Received: 3 March 2022 | Revised: 24 February 2023 | Accepted: 1 March 2023

DOI: 10.1111/puar.13631

RESEARCH ARTICLE

PAR PUBLIC ADMINISTRATION REVIEW ASPA

The impact of social impact bond financing

Debra Hevenstone[1] | **Alec Fraser**[2] | **Lukas Hobi**[1] | **Wojtek Przepiorka**[3] | **Gemma G. M. Geuke**[4]

[1]Social Work, Bern University of Applied Sciences, Bern, Switzerland

[2]King's Business School, King's College London, London, UK

[3]Social and Behavioral Sciences, Utrecht University, Utrecht, The Netherlands

[4]Department of Sociology, Utrecht University, Utrecht, The Netherlands

Correspondence
Debra Hevenstone, Social Work, Bern University of Applied Sciences, Hallerstrasse 10, 3012 Bern, Switzerland.
Email: debra.hevenstone@bfh.ch

Present address
Gemma G. M. Geuke, Department of Tobacco Control, Trimbos Institute, Utrecht, The Netherlands.

Abstract

Social impact bonds (SIBs), also known as Pay for Success, are an innovation in Payment by Results contracting. Investors finance programs and are repaid based on the "SIB effect," which includes changes in outcomes attributable to financing. We generate a quantitative estimate of this part of the SIB effect for two active labor market programs in the Netherlands and Switzerland. Comparing program impacts within providers using SIB and non-SIB contracts suggests financing has positive impacts on public benefit receipt, employment, and income. Qualitative research suggests this is because SIB contracts increased pressure for all involved parties, leading to the institutionalization of selection and greater resources for SIB-financed services. Contracts with high pressure, like SIBs, may compromise both performance requirements and the potential to measure performance. We examine the implications of these findings in relation to agency and stewardship theories and highlight the significance of SIBs as multilateral as opposed to bilateral contracts.

Evidence for practice

- Social Impact Bond (SIB) financing for active labor market programs in the Netherlands and Switzerland improved employment, earnings, and benefit receipts, through the institutionalization of selection processes and greater resources for SIB-financed services compared with non-SIB-financed services.
- Greater emphasis on outcomes in government contracting, as in SIBs, may weaken the effectiveness of contract incentives if the cost of failure feels too high.
- The outcomes-focus of contracts like SIBs increases scrutiny but may inadvertently undermine the accurate measurement of program and financing impacts.

INTRODUCTION

Contracting is challenging in public administration and social policy, where program goals are long-term and complex (Larsen & Wright, 2014; Jahn & Ochel, 2007; Van Slyke, 2007; Brunjes, 2020). This was true in earlier government contracting (Boyne, 1998), with accountability and incentives still flawed following Payment by Results (PbR) innovations (Romzek & Johnston, 2005). The shortcomings of PbR are partly due to the complexity of writing a contract accurately incorporating all relevant incentives (Koning & Heinrich, 2013).

Social Impact Bonds (SIBs), known as Pay for Success (PFS) initiatives in the US, aimed to improve contracting. Under a SIB, investors front money for a program, a provider delivers services, an auditor or evaluator assesses it, and the end payer (often government) repays investors, with terms contingent on achieved outcomes or impacts (Liebman, 2011). SIBs were first established in the UK in 2010, with a current total of 225 SIB-financed projects worldwide (GO Lab, 2020). The most common type of SIB focuses on employment and training, likely because employment and benefit outcomes are easily measured, as are related state savings, and because of the strong existing evidence showing how these programs impact employment and benefit receipt (Card, Kluve, & Weber, 2018).

SIBs need contracts detailing payment terms to providers and investors with investor payment partly

depending on the "SIB effect." The literature has used diverse definitions of the SIB effect, ranging from broad definitions of how financing impacts the practices of those involved (Lowe, Kimmitt, Wilson, Martin, & Gibbon, 2019) to definitions including high social returns due to the redirection of public resources toward under-resourced areas of social policy (Liebman, 2011; Liebman & Feller, 2014). Here we use the definition emphasizing improvements in outcomes due to financing (Edmiston & Nicholls, 2018; Giacomantonio, 2017). Using this definition of the SIB effect requires distinguishing the causal impact of financing from the "program effect," the causal impact of the intervention.

The literature has suggested several mechanisms to motivate a SIB effect, like greater effectiveness through better provider contracts (Dayson, Fraser, & Lowe, 2020; FitzGerald, Rosenbach, Hameed, Dixon, & Blundell, 2021; Heinrich & Kabourek, 2019; Maier, Barbetta, & Godina, 2018) and inter-organizational collaboration (Leventhal, 2012) across public, private, and philanthropic partners (Fox & Albertson, 2011, 2012; Joy & Shields, 2013). Despite the relevance of this SIB effect, there has been no quantitative estimate to date (Edmiston & Nicholls, 2018; Tan, Fraser, McHugh, & Warner, 2021); only program effects have been estimated (Anders & Dorsett, 2017). With this lack of knowledge, the literature consistently calls for empirical evidence on the SIB effect (Chamaki, Jenkins, & Hashemi, 2019; Fox & Albertson, 2011).

In this paper, we answer three questions:

1. Does SIB financing improve the impact of active labor market programs on benefit receipt, employment, and earnings?
2. What are the underlying conditions that lead to any impacts?
3. What do the findings imply regarding the feasibility of measuring program and contract impacts?

We estimate the SIB effect for two Active Labor Market Programs (ALMPs) in the Netherlands and Switzerland, comparing program effects at providers offering the same services to the same target groups in two periods: a pre-SIB period and a period with SIB financing. The SIB effect is estimated by taking the difference between two program effects. These program effects are estimated using difference in differences method, comparing participants in the pre- and SIB periods with two comparable groups of eligible non-participants in the same periods, taken from administrative data. We approximate program impacts using difference in differences methods and longitudinal administrative data, comparing participants in the pre- and SIB periods with two comparable groups of eligible non-participants in the same periods.

Our quantitative analyses find that SIBs reduce public benefit receipt and increase earnings and employment. These analyses are complemented by qualitative data suggesting the results were due to collective pressure for SIB programs to be seen as "successful," which led to institutionalized selection processes and greater funding for the SIB-financed services compared with non-SIB-financed services. SIB financing thus generated two unanticipated consequences: (1) De facto weaker contract incentives and (2) The undermining of impact measurement. Further quantitative analyses motivated by the qualitative findings, using limited additional data from language testing, suggest that estimates using standard approaches may miss selection procedures on the ground, mistakenly suggesting a positive SIB effect. Findings suggest the presence of an investor within a multilateral contracting relationship with government and providers increases the pressure on all actors to ensure that SIB-financed programs "succeed", promoting a model, at least in the short-term, that may be closer to the principles of stewardship theory, fostering collaborative relationships, than agency theory, stressing performance control mechanisms.

THEORY

Contracting

Outsourcing is a principle-agent problem (Van Slyke, 2007; Williamson, 1985). The principle (government) cannot observe the agent's (contractor's) actions. Ideally, the contract aligns the contractor's incentives to the government's. This is challenging in the social sector, where tasks are complex and time horizons are long. In designing a contract, the government makes many decisions, such as: (1) Whether to pay by clients served, services delivered, or results; (2) Which measures to use and, if paying by performance, what estimation strategy to use; (3) Whether incentives should use discrete or continuous targets; (4) How to weight payments by measure; and (5) Which time frame to use.

Each decision has its pitfalls. Payments based on process or raw outcomes often lead to "creaming" and "parking", perverse incentives to serve those who need the least help most and to enroll the neediest while offering minimal assistance. Incentivizing certain goals can inadvertently disincentivize achievement of other goals—e.g., including job placement can reduce the quality of job placements (Lu, 2016). Further, estimation strategies are often imperfect, leading to rewards that are uncorrelated with true impact (Barnow, 2000; Heinrich, 2007) or generating similar perverse incentives as using payment for services or raw outcomes (Courty, Kim, & Marschke, 2011). In addition, contract timelines are often too short since providers need revenue to operate, leading to payments that are unrelated or even negatively correlated with long-run impacts (Heckman, Heinrich, & Smith, 2002; Koning & Heinrich, 2013). Finally, discrete targets make for simpler contracts but can induce providers to reduce effort when target achievement (or non-achievement) is certain for a client (Dixit, 2002).

Although writing a good contract is difficult, in particular for the social services (Blank, 2000), the government has alternative ways to ensure quality, like field audits, customer surveys, and monitoring complaints (Brown & Potoski, 2003), as well as relationships, counseling, and informal conflict resolution (Girth, 2017). Further, there is evidence that front-line workers serve the public good in opposition to perverse contract incentives (Burgess & Ratto, 2003; Heinrich & Marschke, 2010; Van Slyke, 2007), motivated by altruism and professional values (Dixit, 2002). The extent to which perverse incentives filter down depends partly on management practices (Pihl-Thingvad, 2016) with existing relationships among contracting parties, and experience playing important roles (Brunjes, 2020, 2022).

