staff, including information on birth complications, referrals, deaths, and patient health and pregnancy history. The analysis includes delivery data from July 2016 to August 2017 (14 months) for CMWs and LHVs and from November 2016 to July 2017 (9 months) for the HFs due to delayed trainings. The data captured 4,520 deliveries from 125 providers (959 deliveries by 76 CMWs, 426 by 36 LHVs, 3,135 by 13 HFs). Consistency and quality checks on this data was performed during monthly monitoring visits to each provider.

Observational data. To observe the delivery practices of each provider, trained 'silent' observers documented the essential practices followed by health staff in a standardized tool. One observer always observed the entire delivery process for one delivery patient and recorded essential practices, information on communication among staff and towards the patient as well as referral behaviour. Over a period of eight weeks (June to July 2017), 312 deliveries (246 in seven HFs, 61 from nine CMWs, 5 from two LHVs; 99 in control group, 213 in treatment group) were observed. The observed providers were a non-randomly selected convenience subsample of the larger study sample. Still, this dataset gives us important insights on practices conducted by the providers and hence, investigates the most direct impact of the SCC (see Table 4 for all outcome variables).

Provider survey. We conducted a baseline (January to February 2016) and endline (April to May 2017) survey at the provider level. The baseline survey captured the number of deliveries conducted by each provider, which was used to identify the eligible sample, and variables for the pair-wise matching approach. The endline survey additionally included self-reported perceptions of the providers (see Table 5).

2.4. Outcome variables

Patient health outcomes. The primary outcomes are patient health outcomes based on provider records. These include indicators for the occurrence of maternal, newborn, and general birth complications, indicators whether the mother or newborn was referred to a higher-level, better equipped provider as a response to actual or expected complications, and an indicator for the initiation of breastfeeding (see Table 3). Within complications, we only consider complications that can be directly linked to SCC items, such as maternal infections and postpartum hemorrhage.

Adherence to essential practices. Secondary outcomes include the observed delivery practices conducted by each provider, capturing a pre-requisite for impacting the primary outcome, patient health. The outcomes follow the structure of the SCC, with one outcome for each essential practice performed. Additionally, we analyse the number of practices adhered to by each pause point (see Table 4).

Self-reported provider behavior and attitudes. Further secondary outcomes capture self-reported behaviour and attitudes of providers and are based on provider survey data. These outcomes include indicators of the SCC as a reminder and awareness-raising tool, documentation practices, information exchange, knowledge gain, the perceived error rate, and provider empowerment (see Table 5).

2.5. Estimation strategy

By comparing the averages in the outcome variables in treatment and control groups, we can measure the intent-to-treat effect (ITT), using the following OLS equation (here specified for provider-level data from HFs):

$$Y_{fi} = \beta_1 + \beta_2 Treat_f + \beta_3 X_i^{'} + \beta_4 Strata^{'} + u_{fi} \quad (1)$$

Y_{fi} is the outcome for delivery staff i from HF f at endline. β_1 is the intercept and $Treat_f$ is a dummy taking the value of one if the health staff works at a treatment HF. β_2 is the intent-to-treat effect and herewith our main coefficient of interest. To consistently estimate treatment effects and to increase precision of our estimates we account for the stratification procedure applied by including the strata dummies as covariates in our analysis (represented by $Strata^{'}$) as suggested by (Bruhn & McKenzie, 2009). A successful randomization eliminates potential bias from omitted variables. To increase the precision of our estimates, we add a vector of control variables at the HF level (represented by $X_i^{'}$). Control variables include indicator variables for whether the health facility is open 24 h 7 days per week to differentiate higher-level from lower-level facilities, whether the provider is located in a rural or urban setting (as perceived by the enumerator), for districts and for the cadre type (CMW, LHV or HF). Further, we include categorical variables capturing the average professional experience of the staff in completed years, the highest staffing position present during the observed pause point (Midwife, Nurse, Lady Health Visitor, Female Medical Technician, Female Medical Officer, Doctor), a composite measure of the availability of medication and equipment at baseline as well as the number of deliveries conducted in 2015. The error term u_{fi} is clustered at the HF level and in cases with small number of clusters we employ wild cluster bootstrap procedures following Cameron et al. (2008).

To estimate health outcomes (birth complications and referrals) we use the logit-link function to ensure that the predictions lie in the support of the [0; 1] domain. This is especially important when dealing with rare events (here event occurrence of 1–2%) as they lie closely to the lower bound of support of the predicted probabilities. We estimate the logit models using a similar form as equation *(1)* with outcomes at the patient-level. For all estimations we report average marginal effects (AME).

The ITT includes providers that regularly used the SCC and those that never or only selectively applied the SCC. The ITT reflects the true treatment effect in a real-world setting, where full monitoring is

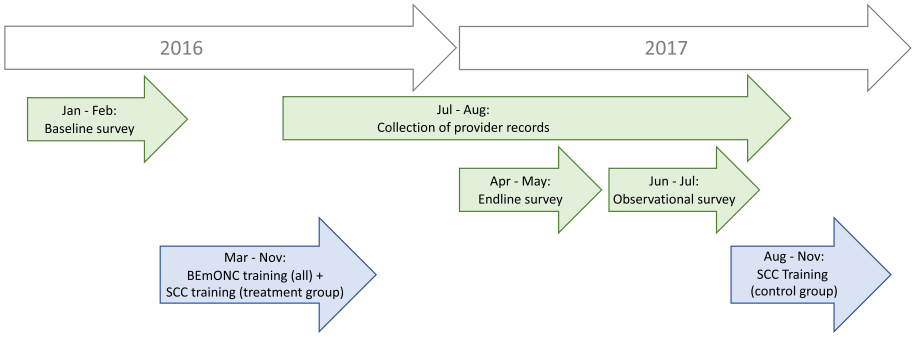

Fig. 1. Study timeline.

Table 3
TT and CACE for CMW sample – Patient health outcomes.

