Registered report Stage 1: Proposal
Healing by hair: local services provision and mental health first response

Björn Nilsson,
Université Paris-Saclay

Clémence Pougué Biyong,
Université Paris 1 Panthéon-Sorbonne

Date of latest draft: 9/06/2023

Abstract

Despite large economic effects, care for mental illnesses is severely underfunded in the developing world, with few signs of increasing commitment towards the creation of mental health infrastructures. This paper describes the evaluation of an innovative program training hair-dressers in Abidjan, Côte d'Ivoire to act as first responders to customers manifesting mental health issues. The program, which aims to improve customers' well-being and produce positive externalities in the labor market and in female decision-making, has the potential to act as a first line of defense against mental illness in low-income settings. 74 hair-dressers underwent a three-day training event in February, 2023, run by medical professionals. The study, set up as a randomized controlled trial, will assess the impact of the program on the quality of hair-dresser - customer interactions, customer mental health and well-being, and secondary outcomes related to the labor market and female decision-making.

Keywords: mental health, well-being, Sub-Saharan Africa, female empowerment.

JEL codes: I12,I15,J16.

Timeline: The data collection is expected to finish in February, 2023. Data analysis and write-up of a final article will be finished by June, 2024.

Pre-registration status: The trial is pre-registered at the AEA RCT registry under the reference AEARCTR-0009989, and can be accessed at https://www.socialscienceregistry.org/trials/9989

1 Introduction

Globally, the burden of mental health is increasing and most likely understated. If migraine and epilepsy are included as mental illnesses, the disability-adjusted life years lost to mental illness surpass those lost to cancer (Vigo, Thornicroft, and Atun 2016). Developing countries are not excluded from this epidemic. For Africa, the WHO's *Global burden of disease* report estimated some 17.9 million years lost to disability due to mental health problems—a figure that comes close to the 18.5 million years lost to disability due to infectious and parasitic diseases (WHO 2016). Yet, in 2014, 46% of African countries did not have a standalone mental health policy framework (Sankoh, Sevalie, and Weston 2018), and in 2005 some 40% of African governments did not allocate any resources at all to mental health (Skeen et al. 2010). In the donor community, mental health is also by all means underfunded. Gilbert et al. (2015) show that between 2007 and 2013, less than one percent of development assistance for health targeted mental health.

Mental illness has economic consequences (Ridley et al. 2020). Back of the envelope calculations from Richard Layard suggest that mental illness is responsible for a 7.5% output gap in Britain (Layard 2013), and that curing depression and anxiety illnesses alone would produce a 4% GDP increase (Layard 2017). His calculations also suggest that at given cost and success rates of therapy and medication respectively, funding treatment would essentially pay for itself. Excellent returns to therapy have also been found in lower income settings: Patel et al. (2017) finds that adding lay-worker therapy to usual care improves depression outcomes in a group of depressed Indians in Goa, adding that at average cost the program pays for itself in as short a time interval as a couple of months. A recent meta-analysis of the economic impacts of mental health interventions in low and middle-income countries reports results from 39 pharmaceutical and psychosocial interventions over the period 1990 to 2018 (Lund et al. 2020). Their overall findings suggest that half of the reported estimates are positive and statistically significant, and that effects are strongest for education related outcomes, and positively correlated with country income levels. The reported studies however most often rely on health professionals delivering care in clinical settings. Given the near total lack of funding of mental and the donor community, combined with the absence of infrastructure, such interventions are unlikely to scale up any time soon. Searching for alternative solutions may thus constitute a better option in many poor and fragile contexts. The intervention described in this

paper has the prominent feature of training local service providers to provide first response to mental distress during interactions arising in already existing settings, limiting infrastructure costs and enabling quick and cheap scaling up. Furthermore, most of the clinical trials surveyed focus on curing the already sick and report no evidence on the impact of therapy on the risk of becoming ill among healthy subjects. Finally, most papers cited in Lund et al. (2020) focus on employment and educational decisions. While these are indeed very relevant outcomes related to mental health, they are not the only dimensions in which positive mental health shocks may produce externalities, in particular when interventions target women. Externalities related to female decision-making and gender norms may also arise from improved mental health. Such impacts have generally not been the focus of the literature. A notable exception here is the cluster randomized study in Pakistan (Maselko et al. 2020; Baranov et al. 2020) on mothers with perinatal depression. The intervention provided cognitive behavioral therapy by community primary health workers, and the treatment group showed a 33 percentage point stronger reduction in depression prevalence after one year than the control group. Furthermore, seven years after the intervention, treated women showed higher control over household spending than women in the control group, with increased money and time investment toward children as a result.

In the following study, we plan to evaluate the Heal-by-hair program, deployed in Côte d'Ivoire and with planned deployments in Togo and Cameroon. It focuses on enabling hair-dressers to become active listeners and "first responders" in mental health, through a three-day-training program led by medical professionals. The training content is based on mhGAP, the Mental Health Gap Action Programme of the WHO, a guide to act against mental, neurological and substance disorders intended for practitioners in non-specialized health settings. The three-day training given through the program aims to enable hair-dressers to observe and recognize the first manifestations of mental health problems in their customers, and if needed, be able to orient them towards mental health professionals. The rationale of the project lies in the fact that hair-dressers spend a lot of time with their customers, who often share stories from their personal lives. A non-representative survey run by the Bluemind foundation (Bluemind Foundation 2021) on 714 women in six African countries showed that 64% of interviewed women at least *sometimes* confide in their hair-dresser, and 86% believe that other women do so. In our baseline survey on 786 customers of hair dressers in three municipalities of Abidjan, seven out of ten women declare at least sometimes confide in their hair dressers. The duration and repetition

of interactions, added to the neutral status of the hair-dresser, could in combination with training constitute a powerful first line of defense against depression and anxiety. Using again data from our baseline survey, we find that 74% of customers visit their hair dresser more than once a month.

Previous studies in development economics have mainly focused on the links between poverty and mental disorders (Lund et al. 2020; Chemin, De Laat, and Haushofer 2013; Ridley et al. 2020; Frasquilho et al. 2015; Kuhn, Lalive, and Zweimüller 2009; Haushofer, Mudida, and Shapiro 2020), with the exceptions of Baranov et al. 2020; Barker et al. 2022. Our intervention differs from the two above interventions in several aspects. Baranov et al. 2020's intervention in Pakistan was one of the largest psychotherapy interventions run in developing countries, directly targeting 903 women affected by prenatal depression disorders. Contrary to the Pakistan intervention, where peer recruitment was carried out for the purpose of the intervention, our intervention innovates by relying on a social structure that is the trustful and pre-existing relationship between hair-dressers and their customers (Cowen, Gesten, Boike, et al. 1979; Cowen, Gesten, Davidson, et al. 1981; Wilson et al. 2008), and the cultural routine in West and Central African to cultivate a relational capital between hairdressers and clients (Essah 2008; Xiao et al. 2020). This intervention aims at improving women's access to mental health care by finding a scalable non-medical alternative to identify and diagnose people at risk as well as by offering active listening by hairdressers to better manage depression and anxiety symptoms.

Our study also builds on other strands of the literature. It contributes to a growing literature on women's empowerment. Hairdressers gain new socio-emotional and cognitive skills as well as receive mental health training and support that can be useful in managing their own anxiety and depressive disorders. In the framework of the behavioral economics theory of scarcity (Mullainathan and Shafir 2013; Schilbach, Schofield, and Mullainathan 2016), this intervention would allow them to increase their mental resources and widen the bandwidth to multi-task and improve their productivity.

Finally, our study provides information on the mental health of women in West and Central Africa in a new setting. First, contrary to Baranov et al. (2020) and Barker et al. (2022), our intervention takes place in an urban setting—three municipalities of Abidjan, the largest city in Côte d'Ivoire. Second, our population of interest is all

women, and we do not focus on cases related to pre- or post-natal depression, like in Baranov et al. 2020; Bindt et al. 2012; Wemakor and Mensah 2016. Third, by focusing on a general population including healthy individuals our results will add to Barker et al. 2022 who study the impact of a group-based cognitive behavioral therapy (CBT) in a sample of 7,227 from the general population of poor households in rural Ghana. They show that any CBT intervention has an effect regardless of baseline mental distress. Fourth, we show results in both the short and medium term, unlike Baranov et al. 2020; Haushofer, Mudida, and Shapiro 2020; Bhat et al. 2022 who bring long-term results or Barker et al. 2022 who present only short-term results.