The desire to properly incentivize a contract is linked to New Public Management (NPM) reform ideas arguing that public sector goals can be formulated as measurable objectives that can be used to motivate a competitive market between providers through outsourcing (Dunleavy, Margetts, Bastow, & Tinkler, 2006). The literature on contracting and NPM has parallels with public administration debates on agency and stewardship theories (Van Slyke, 2007). Agency theory emphasizes goal divergence between the government and the contractor, while stewardship theory emphasizes the intrinsic motivation of all parties to achieve the same goals. Agency theory thus aligns more with incentives and monitoring to align principal–agent relationships as in PbR contracts and NPM theory, while stewardship theory aligns more with altruism, relationships, and trust, as in pay-for-service contracts or post-NPM models such as New Public Governance (NPG) (Osborne, 2010).

Strict typologies fail to acknowledge that in practice, contracts may draw on different approaches at different times and that the characteristics of contracting parties and their context play an important role. Further, contractual relationships evolve contingent on context and organizational characteristics. For example, a long-term relationship may first be defined by the government's principle-agent approach but move toward a model more focused on provider organizations' missions, characterized more by trust or stewardship (Van Slyke, 2007).

How social impact bonds change contracting

It is unclear how SIBs impact the problems of PbR incentives, the balance between stewardship and agency models, and how changes may unfold during contracts. Note that conventional (non-SIB) contracts are *bilateral*—between government purchasers and service providers. A key feature of a SIB is that the contract is *multilateral*, introducing investors (Del Giudice and Migliavacca, 2019; Fraser, Tan, Lagarde, & Mays, 2018; Dixon, 2020). These potentially powerful new actors can impact the relationships between contracting parties, generating new political challenges, with implications for contract performance and public administration theory.

Although all SIBs have an investor, there is substantial diversity among contracts labeled as "SIBs" (Maier et al., 2018). Contracts vary in provider contract model, payment criteria, and associated stakes for the provider and investor (Fraser et al., 2018). Combining the unknown impact of investors with this diversity within the SIB label, has led to conflicting interpretations of the theoretical implications of SIBs.

Much of the current discussion on how to theorize SIBs is embedded in a debate referencing New Public Management (Dayson et al., 2020; French, 2019; Heinrich & Kabourek, 2019; Warner, 2013). Some see SIBs as intensifying incentives thus extending NPM principles and exacerbating the problems associated with PbR (Baliga, 2013; Warner, 2013); Others see SIBs as increasing inter-organizational collaboration and aligning or buffering incentives (Joy & Shields, 2013; Selviaridis and Wynstra, 2015), more characteristic of the NPG than the NPM. Translating these arguments into the language of public administration—there are conflicting arguments that SIBs move contracting in either the direction of agency or stewardship theory. In the next subsection, we examine the arguments about how a shift toward an agency or stewardship model may generate a SIB effect.

The "SIB effect"

The SIB effect, defined as the independent impact of financing, is a key element in defining the conditions of the investor contract. Given mounting evidence that SIBs inflate costs and that providers already offer similar programs (Sacristán López de los Mozos, Duarte, & Ruiz, 2016), it is critical to understand if there is a SIB effect and, if so, why. We state two broad (and contradictory) hypotheses that may lead to a positive SIB effect.

First, SIBs could *intensify* incentives as investors increase pressure on providers through the levers they use to steer provider practices (Carter, 2021). This would create a "firmer" NPM, or agency model with more formal performance management, strict contractual adherence, and financial penalties for underperformance (Dayson et al., 2020; Fraser et al., 2018; Warner, 2013). This could dilute the social intent of an intervention but it increase overall effectiveness (Carter, 2021). More high-pressure performance regimes may encourage providers to "professionalize" their practices through improved data and measurement through the inculcation of "market discipline" (Cohen, 2011). Alternatively, given the investors' role in establishing SIBs, the opposite is also plausible: investors may weaken incentives to lower risk. This could exacerbate the problems associated with NPM and PbR, including creaming, diverting resources from services to the professional class, and increasing the complexity and opacity of government (Warner, 2013).

Second, the SIB contract may pass down incentives to providers, while creating a *buffer* allowing providers to work as stewards (Pauly & Swanson, 2017). This argument is linked to the idea that social programs should face low-power, mediated incentives (Heinrich & Marschke, 2010), aligning SIBs with a "softer" post-NPM stewardship model, stressing the importance of inter-organizational collaboration and learning relationships (Joy & Shields, 2013). Investors buffer the drawbacks of performance management contracts (Callanan & Law, 2012), introducing an "informal" and relational performance management style (Fraser et al., 2018). This aligns with a post-NPM, NPG approach (Osborne, 2010), championed by investors with a social ethos (Carter, 2021). The buffer role of investors counterbalances the traditional dominance of state agencies over providers, enabling longer-term and more flexible contracts, improving programming and outcomes (Del Giudice and Migliavacca 2019; Jackson, 2013).

SIB implementation and the SIB effect depend on the national and project contexts. The Netherlands and Switzerland use *hybrid* social policy models with elements of *coordinated markets* (based on collaboration and incremental change) and *liberal markets* (with competition and potentially radical change) (Hall & Soskice, 2001), with the first predicting "buffered incentives" and the second predicting "intensified incentives." Regarding project context, both cases are pilot studies (Ettelt, Mays, & Allen, 2015). As pilots, stakeholders may want to show success to pave new roads for future work (Ansell & Bartenberger, 2016). This may motivate strategies that generate the impression of success, even if these strategies would not be acceptable at scale.

Based on the theoretical literature, we have three hypotheses. We expect the introduction of investors to impact contracts, shifting contracts toward a NPM or agency model with strong incentives or toward a NPG or stewardship model with buffered incentives. Either could cause positive SIB effects—through better goal alignment or more freedom for non-profit agencies to pursue their missions. Under both, a SIB should improve impact measurement since impacts are central to payment criterion. Thus:

- SIB financing should have a positive effect (H1).
- This could arise through stringent NPM or an agency model (H2a).
- Alternatively, this could arise through a NPG or stewardship model (H2b).

METHODS

To tackle the question of whether there is a SIB effect and why, this research takes a mixed-method approach similar to Burch and Heinrich (2015). The research team combined qualitative and quantitative insights throughout the project to collaboratively define the approach, iteratively adjust estimation techniques, and regularly discuss results to better understand the processes generating observed outcomes and to feed these results back into additional analyses. Jointly, we offer a robust objective estimate of the SIB effect and a clear mechanism to explain the result. Alone, each approach would have provided an incomplete, and even potentially false, understanding of the SIB effect observed in the studied contexts.

Case study background

The Dutch provider was a small, for-profit organization that transitioned from non-SIB to exclusively SIB financing. In the pre-SIB period, they served young adults on social assistance, expanding on typical ALMP strategies with volunteer mentors and targeting not only educational and employment goals but also self-employment. Table 2 shows there was no statistically significant change in clientele characteristics following the financing shift with respect to nationality, education, and gender, though age declined slightly. The introduction of SIBs introduced strong incentives. The potential maximum investor returns were in the double digits, while the provider was doubly incentivized, first through a pay by results (PbR) contract and second through the provider's and provider-CEO's investments.

In the Swiss case, the non-profit provider offered ALMPs to refugees. A single government agency financed the provider using non-SIB funding before and concurrently with SIB financing. Programming for all funding streams included job coaching and support classes, but SIB-funded clients received more post-employment support, fewer courses in language and computing, and fewer conversations with job coaches (Table 3). Although overall client flows to the provider were consistent over time, during the period they had SIB funding, they (following contract terms) selected into the SIB those with a higher education level, better German (the local language). and those less likely to be from Eritrea and more likely to be from Syria (implying a higher quality education), as well as more men and those with a refugee visa. Incentives for investors and providers were lower than in the Dutch case. The investor faced a maximum loss or gain of 1 percent annual interest. The provider's contract had a maximum win/loss of $+/-1.5$ percent over 5 years with contract targets designed such that raw enrollment numbers played a role in triggering payments—in practice, a hybrid PbR/per-client contract.

Qualitative methods

For the qualitative part of the study, we draw on comparative case study methods (Eisenhardt, 1989) to explore

the perceptions and broader narratives offered by informants as they reflect on their experiences of designing and delivering services. The two cases offering similar interventions permit a comparative analytical approach both within and across providers.