	Maternal Complication (Dummy = 1 if occurred)	Newborn Complication (Dummy = 1 if occurred)	Birth Complication (Dummy = 1 if occurred)	Referral of Mother (Dummy = 1 if referral occurred)	Referral of Newborn (Dummy = 1 if referral occurred)	Initiation of breastfeeding (Dummy = 1 if initiated)
	(1)	(2)	(3)	(4)	(5)	(6)
UNADJUSTED						
ITT	**−0.0510***	**−0.0406**	**−0.0715***	**0.00179**	**0.00552**	**0.139****
SE	[0.029]	[0.026]	[0.037]	[0.030]	[0.013]	[0.065]
p-value	(0.075)	(0.117)	(0.052)	(0.953)	(0.670)	(0.037)
RI p-value	(0.112)	(0.090)	(0.088)	(0.964)	(0.651)	(0.052)
ADJUSTED						
ITT	**−0.0483***	**−0.0318*****	**−0.0656***	**−0.00799**	**0.00461**	**0.111****
SE	[0.029]	[0.012]	[0.034]	[0.031]	[0.014]	[0.044]
p-value	(0.097)	(0.006)	(0.052)	(0.799)	(0.738)	(0.013)
RI p-value	(0.154)	(0.100)	(0.112)	(0.813)	(0.731)	(0.057)
CACE	**−0.0647**	**−0.0586**	**−0.0918**	**−0.00877**	**0.00510**	**0.119****
SE	[0.0641]	[0.118]	[0.128]	[0.0336]	[0.0154]	[0.047]
p-value	(0.31307)	(0.61867)	(0.47305)	(0.79423)	(0.73950)	(0.012)
Mean dep. variable	0.06491	0.04260	0.08519	0.04941	0.01014	0.81140
N	899	899	899	946	932	818

***p < 0.01, **p < 0.05, *p < 0.1.
Data is from provider records of CMWs. Point estimates of logit regressions are displayed as average marginal effects. Robust standard errors (SE) are clustered at the provider level. Adjusted regressions include a district dummy and a rural/urban dummy for the LHV and CMW sample and for the HF sample instead of the latter a dummy indicating if a facility is open 24/7. All regressions are based upon the analytical samples. ITT gives the Intent-to-Treat effects and CACE the Complier-Average-Causal effects, where the 'number of filled checklist' is used as complier indication (IV Regression). RI p-value is calculated using randomization inference. Mean dependent variable is based upon the values of the control group over the entire study period. N refers to the number of deliveries.

impossible and where non-compliers are likely to exist in the treatment group. In order to better understand the possible treatment effects of very high compliance rates, we additionally estimate the complier average causal effect (CACE) (Imai et al., 2009).

We approximate compliance with three different measures: First, the observed use of the SCC. Here, compliance ranges between 42% and 48% across the different stages observed during childbirth. Second, we asked providers to name the last item on the checklist to assess their SCC knowledge and proxy SCC use. Only 20% of the providers correctly answered this question. Third, we assess the rate of filled checklists over total deliveries conducted by the health providers on a monthly basis. This continuous indicator is varying at the provider level and indicates the average monthly rate of filled SCCs. Highest compliance rates are among LHVs and CMWs (see Figure A7 in the Appendix). All three compliance measures fulfil the relevance criterion in the first stage regression of the CACE estimation with an F-statistics of at least 10 (Stock & Watson, 2006).

In addition to standard p-values generated through classical inference, we report p-values for our ITT estimates generated through randomization inference using the *ritest* Stata command (Heβ, 2017).

2.6. Ethical considerations

Ethical clearance was provided by the National Bioethics Committee (NBC) Pakistan in October 2015 after a detailed study protocol was revised and approved by them. The trial was not registered in a trial registry prior to implementation. To reduce a potential disadvantage of those providers randomly assigned to the control group, a phase-in design was chosen. Providers in the control group received a training on the SCC after endline data collection in 2017. Informed consent was signed by individuals in all data collection efforts. Providers were asked for consent for data collection by the "silent" observers. Consent to the intervention was provided orally and through a letter of support by all local and regional government officials for the respective providers under their rule. Only the research team had access to identifiable data. Data was anonymized when entered into Stata for data analysis and original data forms were stored in a locked container at the University of Goettingen.

3. Results

3.1. Patient health outcomes

Results for patient health outcomes are presented in Table 2. We do not find any statistically significant effect in the unadjusted ITT analysis, except for increased breastfeeding among treatment providers (statistically significant at 10% level, only based on data from CMWs and LHVs). Controlling for additional covariates gives statistically significant positive coefficients (p-value 0.026) for the referral of newborns indicating that SCC providers refer a newborn 2.5 percentage points more often than providers in the control group. Throughout all specifications, all point estimates indicating the occurrence of complications have negative coefficients. However, only the adjusted ITT estimate of birth complications is statistically significant at the 10% level, indicating 1.9 percentage points reduced birth complications among treated providers. The CACE point estimates are slightly larger than the ITT point estimates. The SCC use (CACE) increases referrals of newborns by 3.1 percentage points (p-value 0.038) and patients' initiation of breastfeeding by 8.7 percentage points (p-value 0.058). While all point estimates related to the occurrence of complications are negative, none of them are statistically significant.

Table 3 shows the results for patient health outcomes for the CMW sample. 78.7% of all birth complications were reported by CMWs, while they only make up 21.8% of all deliveries. CMWs in the treatment group have 4.83 percentage points fewer maternal complications (p-value 0.097), 3.18 percentage points fewer newborn complications (p-value 0.006), and an overall reduced rate of birth complications of 6.56 percentage points (p-value 0.052). These results are no longer statistically significant when randomization inference is applied (though p-values are close to the traditionally applied cut-off values) and also not in the CACE regression. For the initiation of breastfeeding ITT is 11.1 percentage points (p-value 0.013) and the CACE is 11.9 percentage points (p-value 0.012).

Table 4
ITT and CACE results – Adherence to essential practices across pause points.

