Stay in the Game: A Randomized Controlled Trial of a Sports and Life Skills Program for Vulnerable Youth in Liberia

LORI BEAMAN
Northwestern University

SYLVAN HERSKOWITZ
International Food Policy Research Institute

NIALL KELEHER
Innovations for Poverty Action

JEREMY MAGRUDER
University of California, Berkeley

I. Introduction

Across the world, youth are found playing sports wherever there is enough space to kick a ball or swing a bat: on fields, beaches, empty streets, or vacant lots. Few activities capture the excitement, energy, and imagination of youth as much as sports. Beginning in the 2000s, a wave of exuberance brought increased popularity to the use of sports for the promotion of development outcomes. Sport-for-development (SFD) programs take many forms but broadly aim to develop prosocial behaviors and labor force potential among young adults, with a frequent emphasis on at-risk and difficult-to-reach populations.

We are grateful to Mercy Corps for its willingness to evaluate the program discussed in this paper. Without an openness to scrutiny and facilitation of sharing administrative data, this research would not have been possible. We thank Stephanie De Mel for her research assistance and T. Wordploe Marwolo, Joseph Kamara, Dackermue Dolo, Abel Welwean, and other field research staff of Innovations for Poverty Action for their hard work and dedication in carrying out data collection. We are grateful to Innovations for Poverty Action for assistance running the study. We thank the study participants for generously giving their time. We thank Chris Blattman, Julian Jamison, and Margaret Sheridan for sharing questionnaires. Guanghua Chi provided valuable contributions to the qualitative data analysis. This research was funded by the International Initiative for Impact Evaluation (grant OW4/1094), the Swedish International Development Cooperation Agency (award 90954-S-001), the International Growth Centre (grant 1-VCH-VSLE-VXXX-51300), and Mercy Corps. Data are provided through Dataverse (https://doi.org/10.7910/DVN/ERSL7G). Contact the corresponding author, Sylvan Herskowitz, at s.herskowitz@cgiar.org.

Electronically published August 11, 2021

Economic Development and Cultural Change, volume 70, number 1, October 2021. © 2021 The University of Chicago. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits reuse of the work with attribution. Published by The University of Chicago Press. https://doi.org/10.1086/711651
In 2001, the United Nations established the Office on Sport for Development and Peace. Four years later, they declared 2005 the International Year of Sport and Physical Education. “Sport for development” is defined as a broad effort to “engage people from disadvantaged communities in physical activity projects that have an overarching aim of achieving various social, cultural, physical, economic or health-related outcomes” (Adair 2014, 3). A recent review of SFD programs found 955 organizations that engage exclusively in sports programming and more than 2,000 additional organizations that incorporate sports into their programs (Svensson and Woods 2017). Other sources corroborate these estimates, claiming that SFD programs reach tens of millions of youth throughout the world (Adair 2014). Our own efforts to track expenditures on SFD programs suggest that global expenditures exceed hundreds of millions of dollars per year. Both in terms of participation and expenditure, SFD programs operate at a massive global scale.

SFD proponents frequently tout sports-centered programs as an impactful form of direct intervention for at-risk youth as well as an effective entry point for complementary programs targeting difficult-to-reach populations. In particular, it is frequently asserted that these programs can improve psychosocial outcomes and soft skills of participants and that these benefits will, in turn, lead to better labor market outcomes for marginalized youth. Despite this large mobilization of resources, there is little existing evidence to either substantiate or refute these claims. Unlike interventions involving labor force training, there is little experimental evidence to guide policy making in the SFD sector.

1 We follow this definition from Adair (2014) to preserve a focus on sports programming as a form of intervention. “Sports for development and peace,” “sport for change,” “development through sport,” and “sport and development” are other commonly used terms for the types of programs we discuss in this paper.

2 This estimate is based on reports and documentation found online. We examined organizations that self-register through https://www.sportanddev.org and http://www.streetfootballworld.org. The eight members of the steering committee of the International Platform for Sport and Development alone could be linked to more than US$105 million in project budgets for SFD-related activities. Searching for additional financial disclosures led us to US$219.5 million in annual expenditures. This is likely a considerable undercounting of overall SFD expenditures. Further details are provided in app. B.

3 If SFC expenditures are between US$300 million and US$600 million per year, this would be comparable to 0.3%–0.55% of all global aid given by the Organization for Economic Cooperation and Development’s 30 members of the Development Assistance Committee and equivalent to or greater than nine of these countries’ total contributions (OECD 2019). According to a recent World Bank publication, it is also close to the total amount spent in 2014 on noncommunicable diseases (Bendavid et al. 2017).

4 Many organizations claim that these benefits are even wider ranging, with the United Nations asserting that benefits include gender equality, social integration, development of social capital, peace building, conflict prevention and resolution, trauma relief, and economic development (https://www.un.org/sport/content/why-sport/overview).
In Liberia, the site of this study, more than three-fifths of the country’s population is below the age of 25. This foreshadows continued rapid population growth and presents considerable demographic challenges engaging such a large youth population going forward. Although the country’s second civil war ended in 2003, the memories and trauma of one of the continent’s bloodiest and most protracted conflicts still looms large on the national psyche. Policy makers and international actors have therefore been worried that failure to engage with youth and help them find productive outlets would risk destabilizing the country and eruption of civil unrest.

In this paper, we investigate the effect of a life skills program mediated through sports groups on psychosocial well-being and labor force outcomes among youth in Monrovia, Liberia’s largest city. Our evidence derives from a randomized control trial of an SFD program conducted by Mercy Corps, a large international nongovernmental organization with similar programs in more than 25 countries. The study took place in nine communities around Monrovia, with 1,200 youth invited to participate in the Mercy Corps program and another 1,200 assigned to the control group. Similar to the claims of many other SFD proponents and practitioners, the stated aims of Mercy Corps’s SFD program, Sport for Change (SFC), are to improve psychosocial or socioemotional behaviors among participants, increasing their resiliency and readiness for productive labor force participation. The SFC method centers on the formation of youth groups, engagement with these groups through competitive sports, and facilitation of a complementary life skills curriculum. Random assignment of individuals to youth groups and a control group allows us to estimate the causal impact of SFC on a range of psychosocial and labor market outcomes.

Anticipating benefits on psychosocial outcomes, our data collection was designed to allow construction of five measures of psychosocial outcomes, which we aggregate into an overall psychosocial index (PSI). Corroborating the implementer’s theory of change, we see a strong positive baseline correlation between psychosocial well-being and labor force measures. However, our analysis reveals limited evidence of meaningful direct impacts of the program on psychosocial outcomes, with a point estimate of −0.014 standard deviations and a 95% confidence interval ranging from −0.098 to 0.07. We also do not find any evidence of improved resiliency among program beneficiaries faced with challenging life events.

Liberia’s 2014–15 Household Income and Expenditure Survey shows that 76.4% of Liberia’s population is under the age of 35, 62.5% is under 25, and 44.5% is under 15.
The proposed theory of change suggested that psychosocial improvements caused by the program would lead to improved labor force outcomes. And yet, despite a lack of measurable psychosocial impacts, we do see a statistically significant increase in labor outcomes. Our results show an increase in our aggregate labor force index (LFI) among study participants with a point estimate of 0.12 standard deviations and a 95% confidence interval from 0.03 to 0.20 standard deviations. Point estimates for labor supply and earnings both suggest increases of approximately 12%.

Given the lack of evidence of impacts on psychosocial measures and as a hypothesized causal pathway, we examine whether peer effects may have contributed to results on labor market and psychosocial outcomes. We find limited evidence to support the importance of peer effects in this context and are unable to explain the results through exposure to high-performing (or low-performing) peers or the presence of preexisting social ties in assigned groups. Finally, we explore heterogeneity of the treatment effects in order to better understand who is benefitting from the program. It appears that along all the dimensions that we tried (young, uneducated, female, untrained), labor force benefits of the program are bigger for each of these marginalized populations.