If the intervention reduces the long-run prevalence of mental illness, it is likely to be cost-effective. It is difficult to suggest numbers *a priori*, since similar research in this context is nonexistent. We do however know that mental illness correlates with unemployment, absenteeism and productivity (negatively). At baseline and with conservative assumptions—namely that each hair-dresser prevents depression in one woman at all time, and depression is associated with a 50% drop in income, and that the average woman in Abidjan earns the 2021 GDP per capita for Côte d'Ivoire—the income loss avoided per hair-dresser is 6447$ over a five-year period, a sum outweighing training costs by a factor of 10. In the follow-up, the hair dressers' relatively low take-up (see Section 3) and the inclusion of new customers change this theoretical cost-benefit ratio. Our take-up, close to 50%, clearly implies that costs are doubled. On the other hand, hair dressers may draw in new customers or rotate through the same number of customers. Benefits might thus increase thanks to a growing pool of customers, with the hypothesis of 1 woman avoiding depression per hair dresser becoming too conservative. Furthermore, mental health improvement is likely to carry positive externalities in other dimensions, such as in children's health (see Pierce et al. (2020) for a meta-analysis) or child development (Ramchandani et al. 2005). In particular, the finding by Baranov et al. (2020) in rural Pakistan suggested that the women's therapy-induced financial empowerment translated into better time and money investment in children. Finally, if the intervention shows an improvement in mental health and well-being, the scope of proximate services which could be targeted with such training campaigns is large and may generate substantial economic and social benefits at low cost. Finally, since this is a relatively new kind of intervention, a sizable budget is being devoted to communication. It is likely that ensuing training events will cost less, in particular if they are carried out at the neighborhood level.

2 Research design

2.1 Intervention

The intervention consists of three days of mental health awareness and detection training, as part of the *Heal by Hair* programme provided by the Bluemind Foundation, an NGO working to raise awareness about mental health in Africa. Training takes place at the Pullman Hotel in Abidjan, a high-end hotel located in the center of the city. A first batch of 22 hair-dressers were trained in April, 2022 in Abidjan as part of a pilot training to test the logistics and collecting feedback from trained hair-dressers. These 22 hair-dressers are not included in the second batch of training, that is the main study. The second batch of training also took place in Abidjan from January 31 to February 2, 2023, and training sessions are scheduled in Cameroon (April 2023) and Togo (July, 2023), with expansion to other countries (Senegal, Mali, DRC) planned for the future.

During three days, participants receive information and participate in role play mimicking everyday conversations that may arise in interactions with their customers. The curriculum of the training is summarized in Table 1. At the end of the training, a test is carried out, and hair-dressers who score 80% or above on the test will receive a certification as "mental health ambassadors". The training is carried out by a group of health care professionals specializing in mental health.[1] During the training session that took place in January - February 2023, only one hair dresser out of 74 scored below 80% and did not receive certification (more on this in section 3.

Upon completion of the training, hair-dressers are added to a WhatsApp group to stimulate continued interactions between participants at the training. They are also invited to monthly group sessions held with a psychologist, and will be given a list of certified mental health caregivers for referrals. Finally, all certified hair-dressers will be offered help to enroll in the National Fund for Social Security under a new regime targeting independent workers. Under the regime, independents pay a monthly fee proportional to their income, and are entitled to maternity leave, sick leave (when

[1] Most are Psychiatrists, Psychologists or Therapists. In the first training session, all but one were women and the upcoming session will also feature a majority of women, coming not only from Côte d'Ivoire, but also from Cameroon, Mali, Senegal & Togo.

more than 14 days) and pensions. The training is thus a one shot intervention, but hairdressers keep engaging with the NGO over time. Hairdresser–customer interactions are also dynamic, with heterogeneity in the frequency, quality, and length of interactions. Therefore, from the point of view of hairdressers, we believe that it is better to think of the intervention as a program that hairdressers join, rather than as a one-time event. Under this program, they receive continued guidance (through monthly meetings and interactions with medical professionals and among participants in the training) on how to best act in unfamiliar situations, and how to react to new situations (most of these exchanges will take place in WhatsApp groups). These interactions should also contribute to keeping the acquired knowledge fresh. The mental health awareness raising activites of hair dressers are monitored through a chatbot through which hair dressers collect points, for each customer made aware. The chatbot acts as an incentive for hair dressers to raise awareness, since they collect points for each customer informed about mental health. A three-part verification is in place, where both the customer and the hair dresser need to register on the chatbot, and checks that customers have indeed been made aware are carried out by the Bluemind foundation. For various levels of points collected, hair dressers unlock rewards, corresponding to a gift basket of beauty products for the salon (20 customers made aware), to a short training in cosmetics (80 customers made aware). It is important to note that the points given are unconditional on any positive returns to the awareness raising. Customers are not asked if what their hair dresser told them was useful or not, or whether they benefited from it, only whether or not the hair dresser told them about mental health. In our follow-up surveys, in which we clearly state that we are unaffiliated with the NGO, we also make it clear that the individual information gathered will not be relevant for either the foundation or the hair dresser.

These features of the program (repeated interactions and incentives) matter both for inference and for cost-benefit analysis. For inference, we attempt to analyze heterogeneity in two ways: first, by running two post-training surveys, at 4 months and 12 months after the training, assessing whether or not returns to training are increasing in time, or on the contrary decrease (for example through information being forgotten); second, through collecting information on the frequency, quality, and length of interactions between hair dressers and their customers, which we can use in a heterogeneity analysis. However, given that these are both potential outcomes of the program (improved interactions may generate a willingness to visit hairdressers more frequently), and candidates for heterogeneity variables, we prefer focusing on

Table 1: Curriculum of the Heal-by-hair training session

Day	Training content
1	Mental health: introduction, history, origins Deconstructing prejudice and destigmatizing language Becoming a mental health ambassador Ethics and confidentiality Taking care of one's mental health
2	Listening and empathy Depression and signs of psychological distress Intimate violence, grief and trauma Case studies
3	Evaluation Leadership Introduction to personal branding

well-being at 4 months and 12 months as overall outcomes and consider interactions as channels through which the effect runs. Furthermore, we will be able to gain access to the chatbot used in the incentives program, and will therefore be able to check whether any effects found run primarily through hair dressers who were active in raising awareness or not.

Recruitment was done through an information campaign taking place in the field and online, aiming to generate at least 750 applications to the program. Out of the 750 applications, 300 were considered eligible for the program, and the corresponding hair-dressers received word of their eligibility. A stakeholders' committee, composed of representatives for hair-dressers in Abidjan, and a scientific committee composed of mental health professionals, was responsible for selecting eligible candidates. Selection was based on motivation, understanding of the nature of the training, and availability to attend the training event. Due to logistical constraints, only 150 hair-dressers could join the training, initially planned for November 21-23 but finally taking place in January-February, 2023[2], and the final selection of hair-dressers took place through computerized randomization. Prior to randomization, however, all 300 hair-dressers were approached and asked to participate in a baseline survey. They were also asked to produce a list of 10 regular customers, a random subset of which (5 per hair-dresser) will be approached and asked to participate in a survey.

[2]For health reasons on the organizers' side.

The randomization procedure took place after the baseline data collection (see timeline in section 2.4). Strata were created according to the variables deemed most likely to influence impact of the intervention on mental health (Bruhn and McKenzie 2009). Three variables were used: baseline mental health scores of customers, baseline score on a multidimensional scale of perceived social support, and the municipality the hair-dresser operates in. We created dummies based on threshold levels defined for baseline mental health and the multidimensional scale of perceived support. With three distinct municipalities, this creates 12 strata for the randomization. At the level of hair-dressers, there can be no blinding of assignments. However, customers of hair-dressers need not necessarily be aware that their hair-dresser has undergone training. hair-dressers are free to tell their customers that they did (it would be difficult to prevent this information from spreading), but we do not believe that this puts pressure on customers to deliver positive feedback on their hair-dresser through for example reporting a better than actual mental health status. Nevertheless, to limit the risk as much as possible, we dissociate our survey from the intervention, framing it as a survey on women's living conditions and the interactions they may have with their hair-dressers, making no mention to customers of the training event. Another risk of information on the training event is if a selected portion of customers change hair-dressers when they get information on their training status. We however believe that this risk is limited as well: the survey aims to interview well-established customers (regulars) of hair-dressers, who are likely to find it costly to switch. Nevertheless, by comparing baseline and endline data on customer-hair-dresser pairs, we will be able to assess whether or not any significant switching occurred in favor of trained hair-dressers. Furthermore, the size of the three municipalities (regrouping more than 2 million people) implies that most hair-dressers will not apply to or be involved in the program whatsoever, and the likelihood that both a certified and an uncertified applicant be positioned in the very same uptake area (likely to be limited to a street, or a couple of streets) is small.