	Unadjusted			Adjusted						mean dep. variable	N
	ITT	SE	p-value	ITT	SE	p-value	CACE	SE	p-value		
Pause Point 1 – On Admission											
Hand Hygiene Index (0–2 – hand washing and gloves worn)[a]	0.1360	[0.148]	(0.358)	0.1350**	[0.067]	(0.044)	0.2817**	[0.139]	(0.044)	1.8193	264
Referral criteria checked (Dummy = 1 if checked)[a]	0.1878	[0.282]	(0.507)	0.4082***	[0.000]	(0.000)	0.8530***	[0.000]	(0.000)	0.6667	264
Partograph started (Dummy = 1 if started)[a]	−0.2557	[0.344]	(0.458)	0.0427	[0.237]	(0.857)	0.0893	[0.495]	(0.857)	0.6707	265
Mother Temperature Index (0–2 – taken and noted)[a]	−0.2991	[0.538]	(0.579)	0.3793	[0.532]	(0.477)	0.7914	[1.110]	(0.477)	1.0122	265
Mother BP Index (0–2 – taken and noted)[a]	−0.3909	[0.393]	(0.320)	0.2164	[0.305]	(0.479)	0.4516	[0.636]	(0.479)	1.7927	265
Birth companion encouraged (Dummy = 1 if encouraged)[a]	−0.1549	[0.331]	(0.641)	0.1346	[0.119]	(0.258)	0.2808	[0.248]	(0.258)	0.6341	266
Danger signs explained to patient (Dummy = 1 if explained)[a]	0.0544	[0.249]	(0.827)	0.2357*	[0.131]	(0.074)	0.4945*	[0.276]	(0.074)	0.6667	269
FHR Index (0–2 – taken and noted)[a]	0.3323	[0.915]	(0.717)	0.6815	[0.684]	(0.320)	1.4218	[1.428]	(0.320)	1.1463	265
Sum of Practices conducted during PP1 (0–8)	−0.0460	[0.718]	(0.949)	1.5236	[0.966]	(0.116)	3.1746	[2.013]	(0.116)	5.2963	261
Pause Point 2 – Just before Delivery											
Hand hygiene Index (0–2 – hand washing and gloves worn)[a]	0.103	[0.167]	(0.539)	0.170	[0.617]	(0.783)	0.1784	[0.646]	(0.783)	1.8421	282
Mother Temperature Index (0–2 – taken and noted)[a]	−0.224	[0.534]	(0.675)	0.618	[5.280]	(0.907)	0.6470	[5.527]	(0.907)	0.7340	278
Mother BP Index (0–2 – taken and noted)[a]	0.0848	[0.775]	(0.913)	0.322	[1.543]	(0.835)	0.3377	[1.618]	(0.835)	1.2105	281
Assistant identified (Dummy = 1 if identified)[a]	−0.00825	[0.111]	(0.941)	0.133	[0.662]	(0.841)	0.1396	[0.694]	(0.841)	0.5745	280
2nd baby checked (Dummy = 1 if checked)[a]	0.380	[0.471]	(0.420)	0.0487	[0.478]	(0.919)	0.0511	[0.502]	(0.919)	0.6596	281
Oxytocin administered 1 min after Birth (Dummy = 1 if administered)[a]	0.0123	[0.0144]	(0.392)	0.00251	[0.0186]	(0.893)	0.0026	[0.020]	(0.893)	0.9894	280
Baby dried immediately (Dummy = 1 if dried)[a]	0.115	[0.149]	(0.442)	0.0468	[0.189]	(0.805)	0.0491	[0.199]	(0.805)	0.8925	279
Newborn Danger Signs checked Index (0–2 – breathing and movement)[a]	0.708	[0.897]	(0.430)	0.238	[2.278]	(0.917)	0.2497	[2.391]	(0.917)	1.3118	276
FHR Index (0–2 – taken and noted)[a]	0.551	[0.985]	(0.577)	0.617	[2.395]	(0.797)	0.6475	[2.512]	(0.797)	0.8211	280
Sum of Practices conducted during PP2 (0–9)	0.926	[2.309]	(0.689)	1.010	[7.639]	(0.895)	1.0604	[8.019]	(0.895)	6.6452	272
Supplies Index (0–9 – available at bedside)	−0.366	[0.227]	(0.108)	0.172	[0.342]	(0.617)	0.1804	[0.360]	(0.617)	7.9667	279
Pause Point 3 – Just after Birth											
Bleeding checked (Dummy = 1 if checked)[a]	0.103	[0.125]	(0.408)	0.0212	[0.151]	(0.889)	0.0224	[0.160]	(0.889)	0.9043	280
Mother Temperature Index (0–2 – taken and noted)[a]	−0.219	[0.693]	(0.753)	−0.433	[4.253]	(0.919)	−0.4587	[4.502]	(0.919)	0.8936	279
Mother BP Index (0–2 – taken and noted)[a]	−0.0234	[0.305]	(0.939)	0.182	[1.101]	(0.869)	0.1930	[1.168]	(0.869)	1.2553	280
Baby Weight Index (0–2 – taken and noted)[a]	−0.133	[0.669]	(0.843)	0.589	[1.646]	(0.721)	0.6206	[1.734]	(0.721)	0.7949	263
Baby Temperature Index (0–2 – taken and noted)[a]	−0.362	[0.735]	(0.623)	0.114	[8.222]	(0.989)	0.1204	[8.717]	(0.989)	0.8065	279
Baby Respiratory Rate Index (0–2 – taken and noted)[a]	0.222	[0.727]	(0.761)	0.421	[1.755]	(0.811)	0.4458	[1.860]	(0.811)	0.7826	278
Skin to skin initiated (Dummy = 1 if initiated)[a]	0.132	[0.303]	(0.663)	0.450	[1.129]	(0.691)	0.4769	[1.197]	(0.691)	0.6129	279
Breastfeeding initiated (Dummy = 1 if initiated)[a]	0.0111	[0.000]	(1.000)	0.159	[0.435]	(0.715)	0.1687	[0.461]	(0.715)	0.5269	278
Danger signs explained (Dummy = 1 if explained)[a]	0.112	[0.431]	(0.795)	0.0896	[0.503]	(0.859)	0.0950	[0.534]	(0.859)	0.6000	281
Sum of Practices conducted during PP3 (0–9)	1.008	[3.196]	(0.753)	0.781	[5.296]	(0.883)	0.8214	[5.570]	(0.883)	4.3462	261
Pause Point 4 – Before Discharge											
Bleeding checked (Dummy = 1 if checked)[a]	−0.0782	[0.0634]	(0.218)	0.0343	[0.0389]	(0.378)	0.0379	[0.043]	(0.378)	0.9579	271
Mother Temperature Index (0–2 – taken and noted)[a]	−0.221	[0.689]	(0.749)	0.616	[3.673]	(0.867)	0.6814	[4.060]	(0.867)	0.9579	271
Mother BP Index (0–2 – taken and noted)[a]	0.0403	[1.886]	(0.983)	0.201	[1.274]	(0.875)	0.2220	[1.408]	(0.875)	1.1158	271
Baby Temperature Index (0–2 – taken and noted)[a]	−0.265	[0.718]	(0.713)	0.0692	[5.011]	(0.989)	0.0765	[5.538]	(0.989)	0.8511	270
Baby Respiratory Rate Index (0–2 – taken and noted)[a]	0.160	[0.779]	(0.837)	−0.156	[4.613]	(0.973)	−0.1729	[5.099]	(0.973)	0.9149	270
Feeding of Baby checked (Dummy = 1 if checked)[a]	0.0988	[0.511]	(0.847)	0.00791	[2.100]	(0.997)	0.0087	[2.321]	(0.997)	0.5000	270
Newborn Danger Signs Index (0–2 – Movement, Cord)[a]	0.129	[0.520]	(0.805)	−0.123	[0.222]	(0.581)	−0.136	[0.246]	(0.581)	1.6035	270
Family planning discussed (Dummy = 1 if discussed)[a]	−0.216	[0.260]	(0.406)	0.150	[0.251]	(0.551)	0.1661	[0.278]	(0.551)	0.7790	271
Follow-up arranged (Dummy = 1 if arranged)[a]	0.197	[0.418]	(0.637)	0.144	[0.807]	(0.859)	0.1587	[0.892]	(0.859)	0.5895	270
Danger signs communicated (Dummy = 1 if communicated)[a]	0.213	[0.444]	(0.633)	0.0728	[0.840]	(0.931)	0.0805	[0.928]	(0.931)	0.5790	271
Danger Information Sheet given to Patient (Dummy = 1 if given)[a]	−0.0895	[0.260]	(0.731)	0.411	[1.594]	(0.797)	0.4536	[1.760]	(0.797)	0.2747	269
Sum of Practices conducted during PP4 (0–11)	0.102	[0.000]	(1.000)	1.023	[13.80]	(0.941)	1.130	[15.23]	(0.941)	6.1463	267
General – Across Pause Points											
Sum of Practices conducted during all PP (0–37)	1.502	[15.11]	(0.921)	4.982	[4.838]	(0.304)	9.187	[8.921]	(0.304)	23.1704	222

***p < 0.01, **p < 0.05, *p < 0.1.

[a] Marks essential practices included in the Sum of Practices indicators.