II. Related Work

SFD proponents assert a wide range of benefits for program participants, arguing that sports are an effective way to improve prosocial and psychosocial life skills as well as labor outcomes for participating youth. Many organizations trust that benefits result directly and naturally from sports participation: sports may inherently teach valuable life lessons and foster prosocial skills. As a result, some organizations focus exclusively on expanding opportunities for youth to play sports, such as through broad distribution of soccer balls in developing countries. Other organizations have invested considerable time and effort designing and incorporating complementary life skills training into their sports activities.\(^6\) The body of existing evidence for SFD programs of any form is limited. A large portion of the existing literature focuses on improving the design and strategic implementation of SFD programs (e.g., Hartmann and Kwauk 2011; Jeanes 2013).

Alongside these prescriptive suggestions for SFD are implicit claims about expected positive impacts on participants. In one example, Kidd (2008, 371) states “[SFD] has brought considerable benefit to many children and youth in the

\(^6\) See Ball to All (https://www.balltoall.org/) as an example of a sports-only program and Peace Players International (https://www.peaceplayers.org/) and Grassroot Soccer (https://www.grassrootsoccer.org/) as examples of programs with complementary soft skills programming.
countries where it is conducted.” Rookwood (2008) suggests that soccer builds trust, respect, and self-discipline. Others highlight the positive relationship between sports and psychosocial development and resiliency (Petitpas et al. 2005; Berlin et al. 2007; Henley et al. 2007; Perkins and Noam 2007). Frequently studied examples of the SFD movement range from Nairobi to the United States, broadly framing SFD programs as a mechanism to prepare young adults to utilize prosocial skills in a variety of long-term applications.⁷

While these studies provide important psychological and sociological grounding of hypothesized effects and potential benefits, the evidence relies heavily on case studies and theoretical models that are unsubstantiated by empirical evidence. Three recent reviews of this literature have agreed that the evidence base is lacking in rigor and that the few studies with individual-level data are systematically underpowered in sample size and fail to address concerns of endogeneity between program participation and program outcomes (Holt and Jones 2007; Burnett 2009; Coalter 2010).

Since these reviews, there have been a handful of recent studies using more rigorous empirical methods. Panter-Brick et al. (2018) assess the impact of a Mercy Corps program in Syria that is similar to their SFC program in Liberia. Jordanian and Syrian youth were organized into groups centered on a range of activities (often, but not exclusively, sports based), while locally identified facilitators provided life skills and psychosocial support. The authors found modest improvements in some psychosocial outcomes, with bigger effects among youth exposed to higher levels of trauma. However, the study was hindered by high levels of attrition (74% after 1 year) and did not examine economic outcomes. Three other recent studies also focused on sports programming as a way to promote reconciliation across groups. In India, Israel, and Iraq, these studies found that participating in sports increased prosocial cross-group behavior resulting from interactions across castes, ethnicities, or religions (Ditlmann and Samii 2016; Mousa 2020; Lowe 2021).

This paper also relates to the broader body of work linking noncognitive and psychosocial factors to labor market outcomes. Two recent surveys of this literature document how cognitive skills do not fully explain labor market outcomes, attributing a large part of the unexplained variation to socioemotional skills (Heckman and Kautz 2012; Kautz et al. 2014). These “character” skills, or life skills, are seen as necessary for the realization of investments in cognitive

⁷ For example, the Mathare Youth Sports Association in Nairobi (Morris et al. 2004; Kidd 2008; Coalter 2009), the Physically Active Youth program in Namibia (Donnelly et al. 2007), and the midnight basketball leagues across numerous locations in the United States (Hartmann and Depro 2006).
skills. The authors go on to suggest that interventions that improve these soft skills may have an impact on economic and welfare outcomes as well.

In response to this body of work linking soft skills and economic outcomes, considerable interest and programming have been put into psychosocial interventions in traumatized populations, although findings have been inconsistent (Underwood 2018). On the positive end, Adoho et al. (2014) present evidence from a randomized control trial of an intensive program on economic empowerment and life skills training for young Liberian women. They find large improvements in psychosocial measures of self-confidence and anxiety, a 47% increase in employment, and an 80% increase in earnings for program participants. Ibarraran et al. (2014) also present positive, but more modest, results on the impacts of a soft skills training program for youth in the Dominican Republic, finding improvements in earnings (8%), formality of work (10%), and measures of noncognitive skills (0.08–0.12 standard deviations) but no increase in overall labor force participation. Calero and Rozo (2016) show mixed results in another experimental study among youth in Brazil’s favelas, finding that training aimed at reducing risky behaviors increased income but improved behaviors only among participants with higher preexisting levels of socio-emotional skills. Groh et al. (2016) find no effects of a soft skills training program on female youth employment among Jordanian community college graduates. And in the setting most similar to that of our paper, Blattman, Jamison, and Sheridan (2017) provide mixed evidence in their study of prosocial programming in a sample of high-risk ex-combatants in Liberia. The authors find that a cognitive behavioral training (CBT) intervention led to a 0.25–0.3 standard deviation decrease in antisocial behaviors 1 month after completing the CBT program. However, this effect persisted in follow-ups 1 year after the intervention only among CBT beneficiaries who also received a substantial cash grant of US$200. They find no persistent impacts of the program on economic outcomes.

This paper bridges the literature on sports and psychosocial behavior with the literature on psychosocial training as related to labor market outcomes. In the context of our study, we examine a setting different from that examined in other sports program evaluations. Namely, the youth population in Monrovia is less segregated than that in Lowe (2021) and not immediately affected by national conflict as in Ditlmann and Samii (2016) and Mousa (2020). We build on this literature by looking at longer-term impacts of an SFC program on a broader set of psychosocial measures (approximately 1 year after the intervention). We also leverage a larger sample of study participants randomized into sports groups. In addition, the data we collected allow us to examine labor force outcomes as well as psychosocial outcomes. While theory driving the SFD
literature presupposes the connection between sports and psychosocial outcomes, it is not a given, and it is not obvious that changes in labor force outcomes are predicated on changes in psychosocial outcomes. In this paper, we present evidence intended to unpack the linkages between sports programs, psychosocial outcomes, and labor force outcomes.

III. Context and Program Design
At the time of the study, the president of Liberia, Ellen Johnson-Sirleaf, expressed concern over high levels of youth unemployment as a destabilizing factor for the country (Dunmore 2013). Overall labor force participation was and remains very low in Liberia, estimated by the International Labor Organization at roughly 60% (http://data.un.org/en/iso/lr.html). The central role of youth as both victims and combatants in the Liberian conflict amplifies a sense of urgency among policy makers and international actors to find ways to engage young populations in positive activities as well as to help those directly and indirectly affected by the war (for a summary of the qualitative research, see app. F; apps. A–F are available online).