2.1.1 External validity concerns

The nature and scope of the intervention (constrained to three municipalities of one major city), implied that we do not seek external validity in this protocol. However, a discussion about the relevance of the target group is still warranted to gauge the importance of the research findings and anticipate future implementations. Table 2 shows summary statistics on female hair-dressers drawn from the

heal by hair pool of hair-dressers selected for training (Baseline data), and from female hair-dressers (and other women) surveyed under the Harmonized Survey on Households' Living Standards (EHCVM), collected in 2018-2019. Columns 1 and [2,3,4] thus enable a direct comparison between hair-dressers surveyed as part of heal by hair, and hair-dressers surveyed as part of a nationally representative household survey. The main drawback is that of low sample sizes of hair-dressers from the EHCVM. Nevertheless, the information gives some perspective to the nature of the recruitment process involved in selecting hair-dressers under the program. First, HBH and EHCVM hair-dressers are similar to a large extent. Comparing column 1 and 2, for Abidjan, only for age are the confidence intervals for the means overlapping at the 10% significance level. This likely has to do with the fact that the foundation sought to enroll hair salon owners primarily. This choice was made in order to ensure that they would be invested in their occupations, and remain as available as possible to engage with the foundation after the training session.

Table 2: Summary statistics, heal-by-hair and EHCVM hair-dressers

	HBH baseline hair-dressers	hair-dressers EHCVM			Other occupations EHCVM	
		Abidjan	Other urban	Rural areas	Abidjan	Other urban
Age	35.9	29	25.8	25.9	34	34.9
	(0.42)	(1.53)	(0.77)	(1.0)	(0.46)	(0.26)
Primary achieved	0.488	0.32	0.18	0.22	0.31	0.18
	(0.03)	(0.1)	(0.04)	(0.06)	(0.02)	(0.01)
Married	0.317	0.4	0.318	0.49	0.475	0.563
	(0.03)	(0.1)	(0.05)	(0.07)	(0.02)	(0.01)
CIV citizen	0.96	0.96	0.87	0.92	0.87	0.84
	(0.01)	(0.04)	(0.03)	(0.04)	(0.01)	(0.01)
Muslim	0.1	0.24	0.25	0.12	0.42	0.48
	(0.02)	(0.09)	(0.05)	(0.05)	(0.02)	(0.01)
Household head	0.2	0.2	0.24	0.18	0.23	0.24
	(0.03)	(0.08)	(0.05)	(0.06)	(0.02)	(0.01)
Observations	246	25	88	49	659	2931

Source: Column (1): Baseline data collection on hair-dressers under the Heal by Hair project. Columns (2) to (6): data from the Harmonized Survey on Households' Living Standards 2018-2019 (EHCVM).

2.2 Hypotheses

The underlying theory of change of the Heal-by-hair program is that of improved listening. Hair-dressers are not intended to provide medical care but to become "active listeners".

The first hypothesis is that hairdressers trained by medical professionals have tools to be more emphatic to customers, to detect issues and if necessary, refer them to appropriate caregivers. The first-stage hypothesis comprises the following outcomes: the extent to which hair-dressers themselves and their customers report a change in the nature and/or frequency of interactions. We will refer to these as *first-stage outcomes*.

Second, the intention of the project is to make a visible difference in women's lives, in terms of well-being and mental health. Thus, our **second hypothesis** holds that the intervention improves customers' well-being and mental health. Customers should feel more listened to, and they will receive new information on how to reach mental health professionals if needed. This should be visible through increases in the psychometric scores detailed below. The first and second stage outcomes constitute our *primary outcomes*.

Last, our **third hypothesis** focuses on the impact of the intervention on women's participation in intra-household decision-making and in the labor market. It builds on the finding of Baranov et al. (2020) in rural Pakistan, that improved mental health led to financially empowered women. We believe that increases in mental health may lead to stronger bargaining power of women in their households. We are also interested in the effects of well-being and mental health in the labor market. If the intervention leads to better health, this should positively affect both employment status and productivity (as measured by income, primarily in self-employment). We refer to all these outcomes as our *secondary outcomes*.

2.2.1 First-stage outcomes

At the level of hair-dressers, we are interested in their perceived importance as significant others in the lives of their customers. Three questions allow us to get a sense of the quality and length of interactions between hair-dressers and their

customers: "Do customers typically talk to you about their personal lives?" (Question C13a), "Do customers typically talk to you about difficulties they are facing in their personal lives?" (Question C13b), and "When customers evoke personal difficulties, do you feel that you can make a difference?" (Question C13c). Answers to all these questions lie on a five-step Likert scale coded as follows: 1—Always; 2—Very often; 3—Quite often; 4—Rarely; 5—Never. At the level of the customers, first-stage outcomes similarly refer to the extent to which they engage with and open up to their hair-dressers, with similar questions on whether they mention their personal lives, and difficulties therein: "Do you discuss your private life with your hair-dresser?", "Do you talk to your hair-dresser about difficulties in your life?" and "When you talk to your hair-dresser about difficulties in your life, does she show support and/or bring forth ideas for a solution?" (Questions C5a,C5b and C5c in the customer questionnaire). Answers to all these questions lie on a five-step Likert scale coded as follows: 1—Always; 2—Very often; 3—Quite often; 4—Rarely; 5—Never. We also include a multiple choice question on the reasons for choosing and sticking with a particular hair-dresser, which includes financial aspects, geographical aspects but also a dimension on her perceived qualities as a human being (question C3 in the client questionnaire). Based on this question, we will create a dummy variable equal to 1 if customers mention "qualities as a human being" as a reason for sticking with their hair-dresser, and 0 otherwise. All the variables using Likert scales will be standardized in the analysis.

Finally, we also include questions relative to "significant others" using the multi-dimensional scale of perceived support (MSPSS), conceived by Zimet et al. (1988). The scale contains 12 items[3], and we are using the French translation, which was found to possess excellent internal consistency among a set of French women 4 month after childbirth (Denis, Callahan, and Bouvard 2015). A Likert scale is used for each item of the scale, coded as follows: 0—Strongly disagree; 1—Disagree; 2—Neutral; 3—Agree; 4—Strongly agree. Summing the answers thus gives a 0 to 48 point scale of perceived support, which we standardize in the analysis. These first stage outcomes

[3]Respondents answer to twelve questions: "Over the last two weeks: (1) There is a special person who is around when I am in need; (2) There is a special person with whom I can share my joys and sorrows; (3) My family really tries to help me; (4) I get the emotional help and support I need from my family; (5) I have a special person who is a real source of comfort to me; (6) My friends really try to help me; (7) I can count on my friends when things go wrong; (8) I can talk about my problems with my family; (9) I have friends with whom I can share my joys and sorrows; (10) There is a special person in my life who cares about my feelings; (11) My family is willing to help me make decisions; (12) I can talk about my problems with my friends."

thus number 5 at the level of customers, and 3 at the level of hair-dressers. All variables are summarized in Table A.1 below.

2.2.2 Second stage outcomes

The second stage outcomes of interest relate to change in well-being and mental health, measured both in customers and in hair-dressers. To capture overall well-being, we rely on the WHO-5 Well-being index, a short version of the WHO-10 well-being index, itself a contraction of a 28-item index conceived by the WHO. [4] The final raw score is totaling the figures of the five answers. "The raw score ranges from 0 to 25, 0 representing worst possible and 25 representing best possible quality of life. To obtain a percentage score ranging from 0 to 100, the raw score is multiplied by 4. A percentage score of 0 represents worst possible, whereas a score of 100 represents best possible quality of life." The WHO-5 has been shown to possess good clinimetric properties (Topp et al. 2015), and validity in a variety of settings both to measure well-being and as a screening tool for depression. Answers again lie on a Likert scale, ranging from 0 to 5: 0—Never; 1—From time to time; 2—Less than half the time; 3—More than half the time; 4—Most of the time; 5—All the time. Each answer gives a certain number of points.