It was important to select informants from across the different SIB stakeholder groups to include multiple perspectives. As illustrated in Table 1, we interviewed representatives from government (administrators with either social or budgetary mandates), investors, and intermediary actors, as well as local evaluators alongside staff from provider organizations. With respect to provider organizations, we interviewed (1) senior strategic managers with a good overview on the rationale for SIB experimentation and delivery in each site, (2) mid-level operational managers with experience of supervising staff working on both SIB and non-SIB-financed programs, and (3) frontline staff delivering either SIB or non-SIB-financed programs, or both.

Interviews were conducted until "data saturation" (Glaser & Strauss, 1967) and were recorded, translated, transcribed, and coded using NVivo 12. Members of the research team discussed and reviewed the interview data alongside documentary material (local evaluations and annual reports) to ensure consistency. The data were interrogated repeatedly in order to understand key emergent issues, drawing on the principles of "constant comparison" (Glaser, 1965). The analytical approach drew on both inductive and deductive reasoning—exploring emergent issues alongside insights from wider policy, management, and economic theory (Langley, 1999). Whilst "recall bias" may be a problem in relation to retrospective interviews with some stakeholders, retrospective interviews can also have benefits—such as encouraging informants to critically appraise the original rationale for decisions relating to policy and practice. We were not granted access to the loan agreements or contracts between parties and thus explored contractual requirements through evaluations, audit reports, and interviews.

The qualitative element of a mixed-method study enables deep exploration of "how" and "why" questions through detailed contextualized accounts (Yin, 2013).

Quantitative methods

We look at the specific SIB effect, defined as the isolated causal impact of using SIB financing on outcomes. One way this can be measured is if providers manage different funding streams with overlapping eligibility and programming. Here we compare the program impact of Active Labor Market Programs (ALMPs) funded by SIB versus non-SIB financing within Swiss and Dutch providers.

Looking at the treatment group, the Dutch provider supplied individual identifiers and start dates for those in the pre-SIB and SIB periods. These were merged with the Dutch Statistical Agency's (CBS) social benefits dataset (education, migration background of self and parents, gender, year of birth, and information about benefit receipt like benefit type, start and end dates) and employment dataset (start and end dates of each employment and unemployment spell, income, and type of income). We merged 24 months of information prior to enrollment and 48 months following enrollment for the earliest cohort.

In Switzerland, the provider had detailed intake data on gender, civil status, children, country of origin, type of visa, mother tongue, German proficiency level, whether they spoke French and English, work experience, childcare, percentage of work sought, and process data on the classes participants attended and meetings with job coaches, as well as individual identifiers for the treated. We linked this to the National Statistics Office data on public benefits (level, type, start and end dates) and social security records (income level, type of earnings, and start and end dates). We again merged 24 months pre-enrollment and 40 months following enrollment for the first cohort.

Given no concurrent non-SIB group in the Dutch case and the incomparability of the concurrent group in the Swiss case, we employ a pre–post design. However, given the SIB and non-SIB funding take place at different times, and that changing economic conditions can affect outcomes at the same time, we create administrative comparison groups in both periods and cases. This approach allows us to separate the SIB effect from both the program effect and the potential effect of changing economic conditions on our outcome measures.

We selected comparison groups from administrative records, choosing those eligible for the programs with similar benefit and employment histories. In the Dutch case, this was those ages 17–27 on social benefits in the same city, while in Switzerland, this was working-age individuals with the same visa types in the same canton. To create comparison groups, we use start dates from each enrolment cohort and calculate whether individuals in the administrative data have an age and year of migration that would qualify them to be in the administrative comparison group for a specific cohort. We use the same

TABLE 1 Qualitative interview informant details.

	Providers	Government commissioners	Investor/Intermediaries	Evaluators	Totals
Netherlands	4	2	2	3	11
Switzerland	6	2	1	1	10
Totals	10	4	3	4	21

approach to ascertain that employment and benefit histories of individuals in the administrative comparison group fell within the observed range of benefit and employment histories of individuals in a specific enrolment cohort. If an individual from the administrative data was an appropriate match for multiple cohorts, the individual was randomly assigned to one cohort (sampling without replacement).[1]

In both cases, SIB clients were more employable, though there was significant overlap between pre-SIB and SIB groups within the provider. As visible in Table 2, in the Dutch case, the pre-SIB group included more women, more Dutch, and those with slightly more education. As visible in Table 3, in the Swiss case, the pre-SIB period included more women, more participants with a permanent visa rather than a refugee visa, and more individuals from Eritrea.

We used entropy balancing so that the administrative comparison and pre-SIB groups had the same mean characteristics as the SIB group (mean monthly benefit and employment histories in the 2 years prior to program participation and control variables) (Hainmueller, 2012). Entropy balancing calibrates unit weights so that the reweighted treatment and control groups match on multiple moments (i.e., mean, standard deviation) of selected characteristics. Results are very similar if we do not apply any weights and, instead, include the variables used for balancing in the difference in difference (DiD) regression as controls. See the Appendix S1 for balancing and parallel pre-trend analyses necessary before calculating DiD.

We take a DiD approach, comparing the baseline in months −24 to −5 with the weighting adjustment to months 1 onward. Skipping the 4 months prior to enrollment excludes the potential for the "Ashenfelter Dip," in which generally the earnings of participants decline just prior to enrollment—though in our data, because of the types of participants, we do not see a significant dip regardless. Then, for each month following program entry, we compare the difference between the provider group in the SIB-era and the pre-SIB era with the administrative group in the SIB-era and the pre-SIB era—a group level approach often taken in the policy literature (Cunningham, 2021). Formally,

$$[(\text{provider\_t2} - \text{comparison\_t2}) - (\text{provider\_t1} - \text{comparison\_t1})]$$

To additionally control for covariates, we use a DiD regression predicting

$$y_i = C + \beta_1\delta + \beta_2\tau + \beta_3\delta\tau + \boldsymbol{\beta_4 X} + \epsilon,$$

where δ still indicates provider (1) versus comparison group (0), and τ indicates the SIB (1) versus pre-SIB period (0). The interaction effect of the period and the provider dummies, β_3, is then the coefficient estimating the SIB effect, cleansed of any program and period effects related to the pre-SIB and SIB eras; δ indicates the program effect and τ the period effect.[2] $\boldsymbol{\beta_4}$ is a vector of coefficients for the control variables—the same variables listed in the description of weighting (education, visa, benefit and employment histories, etc.; see Tables S1 and S2).[3]

The primary assumption in a DiD analysis is that the treatment and control groups have parallel trends in the outcome, i.e., the treatment group, absent the treatment, would have followed the same time path as the control group for the outcome variables of interest (here, employment and social benefit receipt). To check this assumption, we look at pre-trends without weighting in the Appendix S1 (with weighting in the main text) and run a panel regression testing the same interaction as used in the DiD analysis, but treating time parametrically and taking a more conservative approach with individual random effects (Lalive, Morlok, & Zweimüller, 2011).

The second DiD assumption, of no pre-treatment effects, is addressed by not using months −4 to −1 as a baseline, as is standard in the literature—though, as visible in Figure S1, we do not see a pre-participation dip in employment and income for the youth and refugees in these programs anyhow.

A third potential complication is comparing different treatment effects in contexts where conditioning is needed to make the groups comparable (Callaway & Sant'Anna, 2021) —in our case, this was addressed by comparing results with and without conditioning, which did not differ and are thus not presented. That said, the DiD analysis with comparison groups from administrative data could not condition on variables that the qualitative analysis in Switzerland suggested might be important, like language. Therefore, a small separate analysis based on just the provider groups is also presented in Section 3.2, Figure 3. This analysis cannot adjust for

TABLE 2 The Dutch case, SIB and non-SIB groups.

	SIB	Pre-SIB
N	136	29
Male	0.55	0.31
Age at entry	23.7	24.4*
Parents at entry	0.23	0.28
Origin		
Native Dutch	0.21	0.28
1st gen	0.29	0.14
2nd gen	50.41	58.62
Non-western	0.72	0.69
Education		
Primary	0.02	0.00
Pre-vocational/lower	0.38	0.34
Secondary vocational	0.54	0.45
General/Pre-univ	0.04	0.10
University	0.05	0.10

TABLE 3 The Swiss clientele and treatment.