With this backdrop, Mercy Corps launched the Promoting Sustainable Partnerships for Economic Transformation (PROSPECTS) initiative in 2012 in Montserrado County, Liberia, where Monrovia is located. In this paper, we evaluate the impact of one program within PROSPECTS, namely, the SFC program. SFC targeted vulnerable, out-of-school youth between the ages of 15 and 25 with little or no prior formal work experience. This target population was broadly regarded as unskilled and unemployable. The SFC program in Liberia was designed to use sports groups as a means of attracting and engaging vulnerable youth to participate in prosocial activities. The SFC methodology, designed by Mercy Corps, seeks to leverage the beneficial potential of team-based sports with life skills training. As such, Mercy Corps integrated five core life skills into the SFC program: (1) resilience, (2) strategy making and planning, (3) teamwork and trust building, (4) self-esteem, and (5) constructive communication (for details of the SFC program, see app. D; for an example of a session schedule, see fig. A.1; figs. A.1, A.2 are available online). The methodology closely mirrored the international SFC approach that Mercy Corps has implemented in more than 25 countries to engage youth in postconflict settings. Bundling sports team practices with life skills activities, Mercy Corps expected positive impacts on participants’ psychosocial outcomes and improved resiliency to adverse life events. Improved psychosocial well-being and resiliency were expected to create a foundation of workforce readiness for participants to enter into formal employment or launch their own small businesses.
Contemporaneous to the timing of this study, Liberia experienced a national emergency as a result of an outbreak of Ebola in West Africa. Although the first cases of Ebola were reported in Liberia in March 2014, the SFC program had concluded before the first recorded Ebola deaths were documented in Monrovia. Over the course of the following year, a total of 10,675 cases of Ebola were recorded in Liberia, with 4,809 resulting in deaths. Nearly one-third of these deaths occurred in Montserrado County, where this study was implemented. While the Ebola outbreak did not affect implementation of the program, in the months following the end of SFC, it had a considerable impact on daily life in Monrovia, and the crisis was ongoing at the time of the endline survey.

IV. Experimental Design and Data
A. Recruitment and Random Assignment
We worked with Mercy Corps in nine urban communities of Monrovia to identify a pool of eligible youth for participation in the SFC program. Public announcements were made in each community, inviting young adults between the ages of 15 and 25 to attend a meeting organized by Mercy Corps to register for potential involvement in the Mercy Corps program and in the study. We organized one recruitment event per community. At the event, applicants were randomly selected to participate in the Mercy Corps program via a public lottery. To prevent gender imbalance, we stratified the lottery by applicant gender. Women and men formed separate lines. Each individual drew a ticket indicating group assignment from a covered bucket, and they were unable to change their assignment after it was revealed.

In total, 1,200 individuals were assigned to an SFC team while 1,200 were not invited into the Mercy Corps program and thus served as the study’s control group. With labor force outcomes as the primary outcome of interest, power calculations showed that our design allows us to detect a small minimum effect of 0.16 standard deviations with 80% power, assuming that the standard deviation of treatment effects across communities is 0.1 and that baseline covariates explain 20% of the variation in final outcomes.

At the registration events in each community, registration forms were completed by the entire pool of eligible applicants regardless of group assignment. Registration staff recorded basic demographic information, including age,
gender, and schooling, and extensive tracking and contact information in anticipation of the baseline interviews. Additional details on the recruitment of participants into the study and random assignment to youth sports groups can be found in appendix C.9

Registration events proceeded sequentially one community at a time. The initial event occurred in the West Point community in July 2013, and the final event was held in the Logan Town community in February 2014. Table 1 shows the timing of the registration event and baseline survey and the number of study participants by community. The number of registered participants varied by community from 160 to 480 depending on the anticipated size of its youth population. Mercy Corps sought 50% female participants. Although there was some variance of female participation levels across communities, gender balance across treatments within each community was broadly preserved through our stratification procedures. Ultimately, women comprised 52.2% of the treatment group and 51.8% of the control group.

B. Baseline Summary Statistics

In the days immediately following the registration event in each community, Innovations for Poverty Action conducted in-person baseline interviews with all registrants following the schedule in table 1.10 Between completion of the registration events and administration of the baseline survey, there was very low attrition: only five of 2,400 registrants refused to be interviewed or could not be found after the registration event.

9 The initial research and program design of PROSPECTS also included a cash-for-work (CFW) program that involved an opportunity for participants to earn money by collecting recyclable materials. Implementation challenges, compounded by low interest in the program, meant that very few people assigned to the CFW group actually participated in the program. In total, 1,200 individuals were assigned to the CFW program. Of those assigned to CFW, 976 attended at least one CFW meeting, and only 467 attended more than two meetings. As a result, the analysis in this paper focuses solely on the SFC component of PROSPECTS. Random assignment of the full set of 3,000 participants in the original research design assigned 20% of participants to the CFW program only, 20% to the SFC program only, 20% to both programs, and 40% to the control group. As a result of implementation challenges as well as budgetary and logistical constraints on conducting follow-up surveys amid the outbreak of Ebola, it was decided to exclude respondents in the CFW-only treatment from the endline survey.

10 For logistical reasons, it was not possible to both generate a sample frame of individuals and complete baseline surveys before treatment assignment. This design feature did not lead to differential baseline survey response rates. Of the five registrants who did not participate in the baseline survey, two were in the treatment group and three were in the control group, with attrition rates of 0.17% and 0.25%, respectively.
Table 2 shows summary statistics and balance of participants at baseline by program treatment status. The average age in the sample is 21 years old; 83% had completed primary school, and slightly more than 25% had completed secondary school. Just over 43% had some form of employment at baseline. Among those working, respondents worked roughly 28 hours per week and earned approximately US$16.50.

While most baseline variables are well balanced across treatment status, we note an imbalance in five of the 35 variables, which is slightly higher than we would expect from chance. Respondents in the control group had higher baseline measures of self-esteem and numeracy but lower scores on the depression, anxiety, and stress index and lower 3-month income. For these variables, we note that even though differences are statistically significant, the magnitudes of these differences are all very small and unlikely to be economically meaningful.

The sole exception is for 3-month income, which has a relatively large difference in means. However, this baseline imbalance is no longer significant after implementing the inverse hyperbolic sine transformation, which mitigates the influence of outliers in the data with unusually large reported earnings. Still, given that longer recall windows are prone to greater amounts of measurement error, we exclude the 3-month income measure from our primary analysis of labor outcomes and give preference to reported 7-day income, and we show robustness of our main results to including it in appendix A.

Overall, we may have been concerned about whether baseline interviews occurring after assignment to treatment (but before the intervention began) could have affected reported baseline responses. However, the limited extent
| Demographics and household characteristics: | Control | Treatment | p-Value of Difference |
|------------------------------------------|--------|-----------|----------------------|
| Female                                   | .518   | .522      | .85                  |
| Age                                      | 20.926 | 20.798    | .26                  |
| Head of household                        | .131   | .143      | .42                  |
| Household size                           | 6.762  | 6.598     | .26                  |
| Mother known to be living                | .868   | .870      | .91                  |
| Father known to be living                | .713   | .717      | .84                  |
| Has at least one child                   | .454   | .432      | .28                  |
| Christian                                | .876   | .863      | .37                  |
| Muslim                                   | .111   | .125      | .29                  |
| Matched connections in sample            | 2.390  | 2.419     | .72                  |

| Education and cognitive ability:         |        |           |                      |
| Completed primary school                 | .835   | .836      | .95                  |
| Completed secondary school               | .276   | .265      | .57                  |
| Highest grade completed                  | 11.389 | 11.350    | .81                  |
| Numeracy (0–10)                          | 5.951  | 5.784     | .04**                |
| Digits forward (0–8)                     | 5.176  | 5.124     | .42                  |
| Digits backward (0–8)                    | 1.981  | 1.924     | .24                  |
| First word recall (0–10)                 | 3.059  | 3.072     | .91                  |
| Second word recall (0–10)                | 2.578  | 2.558     | .85                  |
| Ravens score (0–3)                       | 1.753  | 1.726     | .51                  |
| Risk aversion (0−6)                      | 3.785  | 3.758     | .79                  |

| Psychosocial measures:                   |        |           |                      |
| Subjective welfare, today (standardized) | −.011  | .011      | .59                  |
| Self-esteem index                        | 22.977 | 22.495    | .00***               |
| Locus of control index                   | 24.089 | 24.031    | .63                  |
| Aggression index                         | 2.603  | 2.546     | .63                  |
| Risky behaviors index                    | 1.501  | 1.468     | .67                  |
| Depression, anxiety and stress score (DASS21) | 21.987 | 23.282    | .02**                |
| Psychosocial index (standardized)        | .000   | −.065     | .12                  |