In addition to the WHO-5, we also use the well-known PHQ-9 (Kroenke and Spitzer 2002), which has been tested and shown to be a valid screening tool for depression in several African countries (e.g. Ethiopia: Gelaye et al. (2013); South Africa: Bhana et al. (2015), Nigeria: Adewuya, Ola, and Afolabi (2006)). In Côte d'Ivoire and Ghana, the only available study found close to acceptable internal consistency of the instrument, and concluded on the appropriateness of the use of the score sum as a screening tool for depression (Barthel et al. 2015). The PHQ9 is based on 9 units and uses a four-step Likert scale (0—Not at all; 1—Several days; 2—More than half of days; 3—Nearly every day). In the analysis, standardized versions of the two instruments will be used to test second stage impacts of the intervention.

The follow-up survey includes new questions on the nature of the interactions between hair-dressers and their customers. Hair-dressers are asked: "How do you

[4] Respondents answer to five questions: "Over the last two weeks: (1) I have felt cheerful and in good spirits; (2) I have felt calm and relaxed; (3) I have felt active and vigorous; (4) I woke up feeling fresh and rested; (5) My daily life has been filled with things that interested me."

react if a customer shares her health issue with you? "; "How do you react if a customer shares her mental health issue with you?". Customers are asked to think about their last three encounters: "Think about your last three encounters with your hairdresser, did you talk about your private life?"; "Think about your last three encounters with your hairdresser, did you talk about your health?"; "Think about your last three encounters with your hairdresser, did you talk about your mental health?"; "If yes to previous, did your hairdresser listen to it?"; "If yes to previous, did your hairdresser show support or give advice?"; "If yes to previous, did your hairdresser refer a health professional?".

Hair-dressers and customers are also asked with whom they may talk about mental health issues they might have had in the previous month: "Last month, did you experience mental health issues?"; "If yes to previous, did you discuss your mental issues with someone?"; "If yes to previous, with whom did you discuss your mental issues?". Since we do not have this information at baseline, we only make use of these questions in a POST estimator analysis (sample NS, see section 2.4).

2.2.3 Secondary outcomes

Beside our primary outcomes, we are also interested in a set of secondary outcomes related to the labor market and to norms. As discussed in the introduction, a broad literature has shown negative labor market effects from mental illness, related to unemployment, absenteeism and productivity. We therefore gather information on employment and absenteeism in the questionnaire, and use last week's reported income as a proxy for productivity. To measure employment, we rely on a simple question: "Do you have a job?" (Question C8 in the client questionnaire). Because of budgetary constraints, we have decided not to base the employment dummy on the full set of standard labor market questions, involving whether or not individuals collected wood, made reparations on their house, etc. that are typically present in labor market surveys. Instead, surveyors' will be specifically trained on the definition of employment, in order to minimize discrepancies between the employment rate found in the survey and the ILO employment rate. The dummy variable on employment constitutes the first of the secondary outcomes. To capture absenteeism, we make use of questions C10a and C10b in the client questionnaire, asking about the number of hours worked last week, and the number of hours normally worked. We then create a

dummy variable equal to one if the declared number of hours is inferior to the typical number of hours. Finally, in order to proxy for productivity, we rely on a measure on self-reported income from last week (Question C14 in the client questionnaire). The variable will be winsorized at the 1st and 99th percentiles, and no imputations will be made for missing values.

Our second dimension of interest, that of gender norms, builds on the finding of Baranov et al. (2020) in rural Pakistan, namely that improved mental health led to financially empowered women. We, therefore, ask married women a series of questions on decision-making in general in the household, and on who decides on the allocation of individual income earned. Questions E2a - E2e and E4 to E6 gather information on decision-making in the household. For each dimension—food expenses, buying expensive items, children's education, health care for the interviewee, health care for children, the use of the interviewee's earnings, the use of the husband's earnings, and when to visit friends and family—four answers are possible: 1—Woman decides; 2—Husband decides; 3—Woman and husband decide together, 4—Other (explain). After examining and coding the 'Other' category, we create a dummy variable for each subquestion, equal to one if the interviewee is involved in the decision making in each dimension. Finally, we also make use of a frequently included question in the norms literature, namely the cases where a husband's beating of his wife are justified. The answers are limited to Yes and No and we create dummies for each of the five circumstances cited: going out without asking permission, neglecting the children, raising arguments with husband, refusing sex, and burning the food. We create a dummy variable equal to one if women find it acceptable for husbands to beat their wives under any of the above scenarios. Summing up, we test three secondary outcomes related to the labor market, and seven related to norms. All are dummies, with the exception of the income level, which we standardize in the analysis.

2.3 Basic methodological framework

The study is set up as a clustered randomized controlled trial. Given strong concerns for endogeneity in the domain of mental health, a randomized allocation of the treatment is the most appropriate tool for impact evaluation. Three waves of the survey are planned—a baseline survey, a short-term survey four months after training is completed, and a medium-term survey, one year after completed training.

2.4 Data

Intended sample

The analysis takes place at the individual level and will be mainly focused on the customers of hair-dressers. The intended population of hair-dressers—and thus of clusters—is equal to 300, and represents hair-dressers interested in attending a three-day training event on first response to mental health issues, divided equally from three municipalities of the district of Abidjan: Yopougon, Cocody and Abobo.[5] Interviewing 5 customers for each hair-dresser implies that an intended sample size of 1500 customers is attained. A cost-benefit analysis based on unitary interview costs, budgetary concerns and a simulation of minimum detectable effect sizes were used to determine the intended sample size.

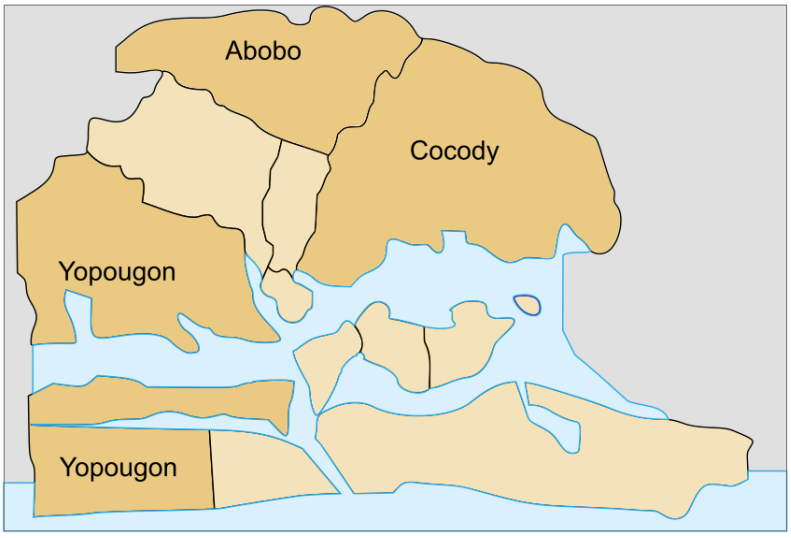

Figure 1: Intervention municipalities and their location in Abidjan

In power calculations, we stick to standard assumptions on power ($1 - \beta = 0.8$) and significance ($\alpha = 0.05$), and assume no additional relevant covariates at this stage. With a set cluster size of five customers per hair-dresser, the minimum detectable effect (MDE) depends on the attrition rate of hair-dressers and of customers, and

[5]These municipalities are large, with estimated populations in 2014 of 447 023 for Cocody, 1 071 543 for Yopougon and 1 030 658 for Abobo

on the intraclass correlation coefficient (ICC). We develop two scenarios: a quite pessimistic one, and a reasonably optimistic one. Under the optimistic scenario, there is 10% attrition in hair-dressers, and 20% in customers, and the ICC stands at 0.2. This implies an MDE of 0.224. Under the pessimistic scenario, there is 30% attrition in hair-dressers and 40% in customers, and the ICC stands at 0.35. This implies an MDE of 0.292. When deciding on the intended sample size, we computed the theoretical number of clusters needed to achieve a MDE of 0.3 under the pessimistic scenario, namely 286. Rounding the figure to 300 thus provides some margin. It is however our belief, and hope, that the ICC and attrition rates will be as low as possible, enabling detection of a an effect lower than 0.2 standard deviations.[6] The pseudo-pilot and the baseline data collection will provide and initial idea of what can be reasonably expected, and will through baseline distributions of outcome variables enable us to get a sense of the percentage increases or decreases we can expect to detect. Furthermore, improvements to power will come from the use of the ANCOVA estimator, and from the stratified randomization as discussed above, and so the above power calculations—which already rely on unknowns—only give an underestimated hint to the effect sizes that can be detected. For the ANCOVA specification in particular, McKenzie (2012) shows that the ratio of variance between ANCOVA and difference in difference is equal to: $\frac{2}{1+\rho}$, with ρ now referring to the autocorrelation in outcomes within individuals between subsequent survey rounds. He also discusses the option of not performing a baseline survey, showing that disregarding it for an additional follow-up survey may improve power, especially in those cases where autocorrelation in outcomes is weak. In cases where autocorrelation is zero, correcting for the baseline outcome value in the regression does not reduce variance, and comes at a cost of one degree of freedom. In our case, we believe that there will be low to medium autocorrelation in outcomes related to mental health. In a previous survey run in Mali by the author, the autocorrelation of a depression dummy[7] in surveys six months apart was 0.25, and if autocorrelations in the present survey are of the same magnitude, including the baseline outcome values will improve precision. Apart from contributing to precision, we find it valuable to conduct a baseline survey in order to explore heterogeneity (see section 3).