	SIB	Pre-SIB	Screened-out SIB	SIB-era, non-SIB
N	172	282	32	400
Age at entry	33.76	33.21	35.5	33
Duration since arrival	3.55	3.33	3.59	3
Male	81.4	69.86***	87.5	73.5***
Education				
<Compulsory	12.21	23.05***	15.62	26.5***
Compulsory	48.84	61.7**	65.62	50.7
Higher	38.95	15.25***	18.75**	22.8***
German language level				
A1	8.14	14.46	53.12***	35.4***
A2	36.05	43.47	37.5	37.3
B1+	55.81	42.17*	9.38***	27.2***
Residence permit				
B-FL	44.19	63.12***	40.62	52.2
F-FL	22.67	29.43	12.5	24.8
F-VA	33.14	7.45***	46.88	23**
Country of origin				
Eritrea	30.81	61.35***	31.25	55.9***
Syria	22.67	6.03***	15.62	14.5*
Afghanistan	8.14	1.77***	9.38	6.6
Other	38.37	30.85*	43.75	23***
Number conversations	4.8	4.71	2.0***	7.00***
DE course	23.84	40.43***	NA	23.2
Computer or job search course	1.16	17.02***	NA	19.3***

Note: *t*-tests compare to SIB group * = .05, ** = .01, *** = .001.

longitudinal changes, as there is no comparison to a concurrent administrative control group.

The Appendix S1 also presents a few additional analyses, including a re-analysis of the Dutch data in which groups are weighted to match the very small pre-SIB program group rather than the SIB group (yielding a better matching), an OLS of outcomes pooled over 2 years, so that program effects can be compared with "middle-term" ALMP impacts summarized (Card et al., 2018), and a short cohort analysis to examine whether program impacts improved over time—a hypothesis raised in qualitative interviews.

RESULTS

Balanced outcomes by group

Figure 1 shows trends in the balanced outcomes in both the Dutch (panel a) and the Swiss cases (panel b).[4] In the Netherlands, the provider group in the pre-SIB period had worse outcomes on every dimension: hours worked, earnings, and benefit receipt—a significant difference despite the small sample.[5] There is also a barely significant increase in benefit receipt among both the provider and comparison groups in the SIB period versus the comparison group in the pre-SIB period. Overall, the picture in the Netherlands is one of underperformance in the pre-SIB period and average performance during the SIB period.

In Switzerland, we see three strong trends. First, the provider group in the pre-SIB period continued to have lower employment and earnings up to 1 year from entry—a continued lock-in effect. Second, during the pre-SIB period, the provider's clients received benefits more often than the comparison group, with a growing divergence between the two. Third, the SIB period provider group had long-term higher employment and earnings. Overall, the picture in Switzerland is one of average performance in the pre-SIB period, followed by over-performance in the SIB period.

Program and SIB effects

Our models estimated monthly program effects, SIB-period effects, and the interaction of the two (capturing the SIB effect), with participants' employment (or hours worked), benefit receipt, and earnings as outcome

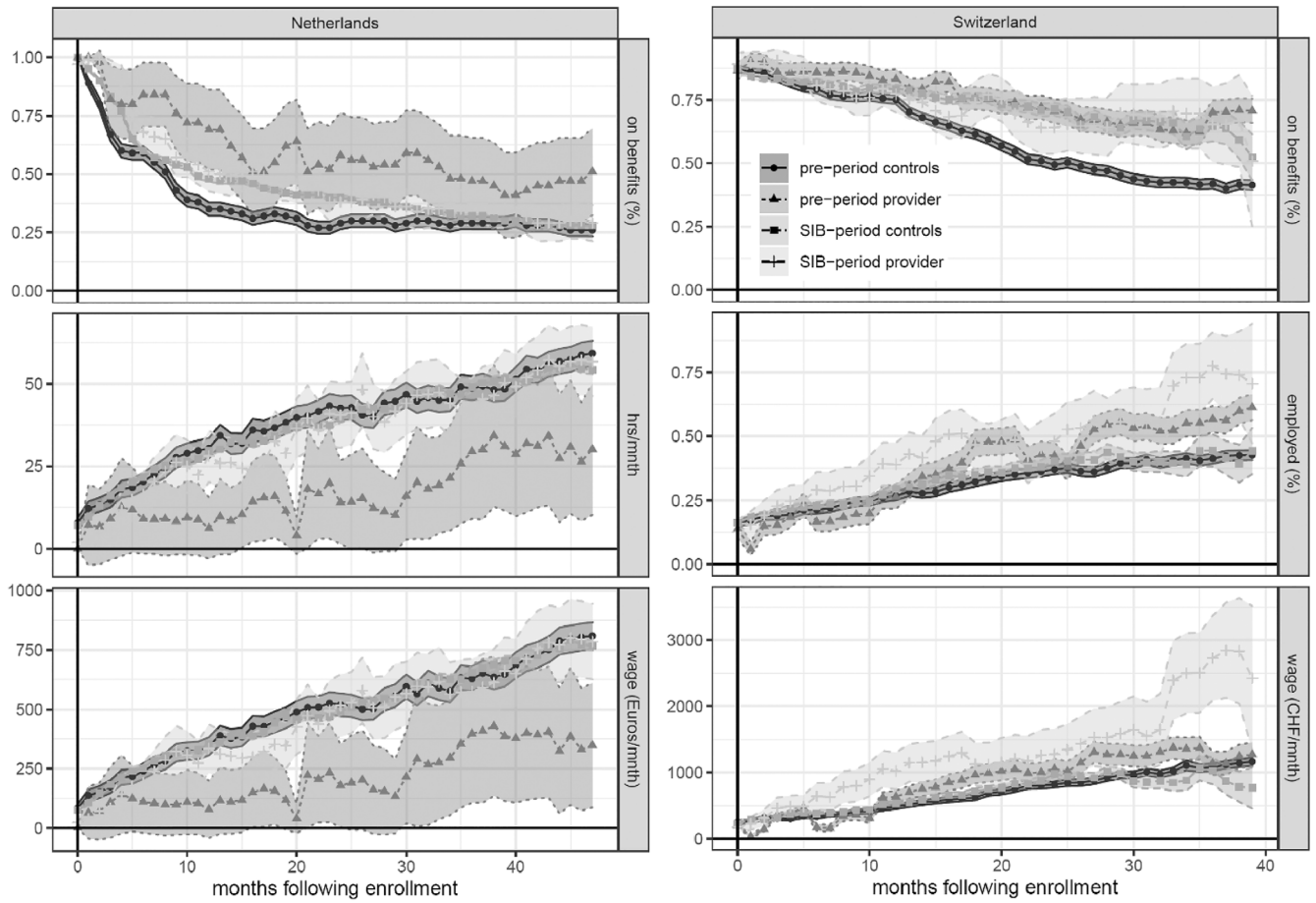

FIGURE 1 Balanced employment and benefit outcomes in Switzerland and the Netherlands.

variables.[6] There is no data on more recent cohorts toward the end of the period, so confidence intervals become larger over time. Results, in DiD units, are presented in Figure 2, with point estimates at 6, 18, and 48 months in Table 4. We first discuss the program effects and then the SIB effects.

Looking at the Dutch case (left panel in Figure 2), there are statistically significant negative *program effects* on all three outcomes in the first 10 months, with higher benefit receipt and lower work and earnings. Over time, negative program effects fade, with program participants converging with the comparison group. These are likely lock-in effects, i.e., the fact that during training job searching declines or ceases, causing an initial drop in the probability of employment for those in training programs. Our time frame of about 20 months is consistent with the literature (Lechner & Wunsch, 2009). It is, however, also possible that negative program effects are due to selection of less work-ready, eligible individuals into the program. Although the literature suggests considering 2-year benefit and employment histories should be sufficient, employment and benefit histories might be less valid indicators for young adults. If this is the case and there is equal selection into the program in the SIB and pre-SIB periods, this is not problematic. On the other hand, if there is more selection not controlled by the employment and benefit variables in the SIB-financed period, then period-specific analyses would suggest significantly higher (false) program effects in the SIB period, and the joint analysis across periods would suggest a strong positive (false) SIB effect.

In Switzerland, program effects were similar to the Dutch case with respect to work. In the first 2 years, the program had a negative effect on employment and earnings, fading to null in the long-term—again, likely due to lock-in effects. With respect to public benefits, the provider increased benefit receipt, perhaps because the program helped participants enroll. It is possible that for this group, two years of employment and benefit histories are insufficient controls. Refugees' short-term employment histories might be inaccurate indications of their potential employability. This problem is similar to that in the Dutch analysis focused on youth.