| Labor force and financial measures:      |        |           |                      |
| Labor force index (standardized)         | −.000  | .021      | .62                  |
| Worked last week                         | .431   | .436      | .83                  |
| Hours worked last week                   | 11.546 | 12.344    | .33                  |
| Hours worked last week (if positive)     | 27.495 | 29.016    | .27                  |
| IHST (hours worked last week)            | 1.513  | 1.563     | .53                  |
| Income earned last week                  | 6.501  | 6.647     | .80                  |
| Income earned last week (if positive)    | 16.342 | 16.999    | .57                  |
| IHST (income earned last week)           | 1.190  | 1.180     | .88                  |
| Income earned last 3 months              | 48.060 | 57.840    | .01**                |
| Income earned last 3 months (if positive)| 80.078 | 96.591    | .00***               |
| IHST (income earned last 3 months)       | 2.625  | 2.738     | .25                  |

| N                                        | 1,197  | 1,198     |                      |

**Note.** Top 1% of values for hours worked, income earned last week, and income earned past 3 months are trimmed and set to missing. IHST = inverse hyperbolic sine transformation.

** p < .05.
*** p < .01.
of baseline imbalance is reassuring and suggestive that, if present, this bias is likely to be small.

C. Negative Life Events, Psychosocial Measures, and Labor Force Measures

The SFC program aimed to improve participants’ psychosocial well-being and resiliency. In addition to their direct benefits, consistent with claims in the SFD community and literature, it was believed that improving psychosocial well-being would impact employment and workforce readiness, leading to higher labor force participation and earnings. To assess this motivation, we collected data on negative life events (to explore resiliency), psychosocial measures, and labor force measures. Without distinct predictions for different individual components of these three groups, we construct aggregate indexes of each of these measures for each individual.

First, we create an index of negative life events. The surveys asked whether respondents or their families had been affected by a set of different types of negative life events over the past year. For example, 27% of the control group reported a serious accident that injured a member of the household, and 28% reported experiences of abuse or a violent crime. Using these responses, we created a life events index by adding together the number of affirmative responses given by the respondent and standardizing this resulting sum. Respondents in the control group and their families had been impacted by an average of 1.5 of these negative life events, with a standard deviation of 2. The standardized index is coded so that higher values indicate that a respondent’s household experienced more negative events.

Second, we collected data on a large number of questions linked to different psychosocial measures and indexes used elsewhere in the literature: subjective welfare, self-esteem, locus of control, aggression, and risky behaviors. Subjective welfare is measured by asking respondents where they see themselves on a six-rung ladder with six as the highest possible response. The mean response in the sample was 2.3. Locus of control is a measure where higher values indicate that a respondent feels that they are more in control of their life outcomes. The self-esteem index measures whether people articulate relatively good or bad feelings about themselves. The aggressive behaviors index captures reported interactions considered to be aggressive, such as disputes with a neighbor or peer. And finally, we collected data to form a risky behaviors index including reported gambling, smoking, alcohol, and drug use. All indexes are standardized and coded so that positive values reflect “better” behaviors. Additional information on the

11 Some life events recorded in the interview were ultimately excluded from the index when we thought that the likelihood of an event could be affected by the program. Table A.9 shows the separate events and indicates with a star which measures were included in the life events index.
subquestions and construction of these indexes is included in appendix D. From these different standardized measures of psychosocial well-being, we then create an aggregate PSI by standardizing the sum of these five components.

Labor outcomes are captured in the survey in the form of reported hours worked and earnings. We create a standardized LFI by standardizing reported hours worked and the inverse hyperbolic sine transformation of earnings (to address skewness in earnings). This sum is then restandardized and used as the LFI in the main analysis.

The program was motivated by a belief that psychosocial well-being and labor force engagement are closely linked. Table 3 shows a significant and positive relationship between three of the five measures of psychosocial behaviors and the LFI. The correlation between the LFI and the aggregate PSI is particularly strong and highly significant. Jointly estimated associations between the psychosocial measures and LFI are included in table A.2 (tables A.1–A.11 are available online) and lead to similar conclusions.

These correlations add credibility to claims of a relationship between psychosocial well-being and labor outcomes but do not constitute a test of causality between the two (in either direction). Importantly, the correlations suggest that in our survey, we are capturing meaningful psychosocial measures that relate to labor market outcomes.

**TABLE 3**

| BASELINE LABOR FORCE INDEX CORRELATIONS WITH PSYCHOSOCIAL MEASURES |
|---------------------------------------------------------------|
| Welfare today       | .068*** (0.021) |
| Self-esteem index  | .109*** (0.020) |
| Locus of control index | .056*** (0.020) |
| Aggressive behaviors index | .031 (0.021) |
| Risky behaviors index | .006 (0.023) |
| Psychosocial index  | .116*** (0.021) |

| N       | 2,310 | 2,310 | 2,310 | 2,310 | 2,310 | 2,310 |
|---------|-------|-------|-------|-------|-------|-------|
| Mean Y  | .01   | .01   | .01   | .01   | .01   | .01   |
| R²      | .061  | .068  | .060  | .058  | .057  | .069  |

**Note.** All indexes are coded so that positive values indicate better behavior. The psychosocial index is the standardized mean of the five (standardized) subindexes. Standard errors are shown in parentheses. The top 1% of earnings and hours are trimmed and set to missing to prevent relationships driven by implausibly large outliers. Regressions include female, age, and age squared covariates as well as fixed effects for educational attainment and community. All variables are coded so that higher values reflect better behaviors or attitudes.

*** p < .01.
D. Program Implementation
Mercy Corps began implementation of the SFC program following completion of a community’s baseline. In total, Mercy Corps established 30 unique sports clubs with 40 members per club. For each club, Mercy Corps recruited two coaches from adults living in the community. All coaches participated in a mandatory 5-day training organized by Mercy Corps. Training curriculum covered facilitation skills, SFC methodology, basic first aid, and the responsibilities of coaches. In addition, Mercy Corps hired four coach mentors to provide ongoing support to coaches throughout the program with continued training, help planning lessons, and assistance problem-solving any challenges that they were experiencing with their teams. Additionally, Mercy Corps conducted audits of sports club meetings to ensure consistency across SFC groups with Mercy Corps’s international standards.

Coaches organized a total of 16 sessions, typically one or two per week. The planned 3-hour meeting comprised 1 hour of introduction and warm-ups, 1 hour of instructional activities, and 1 hour of sports, typically soccer and handball. Table 4 presents the topics covered in each of the 16 SFC sessions along with the targeted skills emphasized in each session. Participants received US$2 for each session that they attended, a sum intended to reimburse participants for transportation expenditures.