Data collection and processing

[6] It suffices that the ICC stand at 0.1, and attrition at 10% for the MDE to be inferior to 0.2.
[7] Defined using the MINI neuropsychiatric interview.

All of the data collection was scheduled to be done by a private contractor based in Abidjan and specializing in field surveys. The hair-dresser and client questionnaires (see Appendix) are coded into ODK and data will be collected using smartphones or tablets. 15 surveyors will carry out the survey, interviewing a target population of 1 800 women (300 hair-dressers and 1500 customers) in the three targeted municipalities. The allocation of surveyors was random, and a new set of surveyors will be used for the follow-up surveys.[8] The estimated time allocated to each interview is 30 - 45 minutes, and surveyors can thus be expected to complete some 6 to 10 interviews per day. It is estimated that data collection in the field for each wave should take no more than 20 days. The approved budget for the contractor funds 1800 questionnaires, and at the level of customers, all will be done to interview five regular customers per hair-dresser. At the level of hair-dressers, however, given that only 300 applications were selected and make up the randomization pool, any refusals to take part in the survey will not be compensated by additional interviews. Customers from these hair-dressers may however still be interviewed, subject to the hair-dressers' approval.

[8]The survey provider will change by the first follow-up, and we will ensure that the new one will randomly allocate surveyors.

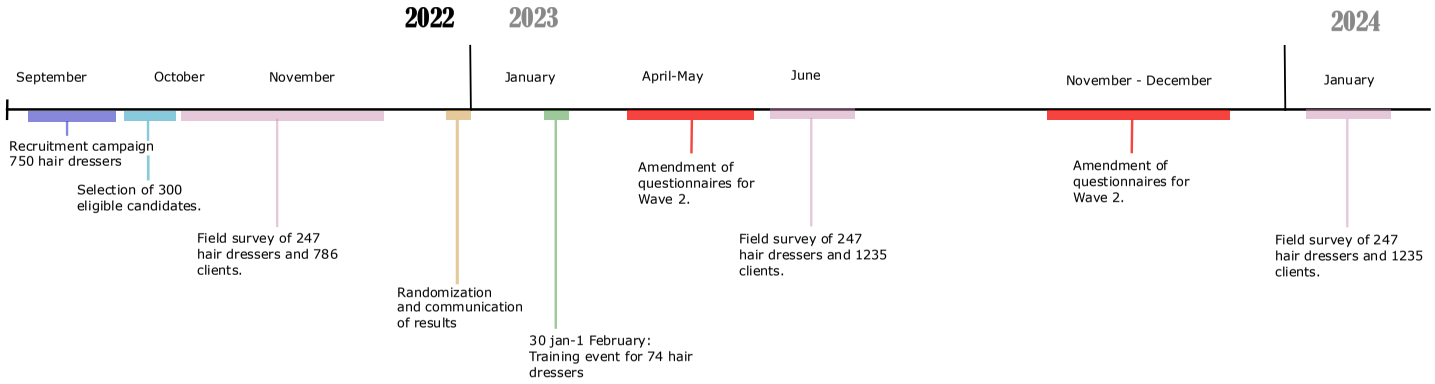

Figure 2: Timeline of research implementation

The timeline for the survey is outlined in Figure 2. Surveyor training took place starting October 5, 2022, with a short two-day pseudo-pilot planned at the end of the two-day training. Analysis of pilot data immediately followed, and baseline data collection begun on October 15, in order to be finished by early November. In the end, due to difficulties finding suitable dates and times for meetings, the last customers were interviewed in December. Randomization then followed in December, and hair-dressers who were selected were notified in early January. The two follow-up surveys will be fielded four months and one year after the intervention.

For the baseline collection, severe communication issues and lack of respect of the agreed-upon numbers of questionnaires in due time means that the initial target numbers were not reached. Specifically, only 247 hair dressers were interviewed (partly due to refusals, which cannot be replaced), and some 786 clients (from 190 hair dressers). Despite numerous attempts to resolve the situation and interview additional customers, we chose to end the collaboration with the contractor and seek another contractor for the following rounds of data collection. The main reason, except for difficulties in collaboration, was a time constraint: selected hair dressers had to be given notice about their acceptance into training with reasonable anticipation, in order for them to make arrangements to attend the training. We did not see it reasonable to interview new customers after hair dressers had been given notice about their participation in the study. These lowered numbers imply that we have made modifications to the initial empirical strategy. Rather than working with a single sample of 1800 customers, we now rely on two samples. The baseline sample (BS) of 786 customers attributed to 190 hair dressers, and a new sample (NS) of an intended number of 1235 customers (five for each of the 247 hair dressers present in the baseline survey). For the upcoming second round of the survey, four months after training, we have therefore partnered with a new contractor, whose role is to re-interview all 247 hair dressers and 786 customers of the baseline sample, and 449 new customers associated with the hair dressers for whom no customers, or too few customers, were interviewed at baseline. While we can interview new customers in the follow-up, at the level of hair dressers—and given that only 300 applications were selected and make up the randomization pool—baseline refusals to take part in the survey will not be compensated by additional interviews. Since hair dressers may believe—falsely and despite stating the contrary in the introduction to the survey—that it is in their interest to show a positive impact of the training to the surveyors, they may wish to push specific customers to the survey team. This would more problematic than

at baseline, since they are now aware of their treatment status. We therefore choose to only recruit new customers at random, outside the salon, without involving hair dressers. Finally, for new customers we also require that they declare being customers of the hair dresser at least since 2022 (and thus before the training).

In the end, we thus have one hair dresser sample, with baseline information from 247 hair dressers, and two customer samples: a baseline-sample (BS), with 786 customers of 190 hair dressers, and (2) a POST sample, with an intended 1235 customers of 247 hair dressers (786 customers from the baseline sample, and 449 customers for whom there is no baseline information). We will treat these samples differently in the empirical analysis below. Specifically, we will use an ANCOVA specification for our "baseline sample", and a simple POST estimator for the "new sample".

The lower-than-expected numbers achieved at baseline have implications for power. With the baseline sample of 786 customers and 190 hair dressers, the minimum detectable effect is 0.264 and 0.343 for the optimistic and pessimistic scenarios, respectively. Using instead a POST estimator (McKenzie 2012) and no baseline survey, on 1235 customers of 247 hair dressers, and using the same assumptions as above, we find a minimum detectable effect of 0.231 under the optimistic scenario, and 0.275 under the pessimistic one.

Given that we now also run a POST estimation without baseline information, we will—as stated in section 2.2.2 make use of additional questions introduced into the questionnaire in the new round of the survey. Specifically, for customers, these concerns the nature of interactions for the last three visits to the hair dresser, as well as a question on whether or not the customer experienced mental health issues in the past month, and if so if she discussed it and with whom. For hair dressers, we also add the question on whether they experienced mental health issues in the past month, and if so if they discussed it and with whom (see Table A1 in the Appendix: round 2 questions are marked with a *).

Given that data collection is computer assisted (CAPI), coding will produce automated skips, minimizing filter errors and omitted conditionals. All variables will be given acceptable intervals to minimize occurrences of responses outside their logical range. To follow the deployment in the field, collected data will be continuously synchronized to a server, enabling the principal investigator to screen filled-out forms

as they arrive, detecting any anomalies and immediately adjusting through the field supervisor.