In both cases, estimates suggest negative program impacts. That said, selection effects cannot be entirely ruled out, as employment and benefit histories might offer insufficient controls for youth and refugees. However, even if there is negative selection, this is irrelevant

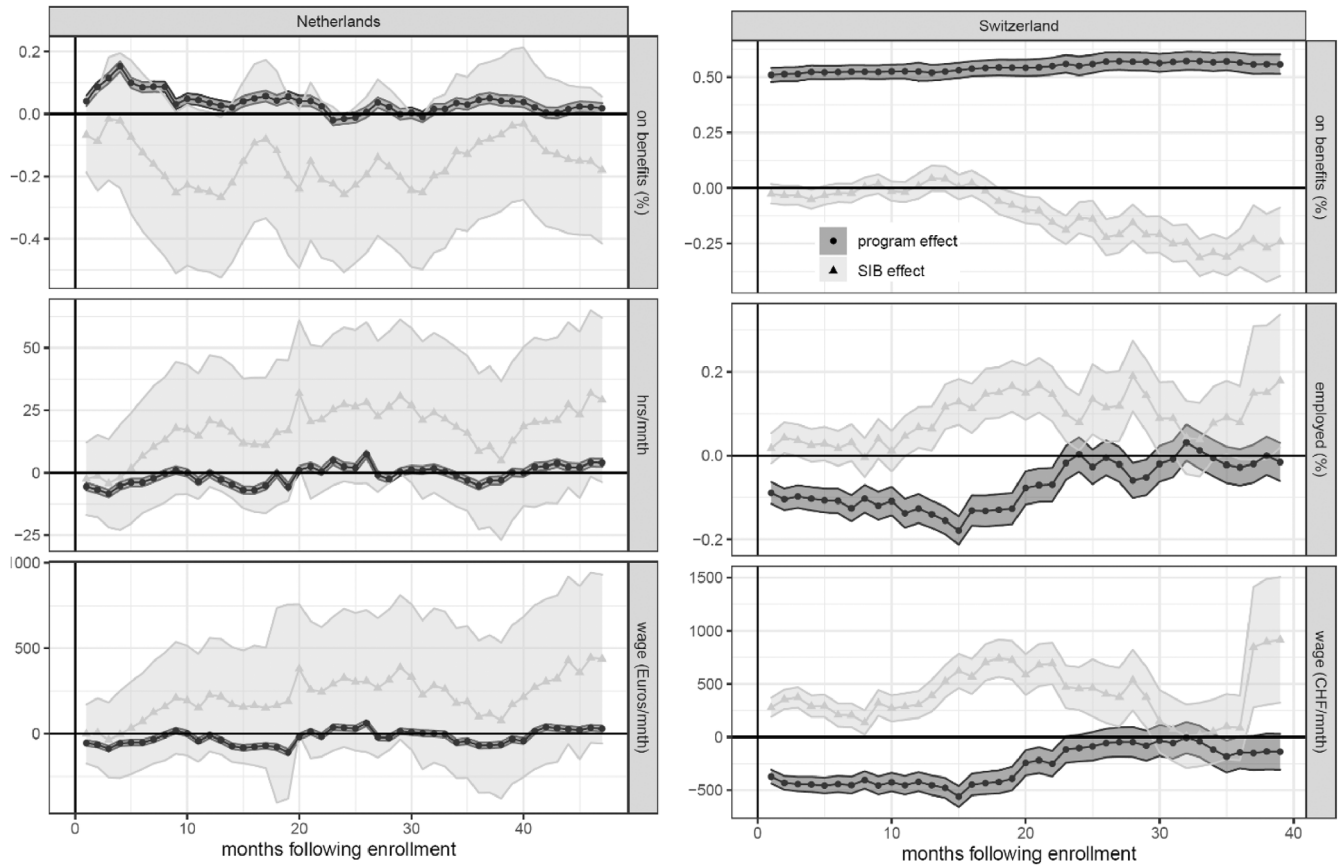

FIGURE 2 The SIB and Program Effects (Difference in Differences) in the Netherlands and Switzerland[17].

TABLE 4 Several point estimates of program and SIB Effects underlying Figure 3.

	Program			SIB		
	Benefits	**Hrs. Worked**	**Wage/mo.**	**Benefits**	**Hrs. Worked**	**Wage /mo.**
NL						
6 months	.085 (.01)	−3.83 (.79)	−50.948 (8.53)	−.124 (.13)	6.618 (11.61)	73.33 (142.67)
18 months	.043 (.01)	−.240 (.80)	−78.209 (9.85)	−.119 (.13)	16.126 (14.88)	166.11 (290.22)
48 months	.031 (.01)	1.406 (.82)	.235 (9.30)	−.191 (.12)	35.700 (17.05)	485.81 (256.29)
	Program			**SIB**		
	Benefits	**Employed**	**Wage/mo.**	**Benefits**	**Employed**	**Wage/mo.**
CH						
6 months	.522 (.02)	−.108 (.02)	−441.09 (39.69)	−.021 (.02)	.017 (.02)	207.81 (55.70)
18 months	.544 (.02)	−.129 (.02)	−421.73 (55.48)	−.059 (.03)	.151 (.03)	741.59 (90.07)
40 months	.543 (.02)	−.049 (.02)	−205.53 (85.81)	−.209 (.08)	.303 (.08)	1272.91 (301.66)

to the SIB effect estimate, *if selection is the same in both periods*. A biased estimate would show no effect in the pre-period and a small positive effect in the SIB period, while the non-biased estimate would show small positive effects in the pre-period and larger positive effects in the SIB period. Either way, the *difference* between the pre-SIB and SIB periods would be the same.

The SIB effect in the Dutch case is statistically significant and positive. SIB financing reduced benefit receipt consistently, and work and wages steadily went up from program entry for 2 years. Although results are statistically significant, the small N for the pre-SIB group means estimates are imprecise. In month 20, wages might have increased anywhere between +1 and +750 €. That said, the results are robust between the methods presented here and those presented in the Appendix S1, and we can confidently say there are positive SIB effects from program entry that have increased over time.

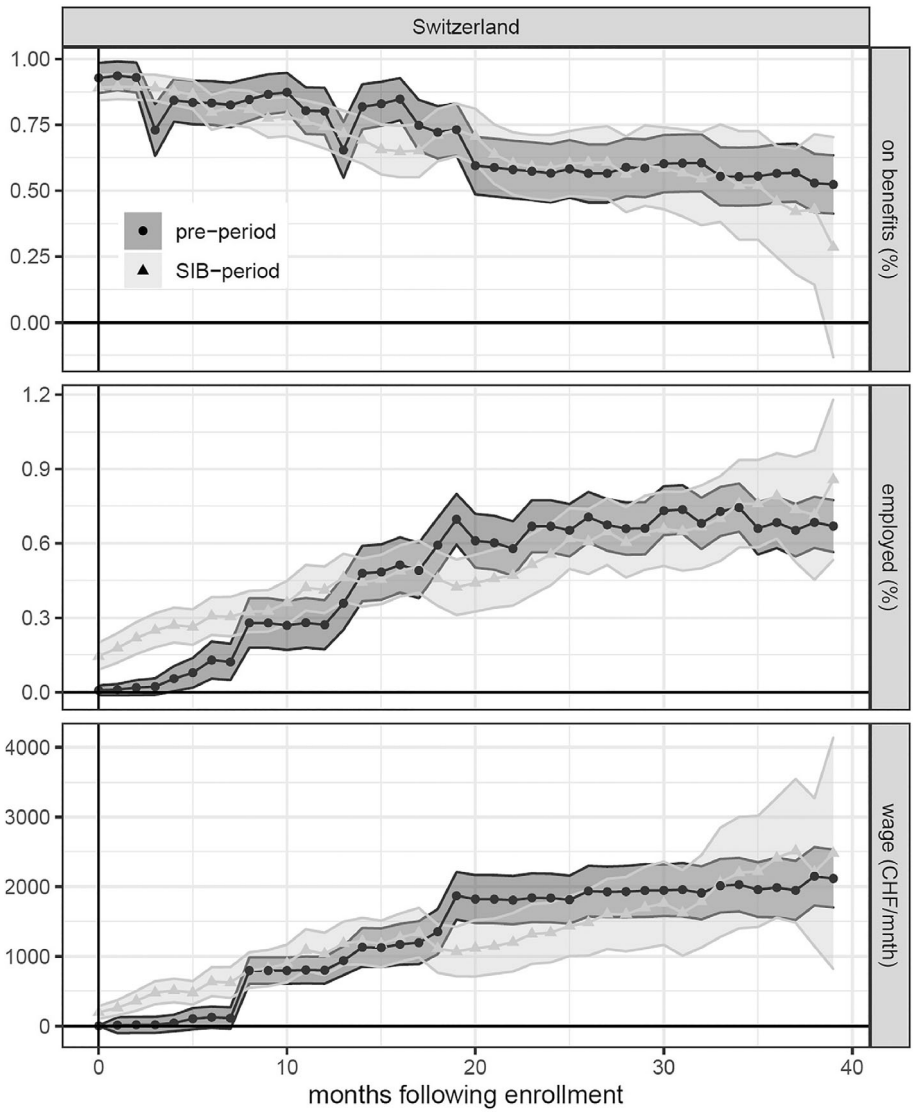

FIGURE 3 Swiss outcomes (benefit, employment, and income) for the Swiss SIB and pre-SIB groups, considering German language level.