Participation in SFC was high in all nine communities; 73% of youth assigned to a SFC group attended at least one session. Figure 1 shows that 65% of all SFC-assigned subjects attended at least 80% of their group’s meetings.

| Session Description                  | Resilience | Planning | Trust | Self-Esteem | Communication |
|-------------------------------------|------------|----------|-------|-------------|---------------|
| 1. Introduction                     | X          |          |       |             |               |
| 2. Me and others                    | X          | X        |       |             |               |
| 3. Understanding emotions           | X          | X        | X     | X           | X             |
| 4. Communication                   | X          |          |       |             |               |
| 5. Relationships                   | X          |          |       |             |               |
| 6. Cooperation                     | X          | X        |       |             |               |
| 7. Believing in me                 | X          | X        |       | X           |               |
| 8. Conflict and violence            | X          | X        | X     | X           |               |
| 9. Collaboration                   | X          |          |       |             |               |
| 10. Motivation                     | X          | X        |       |             |               |
| 11. Dealing with problems          | X          | X        | X     |             |               |
| 12. Making strategies              | X          | X        |       |             |               |
| 13. Applying Sport for Change skills in my life | X          | X        | X     | X           |               |
| 14. Planning graduation event      | X          |          |       |             |               |
| 15. What have we learned?          | X          | X        | X     | X           | X             |
| 16. Review and closing             | X          | X        | X     |             | X             |
On average, the 1,200 individuals assigned to SFC attended 10.35 SFC sessions (out of a maximum of 16). These high attendance rates suggest that the SFC program was desirable in the eyes of participants and add credibility to the view that this program was well implemented. This is corroborated by endline survey responses, where 96% of those who attended at least one SFC session said that they enjoyed the program. Respondents also saw the value of the SFC program applying to a wide range of settings. In response to the question “In what contexts do you think the SFC skills are most useful?” the plurality of respondents, 36%, said when playing sports. However, 21.2% responded that the program was most useful for conflict resolution, while 31.5% said that the program was most useful for finding employment. Additional qualitative interviews revealed that participants saw value in the social and life skills training portions of the SFC program. Further details on these are included in appendix F.

E. Endline Survey
Following completion of implementation of the SFC program in all communities, we attempted follow-up interviews of all baseline respondents assigned to either the control group or one of the SFC teams. The endline survey was intended to be conducted in person with respondents 1 year after completion.
of the intervention. However, because of the risks associated with travel restrictions and quarantines during the Ebola crisis in Liberia, the endline survey was administered through computer-assisted telephone interviews. The endline survey was conducted simultaneously for participants in all nine communities. Endline interviews began on April 3, 2015, and were completed on May 9, 2015.

Despite the challenges of conducting a phone-based survey during the Ebola outbreak, 2,081 individuals were successfully interviewed for the endline survey, a follow-up success rate of 87%. We test for selective attrition in table A.1 and find no evidence of selective attrition in terms of total number of attriters by treatment status (cols. 1, 2). In column 3, we see no overall evidence of selective attrition, and the joint test of significance for baseline covariates interacted with treatment has a p-value of .98. There is one strongly significant baseline characteristic: baseline PSI. Participants in the treatment group with low baseline PSI measures were more likely to attrit from the sample. If anything, this would be likely to positively bias our estimates of SFC’s impacts on PSI.

V. Empirical Results
A. Does SFC Impact Psychosocial and Labor Force Outcomes?
The SFC program had two main objectives. First, Mercy Corps saw the program as a way to improve the psychosocial well-being and resilience of vulnerable youth. Second, Mercy Corps believed that psychosocial improvements would lead to greater workforce preparedness and positively impact labor-related outcomes. Without a strong theoretical foundation for why different dimensions of either psychosocial or labor force measures should be preferred over others and to reduce the number of hypotheses being tested, we focus on the aggregate PSI and LFI detailed in the previous section.

We estimate the direct effects of the program on psychosocial measures and labor outcomes using the following ANCOVA regression specification:

$$Y_{i,t} = \beta_0 + \beta_1 SFC_i + Y_{i,t-1} + \lambda X_i + \delta_i + \epsilon_i,$$

where $Y_{i,t}$ is an outcome of interest for individual $i$ measured at the endline in time $t$, SFC, is an indicator for whether an individual was assigned to the SFC

---

12 Garlick, Orkin, and Quinn (2020) show in a study with microenterprise owners in South Africa that phone-based interviews did not reduce data quality of labor outcomes.

13 Stratification of treatment by community alleviates the concern that inconsistency in time between the program and the follow-up survey bias estimation of the program’s impact. These differences in timing prevent us from making comparisons of treatment effects across communities.

14 We follow Olken (2015) and aggregate multiple measures into aggregate indexes to reduce the number of hypotheses being tested along with the risks associated with testing multiple hypotheses.
program, and \( Y_{i,t-1} \) is an individual’s baseline level of the outcome of interest.\(^{15}\)

The term \( X_i \) is a set of time-invariant covariates that includes age and age squared as well as dummies for female and highest grade level attained. We include a set of community fixed effects, \( \delta_c \), and use robust standard errors to adjust for heteroskedasticity of the error term.

As discussed in section IV, take-up of the SFC program among the treatment group was high, although not universal. We therefore interpret our estimates of \( \beta_1 \) as the average causal effect of the program for those assigned to the treatment group (i.e., the intention to treat estimate). For our main effects, we also present the treatment on the treated estimate by conducting two-stage least squares, with assignment to treatment as an instrument for having ever attended an SFC session. Of those assigned to an SFC group, 73% attended at least one SFC session, whereas no one in the control group ever attended an SFC session. Unsurprisingly, this instrument is highly significant, with a \( T \)-statistic of 53.6 (see table A.3).

Table 5 shows the program’s main impacts. The intention to treat estimates are shown in columns 1 and 2 for PSI and LFI, respectively. For PSI, we see a small negative point estimate with a 95% confidence interval that includes effects between \(-0.098\) and \(0.07\) standard deviations. However, the LFI impacts are positive and statistically significant. Column 2 shows a point estimate of \(0.115\) standard deviations (\(p = .011\)) with a confidence interval from \(0.025\) to \(0.205\). Testing for two outcomes, we perform a sharpened false discovery rate adjustment following Anderson (2008). Adjusted \(q\)-values are reported in brackets beneath the standard errors. The effects of SFC on the LFI in column 2 retain statistical significance with a \(q\)-value of \(0.024\).

Effects of SFC on the psychosocial subindexes are shown in table A.4. Consistent with the effects on the aggregate measure, they are not encouraging. The estimated treatment effects for the different psychosocial subindexes have different signs, and the only outcome showing marginal significance, risky behaviors, has a negative point estimate.\(^{16}\) By contrast, table A.5 shows effects of SFC on the two subcomponents of the LFI, suggesting 11%–12% increases in hours of labor supply and weekly earnings that both survive multiple hypothesis testing (\(q\)-value = \(0.072\)).\(^{17}\)

---

\(^{15}\) With random assignment, two survey periods, and low autocorrelation of our outcome variables, we follow McKenzie (2012) and use ANCOVA as our preferred specification for greatest statistical precision.

\(^{16}\) Table A.8 shows treatment effects on all subcomponents of the PSI subindexes. Unsurprisingly, none of them retain statistical significance after correcting for multiple hypothesis testing within each subindex.

\(^{17}\) Table A.7 shows robustness of the main labor force effects to inclusion of noisier 90-day income recall. Because 50% of randomly selected respondents in the SFC treatment group were also invited to participate in the CFW program, we test whether effects attributed to SFC could have been driven
Columns 3 and 4 of table 5 show the treatment on the treated effect by using random assignment to an SFC group as an instrument for having ever attended an SFC session in a two-stage least squares estimation. As expected given partial uptake of the treatment, magnitudes of both estimates increase. For LFI, the treatment on the treated estimation suggests an average increase of 0.161 standard deviations for those who attended any SFC sessions, with a 95% confidence interval between 0.039 and 0.283.

by CFW instead. Table A.6 shows that point estimates of effects for LFI are larger for those who were not in the CFW treatment group.
B. Does SFC Improve Resilience?

The SFC program was also motivated as a way to improve participants’ resilience to negative events. To explore this, we look at heterogeneous program effects in the presence of negative life events.