3.2. Adherence to essential practices

Table 4 displays the results for the essential practices observed following the structure of the SCC. Most of the practices observed do not differ across treatment and control providers. The point estimates for the number of practices adhered to by the providers (summed at each pause point and over the entire delivery process), are statistically insignificant, however, consistently positive. During the first pause point (on admission) we find statistically significant differences in the frequency of hand hygiene and referral criteria checks. This finding is robust to the Benjamini-Hochberg correction for multiple hypothesis testing (refer to Table A5 in the Appendix).

3.3. Self-reported provider behaviour and attitudes

There are no statistically significant results indicating an impact of the SCC on the health providers' empowerment, documentation practices or information exchange (see Table 5). Neither do we find any statistically significant results with regard to knowledge gain. Only the point estimates depicting the effect of the SCC as a reminder and awareness-raising tool show a more robust pattern. Using the SCC (CACE) increases the use of the partograph by 1.13 points (p-value 0.033) and the availability of danger sign sheets by 70.4 percentage points (p-value 0.003). This pattern is also valid in the adjusted ITT regressions. In line with this, treated providers more often hand danger sign information sheets to patients after delivery (p-value 0.07). Further, providers in the treatment group had the perception that they made fewer errors due to excessive workload. These results hold when applying randomization inference. However, only the availability of danger signs and the perceived reduced error rates are statistically significant at the 5% and 10% level, respectively, when correcting for multiple hypothesis testing.

3.4. Attrition

If attrition is correlated with the treatment assignment (so-called differential or non-random attrition) it might bias our estimates and threaten the internal validity (Duflo et al., 2010). In the pooled sample balance on matching variables and important covariates is still given after attrition. Table 6 shows balancing for treatment and control group across important baseline variables (including joint matching variables) with and without the attrited providers. There do not seem to be any systematic differences before and after attrition, which suggests that our study did not suffer from non-random drop-outs.

4. Discussion

While the majority of studies on the SCC were accompanied by a high intensity peer-to-peer coaching (see for example the BetterBirth Trial with peer-to-peer coaching over a 8-month period, starting with visits twice a week (Delaney et al., 2017; Semrau et al., 2017)), we study the potential of the SCC as a low-cost intervention with limited training and monitoring. Hence, we follow a light monitoring approach with monthly visits refraining from a comprehensive and costly coaching approach, which would be difficult for most health systems in resource constrained settings to implement at scale. Kabongo et al. (2017) report that after the coaching ended, a decline in compliance with practices was observed. Also Delaney et al. (2017) mention that without the presence of a coach, the adherence rate dropped by 24% and without an intensive peer-to-peer coaching. Hirschhorn et al. (2015) find very small changes for essential practices conducted. While the presence of coaches might influence the compliance behaviour of the staff (as suggested by e.g., Dharampal et al., 2016; Hales et al., 2008), this does not seem to be the only channel impacting adherence to checklist items. As we are taking compliance rates into account (CACE estimations), it seems that coaching has an additional effect on behavioural change, e.g., transmits additional knowledge, challenges existing views, or supports changes in the organizational structure. Hence, previous studies using a coach-based checklist approach might have partly captured a 'coaching' effect rather than the 'checklist' effect. There is also evidence that a more comprehensive team training or coaching intervention might have played an important role for the results reported by some interventions that studied the Safe Surgical Checklist (Urbach et al., 2014).

Our baseline results indicate that several providers lack basic supplies necessary to fulfil all SCC practices. This information was shared with the responsible local public health offices which may or may not have taken action to improve the provision of supplies after receiving this information. Nevertheless, shortages of supplies have likely contributed to the limited effects of this study. While we do not find any robust effects of a heterogeneous treatment effect by different medication or equipment availability, this study might be underpowered to detect such effects. The endline assessment also revealed that a substantial number of providers did not have access to basic medication and equipment required to effectively use the SCC. 11.92% of the providers had none of the required medication available and more than 40% had less than half of essential medicine in stock at their facilities. None of the CMWs or HFs had all medication available (including Magnesium Sulphate, skin disinfectant, IV fluids, antibiotics for mother and newborn, Oxytocin, Vitamin K for the newborn). Providers were better endowed with basic equipment. Less than 20% of the providers had fewer than half of the necessary equipment available, and 20% of the HFs had access to all basic equipment (including suction machine, mucus extractor, oxygen cylinder, needle or syringe, baby ambubag, fetoscope, BP apparatus, thermometer, stethoscope, baby scale, partograph sheets, scissors, cord clamp, clean pads for mother, baby towels).

Several essential practices are not regularly performed by the majority of health staff, such as the provision of essential medication to the newborn (Vitamin K), the use of important delivery tools (partograph), or the monitoring of basic maternal and neonatal health indicators (temperature, blood pressure). However, comparing this to the baseline situation in other checklist studies in lower middle-income countries, delivery standards among our study's health providers seem substantially higher across several practices (Semrau et al., 2017). In particular, hygiene practices – directly related to infections, the leading newborn complication – are regularly performed among our study population (in about 90% of deliveries) as compared to the Indian health staff (in less than 1% of deliveries). Health indicators, like maternal temperature, are taken in about 50% of the deliveries in Pakistan versus in less than 1% of deliveries in India. While partograph use is as low as 40% in our setting even after the checklist intervention, the partograph was only started in 1% of the deliveries observed in India.