In Switzerland, the SIB effect was also statistically significant and positive, though it took longer to surface. SIB financing decreased benefit receipt starting around month 20 after program entry with the effect increasing slightly over time. With respect to work, positive effects on employment and wages began around month 10 and increased over time. Though not depicted here (as estimates are less precise relying on the first few cohorts), in years 3 and 4 following program entry, SIB effects do not fade out, as is normally the case with ALMPs, but instead endure.

An important caveat is that if selection changed between the two periods, SIB effect estimates are inaccurate. The Swiss provider had information on language testing for 96 percent of SIB clients and 28 percent of pre-SIB clients. An analysis using language is less trustworthy as there is no administrative comparison group (so one cannot control for time effects) and there is likely selection on the language measure for the pre-SIB group, where the variable was largely missing. Still, incorporating language into weights, differences in employment and benefit trajectories disappear, as illustrated in Figure 3. This suggests the estimated SIB effects might be attributable to a shift in selection between the two periods.

Explaining SIB effects

In this section, we use qualitative evidence to explain the found positive SIB effects. Although we find that the specific program contexts were quite different, we find similar underlying causes for the SIB effect in the two cases.

The Netherlands

In the Netherlands, since 2004, local governments have been responsible for social benefits, and since 2007, they

have also been responsible for labor market integration. Provider informants reported that before using SIBs, they were in a weak bargaining position compared with the local government, with a lot of economic risk, short-term bilateral contracts, and an uncertain referral stream. Local government informants suggested that, particularly following the 2008 recession, outsourced programs were not placing enough people in work—potentially because of poor incentive systems.

The SIB experiment was led by the local municipality with significant local political support, challenging the existing low-trust NPM, principal-agent outsourcing model by introducing a more collaborative, neo-corporatist, principal-steward one. It was viewed by those involved as an innovative and exciting pilot to challenge institutional inertia and long-standing problems of youth unemployment. It was a high-profile undertaking, being one of the earliest SIB-financed programs in mainland Europe. Local politicians, government officials, investors, providers, management consultants, and local evaluation experts came together over a prolonged period to collectively pursue their shared goal. All actors wanted to work collaboratively to overcome problems and frame the project as a success. This shared focus on success impacted (1) selection and (2) resources, as well as substantive programing.

The shift in power dynamics is highlighted by the fact that for the first time under the SIB, the provider was able to negotiate for the government to guarantee a specific number of referrals and a longer term multilateral contract, resulting in a more work-ready pool of clients.

> When you get investors to the table, they look at this… power difference and they say 'but you have been signing a contract that may deliver… 100 referrals, it could also be zero if we put our money on the table we are not going to put our money in a project that might not get any people going in because then we'll never create a success," so that's the first thing that happens, and the investors create a sort of level playing field… so all of a sudden you have a two against one situation where there was always one looking down on the provider or providers, so it's a whole different ballgame. Provider strategic manager

A second selection mechanism was not visible in contracts but surfaced in interviews. The provider reported they were not required to "enroll" all referrals into the SIB but could instead select the more work-ready ones before official enrollment.

> What I did was I did not start with groups of 40, I started with more – sometimes so that I had – we had like two or three weeks to say to the persons, "OK, just try," and then we could also give them a chance but also say "OK, it doesn't work, we tried" and we had three weeks more to say "OK, we have another spot for other young people" and the nicest way, what I liked, is to start with 50 and then you have more chance to hit your targets. Provider operational manager

The SIB also changed the provider's budget. Although we do not know the exact budget per client, informants implied that it was higher than traditional ALMP programs funded by the local municipality.

> [The political leader] said "well if we're overpaying or not, that's not the point, because we can show, we show the world that we save far more than that we paid." And if you maybe overpaying that's different because the spread, I think we saved 3 or 2.5 m euros and we paid out 1.2 m, so the spread was very big. And then he said "well, as long as there is a spread I'm not concerned if we are overpaying this or not." Local municipality commissioner

More generous financing was politically palatable as the cost of the SIB was framed not in terms of the cost of delivering the program but rather in terms of its long-term impact. The provider received additional resources both directly through their provision contract and through returns on its investment.

In the Dutch case, the organization and CEO's high stakes created significant pressure. With the start of SIB financing, the CEO decided to focus fully on the provider rather than spread himself across different companies as before.

> I made in fifteen years of social entrepreneurship some money, but a lot of that money was in the SIB, so my pension was in there, so that was really exciting… I think the full team knew, like, OK this is the target we have, the 22 months, and we need to slap time off that… I think it really created an extra layer of tension and focus [for all staff] but, yes, from day one with [this provider] we wanted to help all members, so what I said first year 85 percent out of welfare, yeah, oh this is amazing, so yeah we have that focus but aligning that with the financial side of things then we really feel, like, usually we just say 'OK couple of thousand to the person and do your best' – it's a different situation. Provider CEO

Front-line staff as well as strategic and operational managers reported working under greater pressure

during the SIB program. This pressure likely motivated a visible shift in programming with SIB financing. Provider informants suggested that historically they had focused on helping clients become entrepreneurs. The 2013 (pre-SIB) annual report reflects this orientation with the staff talking about the clients' "dreams" and "ambitions" and client profiles focusing on those who successfully established a business—and they reported 18 percent of placements were in entrepreneurship. The 2018 (post-SIB) annual report shows a more typical ALMP program with general text about helping young people, profiles including educational placements and jobs, and a drop in the percentage of placements in self-employment to 14 percent.

In sum, SIB-financing seems to have impacted the Dutch program by first institutionalizing the selection of more work-ready clients through a larger guaranteed number of referrals as well as by not officially enrolling those who appeared to be less work-ready in the first few weeks. Second, because of the government's reorientation from allocating budget based on cost to potential long-term government savings, more resources were available. Third, the increased pressure and monitoring motivated the organization to weaken its traditional focus on entrepreneurship in favor of more traditional ALMP approaches.

Switzerland

Switzerland was less deeply affected by the 2008 recession with less cost-savings discourse than in the Netherlands. The SIB agenda was promoted by one lone social investor, while local political actors were less enthusiastic. Interviewees suggested that for those on the left, the mixing of private investment, profit, and vulnerable constituents central to the SIB idea was unattractive, while for those on the right, supporting refugees was not a priority. As in the Netherlands, the initial goal was to develop a SIB-financed program (unspecified sector), and the selection of an ALMP only came later, following discussions with local government actors about their local policy priorities. It was also similar in that all involved parties, *apart from the provider*, came together over a prolonged period to collectively design a successful SIB experiment. Qualitative data demonstrated a similar shared commitment to enabling the provider organization (and the SIB project) to be seen as successful, though this desire was framed by Swiss government respondents as "avoiding embarrassment" rather than "showing the viability of a new policy model," as in the Netherlands.

Unlike the Netherlands, where the provider was an active SIB proponent and invested in the SIB, in Switzerland, the provider was recruited through a competitive tendering process after the multilateral SIB contract had been negotiated by the other parties—a model closer to an agency than a stewardship model. Returns were lower in the Swiss case, with the government resisting pressure from investors, fearful of creating the impression that investors were profiting from social programs. Swiss front-line workers, like the government, were also more skeptical about SIBs than in the Netherlands. One front-line worker joked "*This bonus system will change maybe something if the bonus goes in my pocket!*"—the direct opposite of front-line workers' assessment in the Netherlands.

Despite key differences with the Netherlands, SIB financing seems to have also influenced both (1) selection and (2) program resources.

The Swiss had a standardized process for selecting the more work-ready into their SIB-financed program. Selection occurred in three stages: (1) Based on intake data (2) Over a two-week metal workshop where they observed participants; and (3) Cross-referring following official enrollment. This was not perceived as "creaming," but rather as the direct implementation of contract conditions that specified the SIB serve "middle- and high-skill" clients. It was clear to all management and staff that the SIB enrolled only the most work-ready.

> In the non-SIB programs, we have a lower education level from the clients, you see. I think that's the major difference… We take everyone … And in the SIB program clients have better education, better linguistic proficiency, and stuff like this. So, for me it's absolutely clear why they have a higher outcome." Provider operations manager

In addition, there was selection due to a change in the evaluation. Originally, the program was planned to be evaluated using a randomized control trial. However, given the greater resources in the SIB-funded program than in other parallel non-SIB programming, the decision was made to enroll the control group partway through. This meant the provider had a larger pool of potential participants to choose from, so more selection was possible. Canceling the RCT also improved measured program success, as one contract target was raw enrollment numbers and because employment targets used flat numbers rather than rates (Hevenstone, Fraser, Hobi, & Geuke, 2023).