Our specification for measuring heterogeneous treatment effects is the following:

\[ Y_{i,t} = \beta_0 + \beta_1 \text{SFC}_i + \beta_2 (\text{SFC}_i \times \text{Het}_i) + \beta_3 \text{Het}_i + Y_{i,t-1} + \lambda\mathbf{x}_i + \delta_i + \epsilon_i, \]

where \text{Het}_i is a measure of individual-level heterogeneity, such as exposure to negative life events. The primary coefficient of interest is the estimate of \( \beta_2 \), which tests whether program responses differ by this dimension of heterogeneity.\(^1\)

In columns 5 and 6 of table 5, we test whether the program has additional benefits for those who recently experienced negative life events. Because of the Ebola outbreak, all study participants undoubtedly experienced a meaningful disruption to their daily lives, while many likely experienced considerable adversity.\(^2\) Thus, the main results on PSI may already suggest limited impacts on resiliency. The heterogeneity analysis allows us to further explore resiliency in the presence of additional household trauma. First, we note a strong negative correlation between the life events index and psychosocial outcomes. This further increases the credibility of these measures and raises our confidence that our psychosocial measures are not mere noise. However, the coefficient of interest in this specification is the interaction term between SFC and negative life events. For PSI, greater resilience from the program would imply a positive coefficient on the interaction. While positive, our estimate for this term is close to zero, with a 95% confidence interval from \(-0.074\) to \(0.098\). We do not, therefore, find evidence of improved psychosocial resilience in the presence of negative life events resulting from the program.

Column 6 of table 5 again looks at program heterogeneity in the presence of negative life events for the LFI. However, these predictions are less clear than they were for psychosocial outcomes. Greater resiliency in the presence of negative life events may be reflected in increased labor force participation to cope with these shocks, or it may conversely be reflected in a weaker response if the

\(^{1}\) This approach could be problematic if the SFC program directly affected probability of participants experiencing negative life events. Table A.9 looks at the incidence of negative life events in the sample, split by treatment status, and does not find significant differences.

\(^{2}\) Because of the sensitivity of the issue at the time of the follow-up survey, we did not ask directly about respondents or family members being impacted by Ebola. However, some effects of Ebola were likely captured in the life events index in the form of sick or deceased family members.
program increased resiliency by reducing respondents’ vulnerability to shocks. The regression results in column 6 suggest that negative life events are associated with increased labor force participation. However, the interaction term is negative, thus muting this response, with a confidence interval that contains zero. Without a clear prediction and without significant effects, we do not consider this to be evidence either for or against greater labor force resiliency.

SFC was motivated by a belief that the program would improve participants’ psychosocial well-being and that this would, in turn, improve participants’ ability to participate in the labor force. While we do find positive effects on the latter outcome, we do not see evidence of positive impacts on psychosocial measures or improved resiliency. These results suggest that the effects of SFC on labor force outcomes may not be conditional on improvements to psychosocial well-being.

C. Are Treatment Effects Concentrated among Certain Subgroups?

In this section, we present a secondary analysis to better understand who benefited most from the program and whether impacts are reflected in informative and sensible patterns of heterogeneity. To explore this, we check for heterogeneous treatment effects across a number of different dimensions: gender, age, education, and previous vocational training. Each of these dimensions can be sensibly divided into a group that is economically advantaged (male, older, better educated, or with training) or disadvantaged. We additionally use our core set of covariates to predict labor market outcomes to see whether those with high or low predicted labor force outcomes (above or below the median) respond differently to the treatment.20

Figure 2 plots the estimated LFI treatment effects from separately estimated ANCOVA regressions for different dimensions of heterogeneity. For each pair of estimates, we find that impacts of SFC on LFI are larger for the more disadvantaged group. We also utilize equation (2), where Het, is a dimension that we test for heterogeneous treatment effects. These results are shown in regression form in panel A of table 6. Each row shows coefficients from a separately estimated version of equation (2) with the main effect of SFC participation, an interaction term, and the dimension of heterogeneity (listed in col. 1). The first row of panel A shows a positive point estimate of the SFC program of 0.064 standard deviations, with standard errors of 0.066. Column 2 shows an interaction term of 0.098 that is not significantly different from zero. However, the

20 Because of low autocorrelation in our main outcome measures (0.107 for LFI and 0.202 for PSI), we use data from the follow-up survey of the control group to generate predicted endline values in the full sample.
overall effect for females (SFC + SFC × Het in the table) is 0.162 standard deviations, with a \( p \)-value of .01 reported in column 5. Although the interaction terms testing the difference between these groups are not significant, we observe that all five of the disadvantaged groups experience an overall positive treatment effect with significance levels at or below 5%.\(^\text{21}\)

The final pair of results in figure 2 and fifth row of panel A in table 6 show that those with worse predicted labor force outcomes have significantly larger treatment effects than those expected to be doing better in the absence of the program. People predicted to have low labor force outcomes have a treatment effect that is 0.21 standard deviations bigger than those predicted to have higher labor force outcomes in the absence of the program (\( p = .02 \)). Given a lack of clear theoretical predictions motivating who we should have expected to have greater or lesser responses to the program, we consider this analysis to be exploratory. However, consistent patterns across multiple dimensions of heterogeneity

\(^\text{21}\) As this secondary analysis is exploratory in nature, we do not correct these \( p \)-values for multiple comparisons (following Olken 2015).
are noteworthy and may provide a starting point for future research in a limited literature.

We perform a similar set of analyses to look for parallel patterns of program impact on psychosocial outcomes in table 6, panel B. We do not find meaningful positive treatment effects on PSI for any of our subgroups across these dimensions of heterogeneity. Outcomes for those with high predicted PSI may, in fact, be negative, while those with worse predicted outcomes have a positive point estimate that is indistinguishable from zero. We also see slightly
worse PSI outcomes (although not statistically significant) for those with worse predicted labor force outcomes.

Overall, we take these patterns as additional evidence that the SFC program did not induce positive changes in psychosocial outcomes among easily identifiable subgroups. That we do not see effects on PSI for those groups driving impacts on labor force outcomes further undermines our confidence that labor force benefits came through a psychosocial channel. Ultimately, the mechanisms behind the positive labor force impacts may be more nuanced than a simplified theory of change based on these intermediate indicators of psychosocial well-being.

D. Does Group Composition Affect Program Impacts?
A different possibility is that group composition and peer effects play a central role in program effectiveness. After the results of our evaluation were known, Mercy Corps expressed particular concern that random assignment to youth groups may have disrupted the efficacy of the program on psychosocial impacts. Life skills lessons covered many sensitive topics and required trust among group members sharing personal experiences with the other members. Randomization of registrants into sports groups may have made groups less cohesive, with members sharing less similar backgrounds or preexisting relationships than if they had been permitted to choose their own groups.

We explore whether program impacts differed depending on different measures of group cohesion in table 7. Baseline data collection included questions about each respondent’s social network, allowing us to identify preexisting social linkages among study participants. Panel A shows that presence of a friend in one’s sports group significantly increased the likelihood of ever attending one of the sessions as well as total attendance by 9%–10%. However, we do not find evidence that presence of a friend improves either psychosocial or labor force outcomes. Panel B examines whether individual outcomes vary by group ethnic diversity. Group diversity is measured by calculating a Herfindahl index, $H_g = \sum_{i=1}^{N} E_i^2$, where $E_i$ is the share of sports group $g$’s members that belong to a particular ethnic group, $i$. In panels C and D, we test for heterogeneity by whether a respondent has more (or less) age or age-gender mates in their randomly assigned sports group. While the presence of a larger number of teammates

---

22 In our regressions, we standardize the Herfindahl scores using the following procedure. First, we simulate 1,000 draws of each community’s sports groups and calculate a mean Herfindahl index, $\bar{H}_c$, for each community, $c$. Then we compute the deviation from the simulated index for each sport group and divide by the standard deviation of our simulated Herfindahl index, $(H_g - \bar{H}_c)/SD(\bar{H}_c)$. 
### TABLE 7
**EFFECTS OF GROUP COHESION**

| Ever Attend | Total Attendance | Psychosocial Index | Labor Force Index |
|-------------|------------------|--------------------|-------------------|
| (1)         | (2)              | (3)                | (4)               |