Overall, we find no effect of the SCC on patient health outcomes, particularly complication rates. This is in line with results reported by Semrau et al. (2017), who do not find improvements with respect to maternal and perinatal mortality and maternal morbidity. There is some evidence that CMWs that received the SCC treatment experience fewer maternal, newborn, and general birth complications. There are a number of possible explanations for these heterogeneous effects. CMWs are often the first entry point to the formal health system, and hence might have a larger leverage to prevent complications. Higher-level facilities experience a case-mix that is already characterized by complicated deliveries or patients with referred complications. Hence, their ability to prevent complications might be limited as compared to lower-level health care providers. A similar argument derives from the standardized process of deliveries underlying the SCC structure and idea: It is reasonable to suggest that lower-level providers more often conduct deliveries that follow normal procedures as compared to more complex and challenging deliveries at higher-level facilities. Checklists might be more useful and applicable to the former situation (Hales & Pronovost, 2006; Myers, 2016). This might also be a reason for the success of surgical checklists, as surgeries are mostly planned (90%) and follow a more standardized procedure (Helmiö, 2015). Frequent staff turnover

Table 5
ITT and CACE results – Impact channels of SCC.

	Unadjusted				Adjusted								
	ITT	SE	p-value	RI p-value	ITT	SE	p-value	RI p-value	CACE	SE	p-value	mean dep. variable	N
Reminder and Awareness Raiser													
How long does women stay at provider after delivery? (min)	13.0951	[16.223]	(0.421)	(0.398)	9.615	[15.720]	(0.542)	(0.538)	26.280	[38.360]	(0.493)	115.7759	120
How long does provider stay with woman after delivery? (min)	2.0990	[12.494]	(0.867)	(0.886)	2.119	[12.660]	(0.868)	(0.885)	9.833	[28.910]	(0.734)	86.7105	80
Provider uses partograph for all deliveries? (1–4 – "Never" – "Always ")	0.4780**	[0.218]	(0.030)	(0.032)	0.396*	[0.214]	(0.066)	(0.066)	1.130**	[0.528]	(0.033)	2.6129	128
Does provider encourage birth companion? (Dummy = 1 if yes)	0.0762	[0.077]	(0.323)	(0.289)	0.087	[0.074]	(0.246)	(0.195)	0.215	[0.191]	(0.260)	0.7419	128
Danger information sheets available? (Dummy = 1 if yes)	0.2590***	[0.088]	(0.004)	(0.001)	0.262***	[0.085]	(0.003)	(0.003)	0.704***	[0.240]	(0.003)	0.2258	128
Provides patient with danger information sheets? (Dummy = 1 if yes)	0.1409	[0.087]	(0.108)	(0.112)	0.154*	[0.089]	(0.089)	(0.089)	0.423*	[0.233]	(0.070)	0.2833	126
Generally gives Vitamin K to newborn? (Dummy = 1 if yes)	0.0254	[0.055]	(0.643)	(0.643)	0.00029	[0.048]	(0.995)	(0.991)	0.00476	[0.122]	(0.969)	0.0807	128
Generally conducts clean cord care? (Dummy = 1 if yes)	0.0571	[0.139]	(0.685)	(0.911)	0.029	[0.153]	(0.851)	(0.901)	0.0618	[0.324]	(0.851)	0.8000	24
Forgets to note down information (1–6 - "Never" - "Very often")	−0.2480	[0.189]	(0.192)	(0.203)	−0.238	[0.196]	(0.227)	(0.234)	−0.549	[0.475]	(0.247)	1.6230	125
Reported information s/he was unsure of (1–6 - "Never" -"Very often")	−0.0154	[0.713]	(0.983)	(1.000)	0.201	[0.160]	(0.208)	(0.463)	0.384	[0.305]	(0.208)	1.4000	23
Made errors due to being fatigue (1–6 - "Never" - "Very often")	0.1135	[0.134]	(0.399)	(0.432)	0.164	[0.146]	(0.262)	(0.248)	0.340	[0.335]	(0.311)	1.2459	125
Made errors due to excessive work load (1–6 - "Never" - "Very often")	−0.2646***	[0.099]	(0.009)	(0.008)	−0.247**	[0.102]	(0.017)	(0.011)	−0.612**	[0.267]	(0.022)	1.3115	125
Made errors due to distractions (1–6 - "Never" - "Very often")	−0.1186	[0.098]	(0.229)	(0.289)	−0.087	[0.114]	(0.446)	(0.398)	−0.211	[0.281]	(0.453)	1.1967	125
Satisfaction with quality of delivery services (1–5"Not at all satisfied"- "Very satisfied")	0.0288	[0.166]	(0.863)	(0.879)	0.084	[0.162]	(0.604)	(0.618)	0.200	[0.410]	(0.627)	4.2742	128
Self-assessment safe delivery if all supplies available (1–6 – "Not secure at all" - "Very secure")	−0.0020	[0.137]	(0.989)	(1.000)	−0.0381	[0.139]	(0.784)	(0.798)	−0.104	[0.350]	(0.766)	5.5323	128
Would feel safe being treated here"(1–6 – "Disagree strongly" - "Agree strongly")	−0.1903	[0.172]	(0.271)	(0.328)	−0.0745	[0.157]	(0.637)	(0.625)	−0.173	[0.389]	(0.656)	5.3934	125
Empowerment													
I have influence in delivery section (1–6 - "Disagree strongly" - "Agree strongly")	0.329	[0.577]	(0.575)	(0.604)	0.676	[0.652]	(0.322)	(0.371)	1.429	[1.406]	(0.322)	4.6000	24
It is easy to speak up in case of problems with care (1–6 - "Disagree Strongly" - "Agree strongly")	−0.143	[0.455]	(0.757)	(0.726)	−0.0503	[0.300]	(0.947)	(0.940)	−0.106	[1.575]	(0.947)	5.5000	24
Delivery staff members work together as a well-coordinated team (1–6 - "Disagree strongly" - "Agree strongly")	−0.143	[0.571]	(0.805)	(0.753)	0.361	[0.306]	(0.264)	(0.309)	0.763	[0.663]	(0.264)	5.5000	24

***p < 0.01, **p < 0.05, *p < 0.1.

Data is from provider survey. Robust standard errors (SE) are clustered at the provider level. In case of small number of clusters standard errors are wild bootstrapped. Adjusted regressions include strata dummies, a rural/urban dummy and a dummy indicating if a facility is open 24/7. All regressions are based upon the analytical sample. ITT gives the Intent-to-Treat effects and CACE the Complier-Average-Causal effects, where 'correctly answered SCC item' is used as complier indication (IV Regression). RI p-value is calculated using randomization inference. Mean dependent variable is based upon the values of the control group in the endline assessment. N refers to the number of interviewees.

Table 6
Balancing table of non-attrited vs. attrited pooled sample.