Attrition is a cause for concern given the longitudinal data collection that is planned. Although we cannot completely prevent attrition since participation is on a voluntary basis, surveyors will be instructed to clearly explain to surveyed individuals that two follow-ups will take place, and that in order for their contribution to the study to be as complete as possible, they are expected to be available for the two follow-ups as well. Naturally, everyone is free to leave at any moment, and is informed of that option, but by clarifying the expected contribution from the outset, we hope to keep attrition to a minimum.

2.5 Balance at baseline

Tables 3 and 4 show covariate balance at the hair dresser and customer levels, respectively. Table 3 concerns sample 1, making use of the baseline information in an ANCOVA specification. Standard errors here are clustered at the hair dresser level. Except for household size, no discrepancies are statistically significant, and covariates seem on the whole balanced, as would be expected from a randomization procedure.

For sample 2, we do not have baseline information for the full sample (only for the 786 customers in Table 1), and we therefore contend with showing that there is balance among the full set of 247 hair dressers from which all customers in sample 2 will be drawn. Table 4 indeed suggests that this is the case, with only the proportion of married women showing a statistically significant difference, with a larger proportion of married women in the control group as opposed to the treated group.

3 Empirical analysis

Our setup is a clustered randomized trial with one baseline survey and two follow-ups. As shown by McKenzie (2012), the ANCOVA estimator has larger power than the difference-in-difference estimator, all the more so when autocorrelation in outcomes is low. We believe that autocorrelation of about 0.25 is to be expected for our outcomes related to mental health, which would imply a 37.5% lower variance (and sample size) when using ANCOVA rather than difference-in-difference. For some of our outcomes, such as last week's income, autocorrelation may be even weaker and the rationale to

Table 3: Balance table: customers of 190 hair dressers

Variable	(1) 0 Mean/SE	(2) 1 Mean/SE	(3) Total Mean/SE	T-test P-value (1)-(2)
Age	34.28	32.89	33.45	0.12
	(0.74)	(0.50)	(0.43)	
Married	0.30	0.25	0.27	0.25
	(0.03)	(0.03)	(0.02)	
Ivorian	0.98	0.99	0.98	0.52
	(0.01)	(0.01)	(0.00)	
Born in Abidjan	0.48	0.55	0.53	0.11
	(0.03)	(0.03)	(0.02)	
Ever in school	0.78	0.80	0.79	0.51
	(0.03)	(0.03)	(0.02)	
Currently in school	0.13	0.16	0.15	0.24
	(0.02)	(0.02)	(0.01)	
Score dur. goods	4.48	4.43	4.45	0.78
	(0.13)	(0.12)	(0.09)	
Spouse of HH	0.09	0.06	0.08	0.15
	(0.02)	(0.01)	(0.01)	
Household size	5.09	4.72	4.87	0.06*
	(0.16)	(0.12)	(0.10)	
Has children	0.63	0.59	0.61	0.40
	(0.03)	(0.03)	(0.02)	
Elder sibling	0.20	0.17	0.18	0.54
	(0.03)	(0.02)	(0.02)	
Length of haircut	2.56	2.50	2.52	0.77
	(0.17)	(0.16)	(0.11)	
No private talk	0.22	0.25	0.24	0.44
	(0.03)	(0.03)	(0.02)	
Monthly cuts or less	0.26	0.25	0.26	0.72
	(0.03)	(0.03)	(0.02)	
Working	0.58	0.51	0.54	0.11
	(0.03)	(0.03)	(0.02)	
Health issue last 4w	0.19	0.14	0.16	0.26
	(0.03)	(0.02)	(0.02)	
WHO-5 well-being index	16.17	16.58	16.42	0.56
	(0.53)	(0.45)	(0.34)	
PHQ-9 score	5.93	5.81	5.86	0.84
	(0.44)	(0.40)	(0.29)	
MSPSS score	47.51	48.14	47.89	0.53
	(0.82)	(0.55)	(0.46)	
Locus of control score	12.25	11.77	11.96	0.22
	(0.32)	(0.23)	(0.19)	
Quality of interaction	7.54	7.42	7.47	0.54
	(0.13)	(0.14)	(0.10)	
N	317	469	786	
Clusters	79	111	190	

Notes: The value displayed for t-tests are p-values. Standard errors are clustered at the level of hair dressers. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 4: Balance table: 247 hair dressers

Variable	(1) 0 Mean/SE	(2) 1 Mean/SE	(3) Total Mean/SE	T-test P-value (1)-(2)
Age	35.42	36.37	35.98	0.28
	(0.68)	(0.56)	(0.43)	
Married	0.39	0.27	0.32	0.05*
	(0.05)	(0.04)	(0.03)	
Ivorian	0.94	0.97	0.96	0.24
	(0.02)	(0.01)	(0.01)	
Born in Abidjan	0.34	0.40	0.37	0.33
	(0.05)	(0.04)	(0.03)	
Ever in school	0.80	0.86	0.84	0.21
	(0.04)	(0.03)	(0.02)	
Currently in school	0.02	0.02	0.02	0.97
	(0.01)	(0.01)	(0.01)	
Score dur. goods	4.06	3.87	3.95	0.16
	(0.11)	(0.08)	(0.07)	
Spouse of HH	0.09	0.05	0.07	0.32
	(0.03)	(0.02)	(0.02)	
Household size	5.18	4.93	5.03	0.40
	(0.22)	(0.19)	(0.14)	
Has children	0.79	0.86	0.83	0.20
	(0.04)	(0.03)	(0.02)	
Elder sibling	0.18	0.19	0.19	0.79
	(0.04)	(0.03)	(0.02)	
Owns hair salon	0.67	0.74	0.71	0.26
	(0.05)	(0.04)	(0.03)	
Length of haircut	3.85	2.53	3.07	0.14
	(0.88)	(0.17)	(0.38)	
No private talk	0.01	0.03	0.02	0.18
	(0.01)	(0.02)	(0.01)	
Monthly cuts or less	0.87	0.83	0.85	0.35
	(0.03)	(0.03)	(0.02)	
Hours worked typical week	49.87	48.15	48.85	0.63
	(2.75)	(2.35)	(1.78)	
Health issue last 4w	0.19	0.22	0.21	0.55
	(0.04)	(0.03)	(0.03)	
WHO-5 well-being index	17.62	17.29	17.43	0.63
	(0.51)	(0.47)	(0.35)	
PHQ-9 score	6.09	6.67	6.43	0.41
	(0.53)	(0.47)	(0.35)	
MSPSS score	47.11	46.29	46.62	0.49
	(0.91)	(0.78)	(0.59)	
Locus of control score	11.83	11.87	11.85	0.92
	(0.32)	(0.21)	(0.18)	
Quality of interaction	7.70	7.68	7.69	0.93
	(0.15)	(0.13)	(0.10)	
N	101	146	247	

Notes: The value displayed for t-tests are p-values. Standard errors are robust. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

use ANCOVA even stronger. Our first specification thus reads:

$$Y_{i,1} = \alpha + \beta TREAT_i + \gamma Y_{i,0} + \delta STRATA_i + \epsilon_i \quad (1)$$

Where β is the outcome of interest, ϵ a stochastic error term, and $Y_{i,1}$ and $Y_{i,0}$ refer to the outcomes after and before the intervention, respectively. $TREAT_i$ is a dummy variable equal to 1 if the hair-dresser was selected to attend the three-day training, and 0 if the hair-dresser is not selected to attend the training. $STRATA_i$ is a vector of strata dummies as discussed in section 2.1. This specification aims at evaluating impacts at two points in time: at 4 months, and 12 months. We also provide a secondary estimation function, pooling data from the two follow-ups to estimate an average effect over time:

$$Y_{i,t} = \alpha + \beta TREAT_i + \gamma Y_{i,0} + \sigma_1 I(t=1) + \sigma_2 I(t=2) + \delta STRATA_i + \epsilon_i \quad (2)$$

Where $I(t=1)$ and $I(t=2)$ are time dummies. This second specification improves power, and may be useful to detect effects of weaker magnitude. The two above equations will be used to regress the outcomes detailed in section 2.2.