#### A. Friends

|                      | Attendance | Psychosocial Index | Labor Force Index |
|----------------------|------------|--------------------|-------------------|
| Sport for Change     | .705***    | -0.025             | .123**            |
|                      | (.020)     | (.020)             | (.046)            |
| Sport for Change × any friends | .079*** | .008               | -.015             |
|                      | (.028)     | (.067)             | (.068)            |
| Mean any friends     | .365       | .365               | .365              |
|                      | (.020)     | (.283)             | (.040)            |
|                      | .598       | .590               | .076              |
|                      | (.040)     | (.046)             | (.086)            |

#### B. Group Diversity

|                      | Attendance | Psychosocial Index | Labor Force Index |
|----------------------|------------|--------------------|-------------------|
| Sport for Change     | .738***    | -.006              | .100***           |
|                      | (.019)     | (.029)             | (.036)            |
| Sport for Change × Herfindahl SD | .004 | .015               | -.029             |
|                      | (.008)     | (.137)             | (.019)            |
|                      | (.028)     | (.421)             | (.068)            |
|                      | .5900      | .5818              | .0776             |
|                      | (.019)     | (.368)             | (.068)            |

#### C. Similar Age

|                      | Attendance | Psychosocial Index | Labor Force Index |
|----------------------|------------|--------------------|-------------------|
| Sport for Change     | .655***    | -.119              | .377**            |
|                      | (.073)     | (.165)             | (.140)            |
| Number of group members with similar age | .004 | .005              | -.013*           |
|                      | (.003)     | (.053)             | (.007)            |
| Mean group members with similar age | 19.150 | 19.150            | 19.150             |
|                      | .5919      | .5837              | .0753             |
|                      | (.013)     | (.088)             | (.011)            |

#### D. Similar Age and Gender

|                      | Attendance | Psychosocial Index | Labor Force Index |
|----------------------|------------|--------------------|-------------------|
| Sport for Change     | .644***    | -.130              | .379**            |
|                      | (.072)     | (.169)             | (.141)            |
| Number of group members with similar age and gender | -.005 | .004              | -.020*           |
|                      | (.006)     | (.011)             | (.011)            |
|                      | .018**     | .303**             | .002              |
|                      | (.008)     | (.112)             | (.015)            |
| Mean group members with similar age and gender | 19.150 | 19.150            | 19.150             |
|                      | .5992      | .5910              | .0809             |
|                      | (.003)     | (.088)             | (.016)            |

#### E. Summary Statistics

|                      |           |                   |                   |
|----------------------|-----------|-------------------|-------------------|
| N                    | 2,058     | 2,058             | 2,058             |
| Treatment mean       | .3678     | 5.4111            | -.0180            |
| Control mean         | .0000     | .0000             | -.0000            |
|                      | 1,947     | 1,947             | 1,947             |

**Note.** Results from ANCOVA estimation. “Herfindahl SD” reflects standard deviations of the actual group Herfindahl index from simulated community-level means values. “Similar age” is defined as the number of other group members who are similarly either above or below the median age in the sample. “Similar age and gender” restricts this measure further by also requiring that the group member is of the same gender as well as age group. Standard errors are shown in parentheses.

* * p < .10.
** ** p < .05.
*** *** p < .01.
of similar age and gender increases program attendance (panel D), their presence does not appear to improve program impacts on psychosocial or labor force measures.

We adopt two additional approaches to examine whether peer influence is an important channel for program effects. First, we test whether exposure to “better” peers could have positive impacts on participants. In particular, we test whether groups with greater average baseline levels of either psychosocial measures or labor force measures have differential outcomes. We calculate a leave-out mean baseline value for each individual’s assigned sports group (leaving out the individual’s own baseline value in this calculation), subtracting community-level averages from these group measures. Respondents in the control group are assigned a value of 0 and are preserved in this analysis as a reference point to determine whether peer effects could be driving the overall effects of the program. With treatment randomly assigned and group membership randomly assigned conditional on community, we take these group composition variables as exogenous.

We find little evidence that peer influence is driving the average treatment effect of the program on LFI. Table 8 indicates that controlling for the leave-out mean of group PSI or LFI (normalized to be centered on zero) leaves the coefficient on LFI virtually unaffected and similarly precise (0.118, compared with a coefficient of 0.115 in table 5, col. 2). There is, however, evidence that peer composition may explain some heterogeneity about this mean treatment effect. In particular, panel A, column 4 of table 8 indicates that an additional 0.1 standard deviation average of baseline PSI is associated with a 0.046 standard deviation improvement in LFI, an effect nearly 40% as large as the main effect of the program. At the same time, we do not find evidence that better groups in terms of baseline LFI are associated with larger treatment effects. We also do not find evidence that having a group with a larger baseline PSI or LFI is associated with improved PSI outcomes. Absent a theory of why better PSI groups may matter (only) for LFI outcomes when better LFI groups do not, we leave this analysis as a suggestive area for further research.23 Ultimately,

---

23 We also note that these peer effect estimates do not appear to explain the heterogeneity of the main effects from sec. V, where those with low predicted labor force outcomes benefit more from the program. In table A.10, we see that peer effects are driven by those who are predicted to have higher LFI outcomes. While leaving the channel of the program’s direct effects on labor force outcomes unexplained, this adds further credence that the effects are coming through the program and not peer effects. We additionally performed an agnostic test of peer effects following Shue (2013) with two-way clustering for dyadic data following Fafchamps and Gubert (2007). Our test reveals that group members are, if anything, more dissimilar in their outcomes following intervention than they are to other study participants. This test provides further support that positive peer effects are not an important driver of the evaluation results. These results are shown in table A.11
while we find some suggestive evidence of peer effects influencing labor force outcomes, these analyses do not provide clear insights into the mechanisms of the program's effects on LFI.

VI. Summary and Conclusion

Using sports as a method of intervention and vehicle for socioemotional and psychosocial training has come increasingly into fashion. SFD is viewed as a potentially transformative approach to engaging and positively affecting the lives of vulnerable youth. These programs involve millions of participants across the globe and constitute hundreds of millions of dollars of expenditures each year. Despite these high levels of participation and expenditure, there is little existing evidence for the efficacy of these programs and their effect on participants.
Our evaluation focused on the SFD programming developed and implemented internationally by Mercy Corps, one of the global leaders in this space. In this context, we found that their SFC program exerted limited impacts on psychosocial outcomes but did increase labor force engagement a year after the intervention by a statistically significant 0.12 standard deviation. We note that it is plausible that similar programs may generate different effects on these outcomes if they use a different method of selecting and training coaches. For example, professional counselors, trained therapists, or job placement experts may be expected to result in different impacts on psychosocial and labor force outcomes if they serve as coaches, a possibility that we cannot speak to in this paper.

Given this lack of effects on psychosocial measures, the motivating theory of change for SFD does not appear to have been the mechanism driving improved labor market outcomes in this setting. While we were ultimately unable to isolate these mechanisms, heterogeneity analysis suggests that more disadvantaged groups (women, less educated, young, without vocational training) benefited most from the program.

Ultimately, this evaluation provides evidence of positive impacts of an SFD program on labor force outcomes. Given the scarcity of positive findings on active labor market programs in developing countries in general and the extent to which SFD programs are not precisely targeted at boosting labor market outcomes, these results are notable in their precision and magnitude (McKenzie 2017). However, the strength of these results is tempered by our inability to identify the mechanism through which this program’s impact works. Given the pervasiveness and scale of resources devoted to SFD programs, we feel that further research should be done to deepen the pool of evidence on SFD programs.