With respect to substantive changes, the primary shift with SIB-financing was resources. Per-client budgets in the 2 years prior to SIBs were 3890 and 5510 CHF, increasing to 5–6000 CHF during the SIB period. In contrast, in the first 2 years, the SIB-funded program had 10,710 and 11,250 CHF available per client.

The higher budget can partially be attributed to the decision to remove wage subsidies from the program while keeping budgets constant. This was a response to front-line workers' experience that employers preferred to offer clients temporary contracts whilst judging the individual's ability irrespective of subsidies. Although the

quantitative data suggests that SIB-clients did not have more contacts with job coaches or classes, interviews suggest this generous funding allowed for additional support not visible in the quantitative data.

> If we say "oh we have this possibility to pay the salary 40 percent for six months" and [most of the employers] say "no, it's not necessary," so it makes our budget… huge… and then at the end we thought yeah then we can use the money for different skills, so I remember the first client comes and says 'oh my boss said I need a driving license', driving license in Switzerland costs about 2000 Swiss francs … And then we had a big discussion internally and say "no, we need to help him because the boss says it's necessary" and then we pay 1000/2000 Swiss francs for that. And this happens a lot… and we pay swimming courses, we pay computer, CAD courses, everything that the client needs if he has a job. Provider caseworker

The Swiss and the Dutch cases differ in how key actors felt about SIBs. In the Dutch case, there was more provider and government buy-in, greater power realignment, and increased pressure due to incentives. In the Swiss case, there was less pressure around financial incentives reported by providers, but more around "face-saving," due to the political sensitivity of both SIBs and refugee policy. Yet the cases were similar in generating pressure to deliver a "successful" program. This pressure led to the institutionalization of selection and an increase in resources. The Dutch case differed in that there was also a programmatic shift toward more typical ALMP approaches.

DISCUSSION

Social impact bonds have been promoted with the aim of improving PbR incentives. Under SIBs, part of investors' reward is based on the additional impact attributable to financing, the definition used here of the "SIB effect." Despite the importance of the SIB effect in designing investor contracts, not to mention the choice to use a SIB, there has been no rigorous evidence of the SIB effect to date. In this article, we set out to rectify this gap in knowledge and to answer whether SIB financing has an impact, and if so, why.

Our first research aim was to estimate a SIB effect for two ALMPs, one targeting young people on social assistance in the Netherlands, and a second targeting refugees in Switzerland. We found significant positive SIB effects on public benefits, employment, and earnings in both, confirming Hypothesis 1. This finding would seem to confirm speculation that SIBs effectively align incentives (Pauly & Swanson, 2017) compared with concerns that they exacerbate the problems seen in PbR (Warner, 2013). Based on the quantitative evidence alone, SIBs seem like an effective new instrument to improve public contracting.

Our second research aim was to examine the conditions leading to the measured SIB effect and consider whether this might be seen as the "intensification" or "buffering" of incentives, i.e., shifting toward models described by agency or stewardship theory, respectively (Van Slyke, 2007).

Qualitative findings suggested actors from the provider organizations working on SIB-financed programs were more aware of contract targets than their colleagues from the same organizations working on non-SIB-financed programs, suggesting an NPM-style intensification of incentives (Warner, 2013; Dayson et al., 2020; French et al. 2022), aligning with agency theory. However, the SIB also buffered incentives for providers because the SIB contract *simultaneously* intensified pressure on governments, who institutionalized the unintended "creaming" often seen in the PbR literature, akin to findings of Koning & Heinrich, 2013. This created a neo-corporate and relational post-NPM NPG type approach (Joy & Shields, 2013; Osborne, 2010) aligning with stewardship theory. In some sense, this could be read as a partial confirmation of hypothesis 2b, insofar as the government offered generous and long-term contracts. However, the pathway was not what was anticipated. It was not that long-term flexible contracts led to improved programming but rather that the initial intensification of incentives led to choices, creating the potentially false impression of improved outcomes.

How are these findings relevant to scholars of public administration? SIBs, as multilateral partnerships (Del Giudice and Migliavacca, 2019; Fraser et al., 2018; Dixon, 2020), made projects more politically significant, raising the stakes, increasing the pressure on government actors, and changing the dynamics described by agency and stewardship theories. In this multi-actor context, local political considerations and inter-organizational actor relationships played important roles in determining how agency and stewardship principles came to be used in concert (Van Slyke, 2007). The presence of investors seems to have improved goal alignment and generated a collective good will, leading to collusion, institutionalized creaming, and increased resource investment. At the same time, SIB contracts increased pressure on providers, so that monitoring requirements may have some characteristics akin to the principal-agent model (French et al, 2022) embedded within a stewardship model. One potential reason for these divergent experiences could be the lack of competition in the investor market compared with the provider market (Del Giudice and Migliavacca, 2019).

An important caveat to our findings is that there is a longitudinal element we could not fully observe. In contrast to Van Slyke (2007), our data seemed to suggest the

contracting relationships in these SIBs began under a stewardship model, with elements more aligned with principle–agent relationships possibly evolving later. The Swiss government did not pursue further SIBs and did not offer the provider further contracts following our study. In the Dutch case, government used stricter and more rigorous contractual terms, more akin to a principle-agent model in further SIBs (Hevenstone et al., 2023). Both cases in this study were early SIB contracts in their respective locations, meaning success was important, both with respect to political liabilities and the potential to develop further SIBs.

Our third research question relates to what these findings imply with respect to the feasibility of measuring impact. Although there is a solid body of evidence suggesting observational methods using several years of benefit and employment data generate impacts like those found in randomized control trials, no statistical method could net out the institutionalized selection procedures available to stakeholders working together on the ground. This means that a SIB model with stewardship or post-NPM characteristics could inadvertently undermine measurement, making it impossible to have the methodologically rigorous evaluations called for in the literature (Heinrich & Kabourek, 2019). Perceptions of evaluation reliability are related to factors like political context, policy area, and the reputation of those conducting evaluations (MacKillop & Downe, 2022; Schmidthuber, Willems, & Krabina, 2022), but actual reliability can also be related to service contract conditions.

Our theoretical contribution is to highlight the political and relational factors explaining why SIB stakeholders may behave as they do and to demonstrate how these factors may influence programs through engagement with agency and stewardship theories (Van Slyke, 2007). This builds on existing debates in the SIB literature, which have frequently been posed by scholars in terms of an extension or diminution of NPM (Warner, 2013, Fraser et al., 2018; French et al., 2022). We highlight the multilateral nature of SIB contracts and relationships (Del Giudice and Migliavacca, 2019; Fraser et al., 2018; Dixon, 2020) in contrast to bilateral contracts and suggest this may explain the seemingly contradictory nature of our empirical findings and general ambiguities in SIBs (Maier et al., 2018).

The practical implications of our study are that multilateral contracting adds complexity to public administration compared with bilateral contracting. Findings suggest multilateral contracting may raise governance or regulatory challenges as well as informational challenges with current institutional safeguards poorly suited to the pressures of new multilateral SIBs, including potentially powerful actors such as investors.

The study has limitations. First, regarding context and generalizability, the Netherlands and Switzerland share similar hybrid "corporatist-liberal" approaches to social policy that may be unique in international comparison. It may be that only in this context do we see 'corporatist' approaches to policy dressed in liberal clothing. Both cases took place in "experimental" contexts with stakeholders arguing future benefits could not be accurately calculated within the projects. This may be specific to pilots, with more scrutiny in scaled projects. Second, the estimation technique used here could overestimate a SIB effect if programs were able to make a greater impact on outcomes during the SIB-financed period—which could be true for the Dutch context, where unemployment rates declined slightly. Third, there is a conceptual limitation. We defined the SIB effect as the isolated impact of the investor. However, the qualitative analysis found SIB financing changed referrals and resources that did not necessarily require the investor. It is unclear whether these belong to the SIB effect or should be netted out. This raises general questions on how to estimate the impact of contract incentives.

Finally, we discuss the implications of this work for future research. We have two main calls. First, clear, empirically testable definitions of the SIB effect need to be established. Second, studies need to generate precise estimates of these SIB effects, with clear evidence of the underlying causes or mechanisms. It is concerning that contracting practices expand without any hard evidence.

ACKNOWLEDGEMENTS

We would like to thank the Swiss Network for International Studies for funding our project. The project has also benefited from working with the Swiss Federal Statistics Office, Statistics Netherlands, and the International Labour Organization's Department of Social Finance. This paper has been long in the making, and we are indebted to more people than we can mention here by name. We received valuable feedback on our early ideas and drafts of the paper from colleagues at BFH in Bern and KBS in London, as well as colleagues affiliated with the GO Lab in Oxford. Finally, we thank the anonymous reviewers and the editor at PAR for very thoughtful and constructive feedback. Open access funding provided by Berner Fachhochschule.