For our second sample (2), where we lack baseline information for 465 customers, we use a POST estimator to measure the impact of the program on outcomes. As previously mentioned, when autocorrelation in outcomes is low, adding baseline values of the outcome to the speficiation does little to improve power (McKenzie 2012). Since we believe that this is the case in our setting, we prefer making use of our second sample to confirm the results of the ANCOVA analysis. We therefore also run the regression:

$$Y_i = \alpha + \beta TREAT_i + \delta STRATA_i + \epsilon_i \quad (3)$$

Where, as in equation 2, Y_i is evaluated at two points in time: 4 months after the intervention, and 12 months after.

Finally, given that take-up was far from universal among selected hair dressers (73 out of 150 hair dressers succesfully underwent training), we believe that a LATE specification may be a succesful addition to our main results. We still consider that the ITT effect is the policy relevant parameter of interest, but that the LATE estimate is an interesting complement. In particular, if attendance among selected

hair dressers was random, the LATE would give us an idea of the impact to expect in a scenario with 100% compliance.

Missing values and outliers

Missing values should in theory only arise when interviewees refuse to answer or don't know the answer to a question. Sensitive questions on mental health, income or decision making may produce missing answers. For income, when respondents refuse to declare their income earned in the reference week, we ask them to declare the range within which that income falls. In all cases, we will ignore missing values in the analysis. If a variable has more than 25% missing values, we discard it from the analysis. Income levels will be winsorized at the 1% level. Remaining outcome variables are either dummies or based on Likert scales, so outliers are unlikely to be an issue.

Multiple hypothesis testing

Section 2.2 detailed the variables tested in the study. A total of 17 (5 first stage, 2 second stage, and 10 secondary outcomes) outcomes will be tested for customers of hair-dressers. With a 5% probability of a Type I error, under a no *true* impact scenario, we would thus expect about one outcome coefficient to be statistically different from zero. A computation of the familywise error rate (FWER) shows that the probability of at least one false positive stands at 0.582.[9] To asymptotically control the FWER, we thus implement the Romano Wolf multiple hypothesis correction (Romano and Wolf 2005), which produces adjusted P-values for the vector of outcome coefficients through a stepdown multiple hypothesis testing algorithm and a bootstrap resampling procedure. Both the unadjusted and the adjusted P-values will be presented in the final article.

Heterogeneity

Recent advances in Machine Learning have produced relevant applications for analyzing heterogeneity in both experimental and non-experimental settings. These part with the hypothesis of linearity, allow for high-order interactions between explanatory variables, and—importantly—alleviate concerns for *ad hoc* choices of candidate variables to analyze heterogeneity through. We make use of the *Causal*

[9] FWER=$1-(1-\alpha)^{17}$

forests (Athey, Tibshirani, Wager, et al. 2019; Athey and Wager 2019) algorithm to assess heterogeneity of responsiveness to the intervention.

Understanding across which dimensions heterogeneity operates is not straightforward, even in a machine learning setting. Although the individual treatment effect distribution can be estimated from the causal forest algorithm, there is no straightforward way of asserting which variables play an important role in shaping heterogeneity. Split frequencies and other measures of variable importance are sometimes used, but their interpretation is difficult due to the fact that causal forests—and other machine learning algorithms—are not constant in terms of model selection. Instead of searching for variables associated with heterogeneity (a difficult exercise in a non-parametric setting), we follow Davis and Heller (2020) who partition their sample into quartiles based on the estimated individual treatment effect obtained from the causal forests algorithm. The procedure, thoroughly described in Davis and Heller (2020), first regresses the outcome on an interaction between the treatment dummy and the predictions from the causal forest. Being able to exclude zero implies that the predictions are indeed relevant. In addition, we also compute the Rank-Weighted Average Treatment Effect (RATE) (Yadlowsky et al. 2021). The statistic, which has a known distribution, can be used as a test for whether heterogeneity exists and is identified by the algorithm. After investigating the existence of heterogeneity, we then partition the sample into quartiles according to the predicted CATEs from the forest algorithm, and present descriptive statistics by quartile on a set of sociodemographic variables. Testing for the differences between first-quartile outcomes and fourth-quartile outcomes, like in Davis and Heller (2020) thus gives us an idea of the variables associated with heterogeneous response.

Testing for the existence of heterogeneity will be done for all outcome variables, but tables of summary statistics by predicted treatment will only be produced for those outcomes (if any) that show significant heterogeneity.

4 References

Adewuya, Abiodun O, Bola A Ola, and Olusegun O Afolabi (2006). "Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students". *Journal of Affective Disorders* 96.1-2, pp. 89–93.

Athey, Susan, Julie Tibshirani, Stefan Wager, et al. (2019). "Generalized random forests". *The Annals of Statistics* 47.2, pp. 1148–1178.

Athey, Susan and Stefan Wager (2019). "Estimating treatment effects with causal forests: An application". Working Paper No. 3786, Stanford University, Graduate School of Business.

Baranov, Victoria et al. (2020). "Maternal depression, women's empowerment, and parental investment: evidence from a randomized controlled trial". *American Economic Review* 110.3, pp. 824–59.

Barker, Nathan et al. (2022). "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress". *American Economic Review: Insights* 4.4, pp. 527–45.

Barthel, Dana et al. (2015). "Screening for depression in pregnant women from Côte d'Ivoire and Ghana: Psychometric properties of the Patient Health Questionnaire-9". *Journal of Affective Disorders* 187, pp. 232–240.

Bhana, Arvin et al. (2015). "The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa". *BMC psychiatry* 15.1, pp. 1–9.

Bhat, Bhargav et al. (2022). *The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes*. Tech. rep. National Bureau of Economic Research.

Bindt, Carola et al. (2012). "Antepartum depression and anxiety associated with disability in African women: cross-sectional results from the CDS study in Ghana and cote d'Ivoire". *PloS one* 7.10, e48396.

Bluemind Foundation (2021). *Heal by Hair: Rapport d'études croisées entre les femmes africaines et leurs coiffeuses*.

Bruhn, Miriam and David McKenzie (2009). "In pursuit of balance: Randomization in practice in development field experiments". *American Economic Journal: Applied economics* 1.4, pp. 200–232.

Chemin, Matthieu, Joost De Laat, and Johannes Haushofer (2013). "Negative rainfall shocks increase levels of the stress hormone cortisol among poor farmers in Kenya". *Available at SSRN 2294171*.

Cowen, Emory L, Ellis L Gesten, Mary Boike, et al. (1979). "Hairdressers as caregivers. I. A descriptive profile of interpersonal help-giving involvements". *American Journal of Community Psychology* 7.6, p. 633.

Cowen, Emory L, Ellis L Gesten, Edward Davidson, et al. (1981). "Hairdressers as caregivers II: Relationships between helper characteristics and helping behaviors and feelings". *The Journal of prevention* 1, pp. 225–239.

Davis, Jonathan MV and Sara B Heller (2020). "Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs". *Review of Economics and Statistics* 102.4, pp. 664–677.

Denis, A, S Callahan, and M Bouvard (2015). "Evaluation of the French version of the multidimensional scale of perceived social support during the postpartum period". *Maternal and child health journal* 19.6, pp. 1245–1251.

Essah, Doris S (2008). "Fashioning the Nation: Hairdressing, Professionalism and the Performance of Gender in Ghana, 1900-2006." PhD thesis.

Frasquilho, Diana et al. (2015). "Mental health outcomes in times of economic recession: a systematic literature review". *BMC public health* 16.1, pp. 1–40.

Gelaye, Bizu et al. (2013). "Validity of the patient health questionnaire-9 for depression screening and diagnosis in East Africa". *Psychiatry research* 210.2, pp. 653–661.

Gilbert, Barnabas J et al. (2015). "Assessing development assistance for mental health in developing countries: 2007–2013". *PLoS Med* 12.6, e1001834.

Haushofer, Johannes, Robert Mudida, and Jeremy P Shapiro (2020). *The comparative impact of cash transfers and a psychotherapy program on psychological and economic well-being*. Tech. rep. National Bureau of Economic Research.

Kroenke, Kurt and Robert L Spitzer (2002). *The PHQ-9: a new depression diagnostic and severity measure*.

Kuhn, Andreas, Rafael Lalive, and Josef Zweimüller (2009). "The public health costs of job loss". *Journal of health economics* 28.6, pp. 1099–1115.

Layard, Richard (2013). "Mental health: the new frontier for labour economics". *IZA Journal of Labor Policy* 2.1, pp. 1–16.

— (2017). "The economics of mental health". *IZA World of Labor*.