Variables	Non-Attrited Baseline Sample								Attrited Baseline Sample							
	Pooled Sample			Control		Treatment		p-value of difference in means test	Pooled Sample			Control		Treatment		p-value of difference in means test
	N	Mean	SD	Mean	SD	Mean	SD		N	Mean	SD	Mean	SD	Mean	SD	
# of doctors	146	0.041	0.231	0.014	0.118	0.068	0.302	0.162	117	0.051	0.258	0.018	0.134	0.082	0.331	0.180
# of delivery staff	17	4.529	2.672	4.250	3.370	4.778	2.048	0.698	13	5.000	2.828	5.500	4.509	4.778	2.048	0.690
Equipment Index	145	8.959	1.654	8.592	1.841	9.311	1.374	0.008***	116	9.155	1.454	8.873	1.454	9.410	1.419	0.047**
Medication Index	143	3.888	1.290	3.915	1.273	3.861	1.314	0.802	115	3.809	1.228	3.818	1.156	3.800	1.299	0.937
Disinfection Index	146	2.637	0.742	2.556	0.837	2.716	0.631	0.192	117	2.607	0.742	2.536	0.808	2.672	0.676	0.323
Amenities Index	146	2.637	1.275	2.653	1.313	2.622	1.246	0.883	117	2.726	1.229	2.911	1.180	2.557	1.259	0.121
% Newborn Death	142	0.009	0.037	0.006	0.020	0.011	0.049	0.412	114	0.010	0.041	0.007	0.021	0.013	0.053	0.446
% Stillbirth	143	0.013	0.053	0.011	0.025	0.015	0.071	0.727	115	0.014	0.059	0.012	0.025	0.016	0.078	0.695
% Maternal Death	142	0.005	0.042	0.002	0.006	0.008	0.059	0.383	114	0.006	0.047	0.002	0.007	0.009	0.064	0.395
% Newborn Complications	144	0.162	0.370	0.154	0.423	0.170	0.312	0.798	116	0.154	0.327	0.117	0.320	0.187	0.332	0.252
% Maternal Complications	143	0.444	1.404	0.354	0.746	0.530	1.828	0.455	115	0.367	0.713	0.360	0.762	0.374	0.674	0.921
% Birth Complications	143	0.607	1.535	0.510	1.022	0.700	1.905	0.462	115	0.523	0.948	0.480	0.978	0.561	0.926	0.650
% Incoming Referral	144	0.256	0.527	0.270	0.561	0.243	0.496	0.754	116	0.244	0.451	0.196	0.339	0.286	0.532	0.287
Total deliveries	145	176.766	610.689	143.319	455.449	209.753	734.24	0.514	117	179.402	669.594	139.679	500.660	215.869	796.383	0.541
Educational level	129	4.116	0.915	4.219	1.061	4.015	0.739	0.208	104	4.087	0.936	4.192	1.085	3.981	0.754	0.251
Rural Dummy	146	1.863		1.917		1.811		0.064	117	1.872		1.893		1.852		0.518
Open 24/7	17	0.471		0.375		0.556		0.488	13	0.538		0.500		0.556		0.867
International Funding of HF	17	0.824		0.750		0.889		0.485	13	0.846		0.750		0.889		0.561
Access emergency transport	146	0.363		0.347		0.378		0.698	117	0.410		0.411		0.410		0.992
BEmONC	146	0.034		0.014		0.054		0.185	117	0.043		0.018		0.066		0.206
CEmONC	145	0.062		0.056		0.068		0.781	117	0.060		0.054		0.066		0.787
Partograph use	144	0.139		0.125		0.153		0.633	115	0.122		0.125		0.119		0.918
Vitamin K provision	146	0.240		0.292		0.189		0.149	117	0.214		0.268		0.164		0.174
Birth companion encouraged	146	0.774		0.833		0.716		0.092*	117	0.752		0.804		0.705		0.220
Deliveries at patient home	129	0.775		0.766		0.785		0.798	104	0.740		0.712		0.769		0.507
Delivery room	117	0.906		0.914		0.898		0.776	94	0.926		0.936		0.915		0.698
Provider was born in area	129	0.612		0.531		0.692		0.061*	104	0.635		0.558		0.712		0.105
Access to HF drugstore	44	0.432		0.524		0.348		0.249	30	0.433		0.625		0.214		0.023**
At patient home: running water	126	0.944		0.951		0.938		0.764	101	0.960		0.980		0.942		0.342
At patient home: Electricity	125	0.744		0.738		0.750		0.876	100	0.710		0.714		0.706		0.927
At patient home: Improved toilet	125	0.968		0.951		0.984		0.290	100	0.980		0.980		0.980		0.977

Asterisks indicate p-values according to: ***p < 0.01, **p < 0.05, *p < 0.1.
Data is from provider survey. Sample is based upon all providers included in the baseline sample and variables are from the baseline provider survey. Attrition refers to drop-outs throughout the study period leading to the final endline sample. SD gives the standard deviation, N the number of observations, i.e. providers.

seen throughout the study period at most of the HFs, might have further limited the possible effect of the SCC as it reduced the general SCC experience and know-how of each involved staff member.

5. Conclusion

Our study aimed at generating rigorous evidence on the effectiveness of the WHO Safe Childbirth Checklist (SCC). This is the first study to include non-facility based health providers in addition to primary and secondary health facilities. Overall, our study finds only very limited evidence for positive effects of the SCC on a range of provider- and patient-level outcomes. There is some indication that the SCC serves as a mnemonic device. Suggested SCC mechanisms, including improved knowledge and documentation practices, are not statistically significantly related to the SCC use. Findings from the provider records show increased newborn referral rates to higher-level facilities, as well as suggestive evidence of decreased complication rates among community midwives, a non-facility based group of health workers. However, overall we find no statistically significant effect on complication rates, which is in line with other rigorous evaluations of the SCC (Kaplan et al., 2021; Semrau et al., 2017). In comparison to other – mostly pre-post-studies, our results are less positive with regard to adherence to essential practices by health staff. Our findings do not support an overall increase in practices followed by the providers. To test the SCC in its effectiveness as a low-cost and easy-to-implement tool, we followed a light monitoring approach without additional coaching. Our findings are similar when compliance rates of the SCC use are taken into account (CACE estimations). Hence, coaching seems to affect outcomes through additional channels. In order to better understand the interdependency of these aspects, it would be necessary to study interventions with cross-cutting designs by assigning different treatment groups. While general delivery standards were above those of other studies (possibly already limiting our potential to generate significant improvements), we were still confronted with a lack of basic delivery knowledge among study participants essential for a successful adoption and use of the SCC. In other studies, with equal or even lower delivery standards, this necessary know-how was transmitted through the coaching component in treatment facilities. In our study, a delivery training of all study participants was conducted prior to the intervention. This ensured an equal knowledge base, enabling us to relate differences in outcomes between treatment and control group to the SCC use.

Ethical statement

Ethical clearance was provided by the National Bioethics Committee (NBC) Pakistan in October 2015 after a detailed study protocol was reviewed and approved by them.

Financial disclosure statement

This research was supported by the Reproductive, Maternal & Newborn Health Project (RMNHP) of Deutsche Gesellschaft für INTERNATIONALE ZUSAMMENARBEIT (GIZ) GmbH and funded by the German Federal Ministry for Economic Cooperation and Development.