References
Adair, Daryl. 2014. *Global Sport-for-Development: Critical Perspectives.* London: Palgrave Macmillan.
Adoho, Franck, Shubha Chakravarty, Dala T. Korkoyah, Mattias K. A. Lundberg, and Afia Tasneem. 2014. “The Impact of an Adolescent Girls Employment Program: The EPAG Project in Liberia.” World Bank Policy Research Working Paper no. 6832, World Bank, Washington, DC.
Anderson, Michael L. 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association* 103, no. 484:1481–95.
Bendavid, Eran, Trygve Ottersen, Liu Peilong, Rachel Nugent, Nancy Padian, John-Arne Rottingen, and Marco Schaferhoff. 2017. “Development Assistance for Health.” In *Disease Control Priorities: Improving Health and Reducing Poverty*, 3rd ed., ed.
D.T. Jamison, H. Gelband, and S. Horton. 2006. “Reduction Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia.” *American Economic Review* 107, no. 4:1165–206.

Burnett, Cora. 2009. “Engaging Sport-for-Development for Social Impact in the South African Context.” *Sport in Society: Cultures, Commerce, Media, Politics* 12, no. 9:1192–205.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. “The Effects of Youth Training on Risk Behavior: The Role of Non-Cognitive Skills.” *IZA Journal of Labor and Development* 5, no. 1:12.

Coalter, Fred. 2009. “Sport-in-Development: Accountability or Development?” In *Sport and International Development*, ed. Roger Levermore and Aaron Beacom, 55–75. London: Palgrave Macmillan.

———. 2010. “The Politics of Sport-for-Development: Limited Focus Programmes and Broad Gauge Problems?” *International Review for the Sociology of Sport* 45, no. 3:295–314.

Ditlmann, Ruth K., and Cyrus Samii. 2016. “Can Intergroup Contact Affect Ingroup Dynamics? Insights from a Field Study with Jewish and Arab-Palestinian Youth in Israel.” *Peace and Conflict: Journal of Peace Psychology* 22, no. 4:380–92.

Donnelly, Peter, Simon Darnell, S. Wells, and J. Coakley. 2007. “The Use of Sport to Foster Child and Youth Development and Education.” In *Harnessing the Power of Sport for Development and Peace*, ed. Sport for Development and Peace International Working Group. Toronto: Right to Play.

Dunmore, Charlie. 2013. “Liberian President Says Youth Unemployment a Threat to Peace.” *Reuters*, November 25, 2013.

Fafchamps, Marcel, and Flore Gubert. 2007. “The Formation of Risk Sharing Networks.” *Journal of Development Economics*, 83, no. 2:326–50.

Garlick, Robert, Kate Orkin, and Simon Quinn. 2020. “Call Me Maybe: Experimental Evidence on Using Mobile Phones to Survey Microenterprises.” *World Bank Economic Review* 34, no. 2:418–43.

Groh, Matthew, Nandini Krishnan, David McKenzie, and Tara Vishwanath. 2016. “The Impact of Soft Skills Training on Female Youth Employment: Evidence from a Randomized Experiment in Jordan.” *IZA Journal of Labor and Development* 5, no. 1:9.

Hartmann, Douglas, and Brooks Depro. 2006. “Rethinking Sports-Based Community Crime Prevention: A Preliminary Analysis of the Relationship between Midnight Basketball and Urban Crime Rates.” *Journal of Sport and Social Issues* 30, no. 2: 180–96.

Hartmann, Douglas, and Christina Kwauk. 2011. “Sport and Development: An Overview, Critique, and Reconstruction.” *Journal of Sport and Social Issues* 35, no. 3:284–305.
Heckman, James J., and Tim Kautz. 2012. “Hard Evidence on Soft Skills.” *Labour Economics* 19, no. 4:451–64.

Henley, Robert, Ivo Schweizer, Francesco de Gara, and Stefan Vetter. 2007. “How Psychosocial Sport and Play Programs Help Youth Manage Adversity: A Review of What We Know and What We Should Research.” *International Journal of Psychosocial Rehabilitation* 12, no. 1:51–58.

Holt, Nicholas L., and Martin I. Jones. 2007. “10 Future Directions for Positive Youth Development and Sport Research.” In *Positive Youth Development through Sport*, ed. Nicholas L. Holt, 122–32. London: Routledge.

Ibarraran, Pablo, Laura Ripani, Bibiana Taboada, Juan Miguel Villa, and Brigida Garcia. 2014. “Life Skills, Employability and Training for Disadvantaged Youth: Evidence from a Randomized Evaluation Design.” *IZA Journal of Labor and Development* 3, no. 1:10.

Jeanes, Ruth. 2013. “Educating through Sport? Examining HIV/AIDS Education and Sport-for-Development through the Perspectives of Zambian Young People.” *Sport, Education and Society* 18, no. 3:388–406.

Kautz, Tim, James J. Heckman, Ron Diris, Bas terWeel, and Lex Borghans. 2014. “Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success.” NBER Working Paper no. 20749, National Bureau of Economic Research, Cambridge, MA.

Kidd, Bruce. 2008. “A New Social Movement: Sport for Development and Peace.” *Sport in Society* 11, no. 4:370–80.

Lowe, Matt. 2021. “Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration.” *American Economic Review* 111, no. 6:1807–44.

McKenzie, David. 2012. “Beyond Baseline and Follow-Up: The Case for More T in Experiments.” *Journal of Development Economics* 99, no. 2:210–21.

———. 2017. “How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence.” *World Bank Research Observer* 32, no. 2:127–54.

Morris, Leesa, Jo Sallybanks, Katie Willis, and Toni Makkai. 2004. “Sport, Physical Activity and Antisocial Behaviour in Youth.” *Youth Studies* 4, no. 1:47–52.

Mousa, Salma. 2020. “Building Social Cohesion between Christians and Muslims through Soccer in Post-ISIS Iraq.” *Science* 369, no. 6505:866–70. https://doi.org/10.1126/science.abb3153.

OECD (Organization for Economic Cooperation and Development). 2019. *Development Aid at a Glance*. Paris: Organization for Economic Cooperation and Development. https://www.oecd.org/dac/financing-sustainable-development/development -finance-data/World-Development-Aid-at-a-Glance-2019.pdf.

Olken, Benjamin A. 2015. “Promises and Perils of Pre-Analysis Plans.” *Journal of Economic Perspectives* 29, no. 3:61–80.

Panter-Brick, Catherine, Rana Dajani, Mark Eggerman, Sabrina Hermosilla, Amelia Sancilio, and Alastair Ager. 2018. “Insecurity, Distress and Mental Health: Experimental and Randomized Controlled Trials of a Psychosocial Intervention for Youth Affected by the Syrian Crisis.” *Journal of Child Psychology and Psychiatry* 59, no. 5:523–41.
Perkins, Daniel F., and Gil G. Noam. 2007. “Characteristics of Sports-Based Youth Development Programs.” *New Directions for Student Leadership* 2007, no. 115: 75–84.

Petitpas, Albert J., Allen E. Cornelius, Judy L. Van Raalte, and Tiffany Jones. 2005. “A Framework for Planning Youth Sport Programs That Foster Psychosocial Development.” *Sport Psychologist* 19, no. 1:63–80.

Rookwood, Joel. 2008. “Soccer for Peace and Social Development.” *Peace Review* 20, no. 4:471–79.

Shue, Kelly. 2013. “Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers”. *Review of Financial Studies* 26, no. 6:1401–42.

Svensson, P. G., and H. Woods. 2017. “A Systematic Overview of Sport for Development and Peace Organisations.” *Journal of Sport for Development* 5, no. 9:36–48.

Underwood, Emily. 2018. “Lessons in Resilience.” *Science* 359, no. 6379:976–79.