Lund, Crick et al. (2020). "Economic impacts of mental health interventions in low and middle-income countries: a systematic review and meta-analysis".

Maselko, Joanna et al. (2020). "Effectiveness of a peer-delivered, psychosocial intervention on maternal depression and child development at 3 years postnatal: a cluster randomised trial in Pakistan". *The Lancet Psychiatry* 7.9, pp. 775–787.

McKenzie, David (2012). "Beyond baseline and follow-up: The case for more T in experiments". *Journal of Development Economics* 99.2, pp. 210–221.

Mullainathan, Sendhil and Eldar Shafir (2013). *Scarcity: Why having too little means so much*. Macmillan.

Patel, Vikram et al. (2017). "The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial". *The Lancet* 389.10065, pp. 176–185.

Pierce, Matthias et al. (2020). "Effects of parental mental illness on children's physical health: systematic review and meta-analysis". *The British Journal of Psychiatry* 217.1, pp. 354–363.

Ramchandani, Paul et al. (2005). "Paternal depression in the postnatal period and child development: a prospective population study". *The Lancet* 365.9478, pp. 2201–2205.

Ridley, Matthew et al. (2020). "Poverty, depression, and anxiety: Causal evidence and mechanisms". *Science* 370.6522, eaay0214.

Romano, Joseph P and Michael Wolf (2005). "Stepwise multiple testing as formalized data snooping". *Econometrica* 73.4, pp. 1237–1282.

Sankoh, Osman, Stephen Sevalie, and Mark Weston (2018). "Mental health in Africa". *The Lancet Global Health* 6.9, e954–e955.

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan (2016). "The psychological lives of the poor". *American Economic Review* 106.5, pp. 435–440.

Skeen, Sarah et al. (2010). "Meeting the millennium development goals in Sub-saharan Africa: what about mental health?" *International Review of Psychiatry* 22.6, pp. 624–631.

Topp, Christian Winther et al. (2015). "The WHO-5 Well-Being Index: a systematic review of the literature". *Psychotherapy and Psychosomatics* 84.3, pp. 167–176.

Vigo, Daniel, Graham Thornicroft, and Rifat Atun (2016). "Estimating the true global burden of mental illness". *The Lancet Psychiatry* 3.2, pp. 171–178.

Wemakor, Anthony and Kofi Akohene Mensah (2016). "Association between maternal depression and child stunting in Northern Ghana: a cross-sectional study". *BMC public health* 16, pp. 1–7.

WHO (2016). "Global Health Estimates 2016: Burden of Disease by Cause". *Age, Sex, by Country and by Region.*

Wilson, Tracey E et al. (2008). "Hair salon stylists as breast cancer prevention lay health advisors for African American and Afro-Caribbean women". *Journal of Health Care for the Poor and Underserved* 19.1, pp. 216–226.

Xiao, Wen et al. (2020). "Value co-creation effects on transaction cost, relational capital, and loyalty of hair salon customers: Results and implications of a Ghanaian study". *Journal of Psychology in Africa* 30.3, pp. 217–224.

Yadlowsky, Steve et al. (2021). "Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects". *arXiv preprint arXiv:2111.07966*.

Zimet, Gregory D et al. (1988). "The multidimensional scale of perceived social support". *Journal of Personality Assessment* 52.1, pp. 30–41.

5 Appendix

A Summary of outcome variables

Table A.1: List of outcome variables

Type of outcome	Respondent	Indicator/Question	Answer universe
First-stage	Hair-dresser	Do customers typically talk to you about their personal lives?	Five-step Likert (1-Always; 2-Very often; 3-Quite often; 4-Rarely; 5-Never)
		Do customers typically talk to you about difficulties they are facing in their personal lives?	
		When customers evoke personal difficulties, do you feel that you can make a difference?	
	Customer	Do you discuss your private life with your hair-dresser?	Five-step Likert (1-Always; 2-Very often; 3-Quite often; 4-Rarely; 5-Never)
		Do you talk to your hair-dresser about difficulties in your life?	
		If yes to previous, does she show support and/or bring forth ideas for a solution?	
		Mentions qualities as a human being as reason for sticking with current hair-dresser	Yes (1) / No (0)
First-stage	Customer	MSPSS (multidimensional scale of social support)	Index of 12 items with five-step Likert scale (0-Strongly disagree; 1-Disagree; 2-Neutral; 3-Agree; 4-Strongly agree)

Continued on next page

Table A.1: List of outcome variables (Continued)

First-stage	Customer	Think about your last three encounters with your hairdresser. Did you talk about your private life?*	Yes (1) / No (0)
		If yes to previous, do you feel that your hairdresser listened?*	
		Think about your last three encounters with your hairdresser. Did you talk about your health?*	
		If yes to previous, did your hairdresser show support or give advice?*	
		Think about your last three encounters with your hairdresser. Did you talk about your mental health?*	
		If yes to previous, did your hairdresser refer you to a health professional?*	
Second-stage	Hair-dresser	Last month, did you experience mental health issues?*	Yes (1) / No (0)
		If yes to previous, did you discuss your mental issues with someone?*	

Continued on next page

Table A.1: List of outcome variables (Continued)

Second-stage	Hair-dresser	If yes to previous, with whom did you discuss your mental issues?*	1-A relative; 2-A friend; 3-Another hair-dresser part of the Heal-by-hair program; 4-A mental health professional not related to the Heal-by-hair program; 5-A mental health professional part ot the Heal-by-hair program; 6-Other
	Customer	Last month, did you experience mental health issues?*	Yes (1) / No (0)
		If yes to previous, did you discuss your mental issues with someone?*	
		If yes to previous, with whom did you discuss your mental issues?*	1-A relative; 2-A friend; 3-Another hair-dresser part of the Heal-by-hair program; 4-A mental health professional not related to the Heal-by-hair program; 5-A mental health professional part ot the Heal-by-hair program; 6-Other

Continued on next page

Table A.1: List of outcome variables (Continued)

	Customer	WHO-5 Well being index	Index of 5 items with six-step Likert scale (0-Never 1-From time to time; 2-Less than half of the time; 3-More than half of the time; 4-Most of the time; 5-All of the time)
		Personal health questionnaire (PHQ-9)	Index of 9 items with four-step Likert scale (0-Not at all; 1-Several days; 2-More than half the days; 3-Nearly every day)
Secondary	Customer	Do you have a job?	Yes (1) / No (0)
		How many hours do you work in your main job during a typical week?	[0,120]
		How many hours did you work in your main job last week?	[0,120]
		How much did you make in your main job last week?	[0,9999999]
		In your household, who makes decisions when it comes to buying food?	Dummy=1 if woman involved in decision
		In your household, who makes decisions when it comes to buying expensive items?	Dummy=1 if woman involved in decision
		In your household, who makes decisions when it comes to children's education?	Dummy=1 if woman involved in decision

Continued on next page

Table A.1: List of outcome variables (Continued)

Secondary	Customer	In your household, who makes decisions when it comes to seeking medical care when you are ill?	Dummy=1 if woman involved in decision
		In your household, who makes decisions when it comes to seeking medical care when the children are ill?	Dummy=1 if woman involved in decision
		In your household, who makes decisions concerning the use of your earnings?	Dummy=1 if woman involved in decision
		In your household, who makes decisions concerning the use of your husband's earnings?	Dummy=1 if woman involved in decision
		In your household, who makes decisions concerning visits to friends and family?	Dummy=1 if woman involved in decision
		When is it justified for a husband to beat his wife ? At least one event cited.	Yes (1) / No (0)

6 Administrative information

A Funding

The present study is financed by the Bluemind Foundation, which received funding from the Fund for Innovation in Development, a part of French Overseas Development Assistance. The study has also received funding from Université Paris-Saclay and Université Paris 1 Panthéon - Sorbonne.

B Internal Review Board

The study has been approved by the IRB of the Paris School of Economics. The certificate number is 2022-021 and the protocol was approved on October 20, 2022.

C Declaration of interest

The authors have no conflicts of interest to declare.

D Assistance

The authors were assisted by Marie-Alix de Putter, president of Bluemind Foundation, who provided valuable input on the NGO's activities; Ernest Mobio Akre, responsibly for data collection in the initial phase of the project, and Jonathan Beugre N'guessan, who will be responsible for data collection for the two follow-ups.

I, Björn Nilsson, the corresponding author of this manuscript, certify that the contributors' and conflicts of interest statements included in this paper are correct and have been approved by all co-authors.