Ethical statement

Ethical clearance was provided by the National Bioethics Committee (NBC) Pakistan in October 2015 after a detailed study protocol was reviewed and approved by them.

Declarations of competing interest

None.

Data availability

Data will be made available on request.

Acknowledgements:

We are grateful to the entire RMNHP team of GIZ Islamabad and all project partners from the public health system in Khyber Pakhtunkhwa, particularly Haripur and Nowshera districts, for their support and collaboration. We thank Marina Kuch, Peter Brückmann, Katharina Kreutz, Aristide Djimgou, and Memona Mateen for their excellent work as research assistants and support in data collection as well as the numerous local research assistants without whom this study would not have been possible. We are grateful to Lisa Bogler for her assistance with improving the final manuscript and proofreading. Finally, we would like to thank the anonymous referees for their thoughtful and valuable comments. We acknowledge support by the Open Access Publication Funds of the University of Goettingen.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssmph.2023.101495.

References

Borchard, A., Schwappach, D. L. B., Barbir, A., & Bezzola, P. (2012). A systematic review of the effectiveness, compliance, and critical factors for implementation of safety checklists in surgery. *Annals of Surgery, 256*(6), 925. https://doi.org/10.1097/SLA.0b013e3182682f27

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics, 1*(4), 200–232. https://doi.org/10.1257/app.1.4.200

Burgess, D. J. (2010). Are providers more likely to contribute to healthcare disparities under high levels of cognitive load? How features of the healthcare setting may lead to biases in medical decision making. *Medical Decision Making, 30*(2), 246–257. https://doi.org/10.1177/0272989X09341751

Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P.-A. C., & Rubin, H. R. (1999). Why don't physicians follow clinical practice guidelines?A framework for improvement. *JAMA, 282*(15), 1458–1465. https://doi.org/10.1001/jama.282.15.1458

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics, 90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Cochrane, L. J., Olson, C. A., Murray, S., Dupuis, M., Tooman, T., & Hayes, S. (2007). Gaps between knowing and doing: Understanding and assessing the barriers to optimal health care. *Journal of Continuing Education in the Health Professions, 27*(2), 94–102. https://doi.org/10.1002/chp.106

Courtenay-Hall, P., & Rogers, L. (2002). Gaps in Mind: Problems in environmental knowledge-behaviour modelling research. *Environmental Education Research, 8*(3), 283–297. https://doi.org/10.1080/13504620220145438

Croskerry, P. (2002). Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Academic Emergency Medicine, 9*(11), 1184–1204. https://doi.org/10.1197/aemj.9.11.1184

Deck, C., & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review, 78*, 97–119. https://doi.org/10.1016/j.euroecorev.2015.05.004

Delaney, M. M., Maji, P., Kalita, T., Kara, N., Rana, D., Kumar, K., Masoinneuve, J., Cousens, S., Gawande, A. A., Kumar, V., Kodkany, B., Sharma, N., Saurastri, R., Singh, V. P., Hirschhorn, L. R., Semrau, K. E., & Firestone, R. (2017). 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. *Global Health Science and Practice, 5*(2), 217–231. https://doi.org/10.9745/GHSP-D-16-00410

Delaney, M. M., Miller, K. A., Bobanski, L., Singh, S., Kumar, V., Karlage, A., Tuller, D. E., Gawande, A. A., & Semrau, K. E. A. (2019). Unpacking the null: A post-hoc analysis of a cluster-randomised controlled trial of the WHO safe childbirth checklist in Uttar Pradesh, India (BetterBirth). *Lancet Global Health, 7*(8), e1088–e1096. https://doi.org/10.1016/S2214-109X(19)30261-X

Dharampal, N., Cameron, C., Dixon, E., Ghali, W., & Quan, M. L. (2016). Attitudes and beliefs about the surgical safety checklist: Just another tick box? *Canadian Journal of Surgery, 59*(4), 268–275. https://doi.org/10.1503/cjs.002016

Duflo, E., Glennerster, R., & Kremer, M. (2010). Using randomization in development economics research: A toolkit. In D. Rodrik, & M. R. Rosenzweig (Eds.), *Handbook of development economics* (pp. 3895–3962). Elsevier B.V. https://doi.org/10.1016/S1573-4471(07)04061-2.

Frederiks, E. R., Stenner, K., & Hobman, E. V. (2015). Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour.

Renewable and Sustainable Energy Reviews, 41, 1385–1394. https://doi.org/10.1016/j.rser.2014.09.026

Greevy, R., Lu, B., Silber, J. H., & Rosenbaum, P. (2004). Optimal multivariate matching before randomization. *Biostatistics, 5*(2), 263–275. https://doi.org/10.1093/biostatistics/5.2.263

Hales, B. M., & Pronovost, P. J. (2006). The checklist—a tool for error management and performance improvement. *Journal of Critical Care, 21*(3), 231–235. https://doi.org/10.1016/j.jcrc.2006.06.002

Hales, B., Terblanche, M., Fowler, R., & Sibbald, W. (2008). Development of medical checklists for improved quality of patient care. *International Journal for Quality in Health Care, 20*(1), 22–30. https://doi.org/10.1093/intqhc/mzm062

Haugen, A. S., Søfteland, E., Almeland, S. K., Sevdalis, N., Vonen, B., Eide, G. E., Nortvedt, M. W., & Harthug, S. (2015). Effect of the world health organization checklist on patient outcomes: A stepped wedge cluster randomized controlled trial. *Annals of Surgery, 261*(5), 821. https://doi.org/10.1097/SLA.0000000000000716

Helmiö, P. (2015). Towards better patient safety: The WHO surgical checklist in otorhinolaryngology [University of Helsinki] http://hdl.handle.net/10138/154152.

Heß, S. (2017). Randomization inference with Stata: A guide and software. *STATA Journal: Promoting Communications on Statistics and Stata, 17*(3), 630–651. https://doi.org/10.1177/1536867X1701700306

Hirschhorn, L. R., Semrau, K., Kodkany, B., Churchill, R., Kapoor, A., Spector, J., Ringer, S., Firestone, R., Kumar, V., & Gawande, A. (2015). 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. *Implementation Science, 10*(1), 117. https://doi.org/10.1186/s13012-015-0309-y

Hoffman, M., Mukete, V., Szmanda, J., & Witt, C. (2011). *The effect of increased stress levels on memory test performance*.

Imai, K., King, G., & Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Statistical Science, 24*(1), 29–53. https://doi.org/10.1214/08-STS274

Kabongo, L., Gass, J., Kivondo, B., Kara, N., Semrau, K., & Hirschhorn, L. R. (2017). Implementing the WHO Safe Childbirth Checklist: Lessons learnt on a quality improvement initiative to improve mother and newborn care at Gobabis District Hospital, Namibia. *BMJ Open Quality, 6*(2), Article e000145. https://doi.org/10.1136/bmjoq-2017-000145

Kaplan, L. C., Ichsan, I., Diba, F., Marthoenis, M., Muhsin, M., Samadi, S., Richert, K., Susanti, S. S., Sofyan, H., & Vollmer, S. (2021). Effects of the world health organization safe childbirth checklist on quality of care and birth outcomes in Aceh, Indonesia: A cluster-randomized clinical trial. *JAMA Network Open, 4*(12), Article e2137168. https://doi.org/10.1001/jamanetworkopen.2021.37168

Kara, N., Firestone, R., Kalita, T., Gawande, A. A., Kumar, V., Kodkany, B., Saurastri, R., Singh, V. P., Maji, P., Karlage, A., Hirschhorn, L. R., & Semrau, K. E. (2017). 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, 5*(2), 232–243. https://doi.org/10.9745/GHSP-D-16-00411

Kramer, H. S., & Drews, F. A. (2017). Checking the lists: A systematic review of electronic checklist use in health care. *Journal of Biomedical Informatics, 71*, S6–S12. https://doi.org/10.1016/j.jbi.2016.09.006

Kumari, S., Panicker, R., Jayaram, A., Dumka, N., Sharma, S., Singhal, V., Chatterjee, T., & Varghese, B. (2016). Evaluation of the safe childbirth checklist program in Rajasthan, India: The how and what of the evaluation efforts. *Journal of Public Health in Developing Countries, 2*(2), 212–222.

Kumar, S., Yadav, V., Balasubramaniam, S., Jain, Y., Joshi, C. S., Saran, K., & Sood, B. (2016). Effectiveness of the WHO SCC on improving adherence to essential practices during childbirth, in resource constrained settings. *BMC Pregnancy and Childbirth, 16*(1), 345. https://doi.org/10.1186/s12884-016-1139-x

Lichand, G., & Mani, A. (2020). *Cognitive droughts (SSRN scholarly paper No. 341)*. SSRN. https://doi.org/10.2139/ssrn.3540149

Lu, B., Greevy, R., Xu, X., & Beck, C. (2011). Optimal nonbipartite matching and its statistical applications. *The American Statistician, 65*(1), 21–30. https://doi.org/10.1198/tast.2011.08294

Michie, S., Johnston, M., Abraham, C., Lawton, R., Parker, D., & Walker, A. (2005). Making psychological theory useful for implementing evidence based practice: A consensus approach. *BMJ Quality and Safety, 14*(1), 26–33. https://doi.org/10.1136/qshc.2004.011155

Myers, P. L. (2016). Commercial aircraft electronic checklists: Benefits and challenges (literature review). *International Journal of Aviation, Aeronautics, and Aerospace, 3*(1), 1.

Patabendige, M., & Senanayake, H. (2015). Implementation of the WHO safe childbirth checklist program at a tertiary care setting in Sri Lanka: A developing country experience. *BMC Pregnancy and Childbirth, 15*(1), 12. https://doi.org/10.1186/s12884-015-0436-0

Semrau, K. E. A., Hirschhorn, L. R., Kodkany, B., Spector, J. M., Tuller, D. E., King, G., Lipsitz, S., Sharma, N., Singh, V. P., Kumar, B., Dhingra-Kumar, N., Firestone, R., Kumar, V., & Gawande, A. A. (2016). 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, 17*(1), 576. https://doi.org/10.1186/s13063-016-1673-x

Semrau, K. E. A., Hirschhorn, L. R., Marx Delaney, M., Singh, V. P., Saurastri, R., Sharma, N., Tuller, D. E., Firestone, R., Lipsitz, S., Dhingra-Kumar, N., Kodkany, B. S., Kumar, V., & Gawande, A. A. (2017). Outcomes of a coaching-based WHO safe childbirth checklist program in India. *New England Journal of Medicine, 377*(24), 2313–2324. https://doi.org/10.1056/NEJMoa1701075

Sligo, F. X., & Jameson, A. M. (2000). The knowledge—behavior gap in use of health information. *Journal of the American Society for Information Science, 51*(9), 858–869. https://doi.org/10.1002/(SICI)1097-4571(2000)51:9<858::AID-ASI80>3.0.CO;2-Q

Spector, J. M., Lashoher, A., Agrawal, P., Lemer, C., Dziekan, G., Bahl, R., Mathai, M., Merialdi, M., Berry, W., & Gawande, A. A. (2013). Designing the WHO Safe Childbirth Checklist program to improve quality of care at childbirth. *International Journal of Gynecology & Obstetrics, 122*(2), 164–168. https://doi.org/10.1016/j.ijgo.2013.03.022

Stock, J., & Watson, M. (2006). *Introduction to econometrics* (2nd ed.). Pearson.

UNICEF. (2023). *Neonatal mortality*. UNICEF DATA. https://data.unicef.org/topic/child-survival/neonatal-mortality/.

UNICEF, World Health Organization, World Bank Group, & United Nations. (2020). *A Neglected Tragedy: The global burden of stillbirths*. United Nations Children's Fund. https://data.unicef.org/resources/a-neglected-tragedy-stillbirth-estimates-report/.

Urbach, D. R., Govindarajan, A., Saskin, R., Wilton, A. S., & Baxter, N. N. (2014). Introduction of surgical safety checklists in ontario, Canada. *New England Journal of Medicine, 370*(11), 1029–1038. https://doi.org/10.1056/NEJMsa1308261

Workman, M., Lesser, M. F., & Kim, J. (2007). An exploratory study of cognitive load in diagnosing patient conditions. *International Journal for Quality in Health Care, 19*(3), 127–133. https://doi.org/10.1093/intqhc/mzm007

World Health Organization & UNICEF. (2010). *Countdown to 2015 decade report (2000-2010): Taking stock of maternal, newborn, and child survival*. World Health Organization and UNICEF. https://www.who.int/publications/i/item/9789241599573.

World Health Organization. (2022). *Newborn mortality*. https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-mortality-report-2021.

World Health Organization. (2005). *The World Health Report 2005. Make every mother and child count*. World Health Organization. https://www.who.int/publications/i/item/9241562900.

World Health Organization. (2008). *Safe surgery save lives*. World Health Organization. http://apps.who.int/iris/bitstream/10665/70080/1/WHO_IER_PSP_2008.07_eng.pdf.

World Health Organization. (2023). *Maternal mortality*. https://www.who.int/news-room/fact-sheets/detail/maternal-mortality.

Zafar, Z. (2017). *Development, implementation and validation of an integrated WHO Safe childbirth checklist/maternal and newborn health training package*. Jhpiego.