ENDNOTES

[1] For both client groups—youth on social assistance (NL) and refugees (CH)—ALMPs are widespread, so individuals in the comparison groups might receive ALMP assistance at another provider. Estimated program effects thus compare providers not to no treatment at all but to the average treatment provided to those target groups.

[2] In addition, we ran DiD estimates for program and SIB effects separately, using the information on individual outcomes before enrollment. For program effects, in the SIB era we calculated the interaction between the pre/post dummy and the provider/administrative dummy as a program effect. Then we estimated the SIB effect using only information on those at the provider, with the interaction between the pre/post dummy and the era (SIB vs. pre-SIB) indicating the SIB effect. Both of these are available upon request.

[3] We ran without controls, with the same controls as used for weighting with weighting, and with the same controls but without weighting—with similar results.

[4] Outcomes are slightly different between the two cases because of data availability (hours worked vs. percent employed), each panel must be read independently—y labels cannot be aligned.

[5] Because balancing such a small group with the SIB-period provider group was difficult, in the Appendix S1 we reweighted the administrative pre group to match the provider group. Balancing improved, while the significantly worse outcomes for the pre-provider group remained.

[6] See the Appendix S1 for results using a random effects model pooling months within a single model. Results are similar.

[7] At the time of these results, exchange rates were close to 1CHF = 1 Euro = 1 USD.

REFERENCES

Anders, J. D., and R. Dorsett. 2017. *HMP Peterborough Social Impact Bond – Cohort 2 and Final Cohort Impact Evaluation*." Report. London, UK: Ministry of Justice. https://www.gov.uk/government/publications/final-results-for-cohort-2-of-the-social-impact-bond-payment-by-results-pilot-at-hmp-peterborough.

Ansell, Christopher K., and Martin Bartenberger. 2016. "Varieties of Experimentalism." *Ecological Economics* 130: 64–73.

Baliga, Shifali. 2013. "Shaping the Success of Social Impact Bonds in the United States: Lessons Learned from the Privatization of U.S. Prisons Note." *Duke Law Journal* 63: 437–80.

Barnow, Burt S. 2000. "Exploring the Relationship between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act." *Journal of Policy Analysis and Management* 19(1): 118–41. https://doi.org/10.1002/(SICI)1520-6688(200024)19:1<118::AID-PAM7>3.0.CO;2-4.

Blank, Rebecca M. 2000. "When Can Public Policy Makers Rely on Private Markets? The Effective Provision of Social Services." *The Economic Journal* 110(462): 34–49. https://doi.org/10.1111/1468-0297.00519.

Boyne, George A. 1998. "Bureaucratic Theory Meets Reality: Public Choice and Service Contracting in U. S. Local Government." *Public Administration Review* 58(6): 474–84. https://doi.org/10.2307/977575.

Brown, Trevor L., and Matthew Potoski. 2003. "Managing Contract Performance: A Transaction Costs Approach." *Journal of Policy Analysis and Management* 22(2): 275–97. https://doi.org/10.1002/pam.10117.

Brunjes, Benjamin M. 2020. "Competition and Federal Contractor Performance." *Journal of Public Administration Research and Theory* 30(2): 202–19. https://doi.org/10.1093/jopart/muz027.

Brunjes, Benjamin M. 2022. "Your Competitive Side Is Calling: An Analysis of Florida Contract Performance." *Public Administration Review* 82(1): 83–101. https://doi.org/10.1111/puar.13430.

Burch, Patricia, and Carolyn J. Heinrich. 2015. *Mixed Methods for Policy Research and Program Evaluation*. Newburry Park CA: SAGE Publications.

Burgess, Simon, and Marisa Ratto. 2003. "The Role of Incentives in the Public Sector: Issues and Evidence." *Oxford Review of Economic Policy* 19(2): 285–300. https://doi.org/10.1093/oxrep/19.2.285.

Callanan, L., and J. Law. 2012. *Will Social Impact Bonds Work in the United States?* New York: McKinsey on Society.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, Themed Issue: Treatment Effect 1 225(2): 200–30. https://doi.org/10.1016/j.jeconom.2020.12.001.

Card, David, Jochen Kluve, and Andrea Weber. 2018. "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations." *Journal of the European Economic Association* 16(3): 894–931. https://doi.org/10.1093/jeea/jvx028.

Carter, Eleanor. 2021. "More than Marketised? Exploring the Governance and Accountability Mechanisms at Play in Social Impact Bonds." *Journal of Economic Policy Reform* 24(1): 78–94. https://doi.org/10.1080/17487870.2019.1575736.

Chamaki, Foroogh Nazari, Glenn Paul Jenkins, and Majid Hashemi. 2019. "Social Impact Bonds: Implementation, Evaluation, and Monitoring." *International Journal of Public Administration* 42(4): 289–97. https://doi.org/10.1080/01900692.2018.1433206.

Cohen, R. 2011. "Harnessing Social Entrepreneurship and Investment to Bridge the Social Divide." In *EU Conference on the Social Economy*. Brussels: European Economic and Social Committee.

Courty, Pascal, Do Han Kim, and Gerald Marschke. 2011. "Curbing Cream-Skimming: Evidence on Enrolment Incentives." *Labour Economics* 18(5): 643–55. https://doi.org/10.1016/j.labeco.2011.01.007.

Cunningham, Scott. 2021. *Causal Inference: The Mixtape*. New Haven CT: Yale University Press. https://doi.org/10.2307/j.ctv1c29t27.

Dayson, Chris, Alec Fraser, and Toby Lowe. 2020. "A Comparative Analysis of Social Impact Bond and Conventional Financing Approaches to Health Service Commissioning in England: The Case of Social Prescribing." *Journal of Comparative Policy Analysis: Research and Practice* 22(2): 153–69. https://doi.org/10.1080/13876988.2019.1643614.

Del Giudice, Alfonso, and Milena Migliavacca. 2019. "Social impact bonds and institutional investors: An empirical analysis of a complicated relationship." *Nonprofit and Voluntary Sector Quarterly* 48(1): 50–70.

Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review." *The Journal of Human Resources* 37(4): 696–727. https://doi.org/10.2307/3069614.

Dixon, Ruth. 2020. "Performance Management in Social Impact Bonds: How an Outcomes-Based Approach Shapes Hybrid Partnerships." *International Journal of Public Sector Management* 34(3): 356–76. https://doi.org/10.1108/IJPSM-03-2020-0070.

Dunleavy, Patrick, Helen Margetts, Simon Bastow, and Jane Tinkler. 2006. "New Public Management Is Dead—Long Live Digital-Era Governance." *Journal of Public Administration Research and Theory* 16(3): 467–94. https://doi.org/10.1093/jopart/mui057.

Edmiston, Daniel, and Alex Nicholls. 2018. "Social Impact Bonds: The Role of Private Capital in Outcome-Based Commissioning." *Journal of Social Policy* 47(1): 57–76. https://doi.org/10.1017/S0047279417000125.

Eisenhardt, Kathleen M. 1989. "Building Theories from Case Study Research." *Academy of Management Review* 14(4): 532–50. https://doi.org/10.5465/amr.1989.4308385.

Ettelt, Stefanie, Nicholas Mays, and Pauline Allen. 2015. "Policy Experiments: Investigating Effectiveness or Confirming Direction?" *Evaluation* 21(3): 292–307. https://doi.org/10.1177/1356389015590737.

FitzGerald, Clare, Franziska Rosenbach, Tanyah Hameed, Ruth Dixon, and Jo Blundell. 2021. "New Development: Rallying Together—The Rationale for and Structure of Collaborative Practice in England." *Public Money & Management* 42(5): 1–4. https://doi.org/10.1080/09540962.2021.1981628.

Fox, Chris, and Kevin Albertson. 2011. "Payment by Results and Social Impact Bonds in the Criminal Justice Sector: New Challenges for the Concept of Evidence-Based Policy?" *Criminology & Criminal Justice* 11(5): 395–413. https://doi.org/10.1177/1748895811415580.

Fox, Chris, and Kevin Albertson. 2012. "Is Payment by Results the Most Efficient Way to Address the Challenges Faced by the Criminal Justice Sector?" *Probation Journal* 59(4): 355–73. https://doi.org/10.1177/0264550512458473.

Fraser, Alec, Stefanie Tan, Mylene Lagarde, and Nicholas Mays. 2018. "Narratives of Promise, Narratives of Caution: A Review of the Literature on Social Impact Bonds." *Social Policy & Administration* 52(1): 4–28. https://doi.org/10.1111/spol.12260.

French, Richard D. 2019. Is It Time to Give up on Evidence-Based Policy? Four Answers https://doi.org/10.1332/030557318X15333033508220.

French, Max, Jonathan Kimmitt, Rob Wilson, David Jamieson, and Toby Lowe. (2022). "Social impact bonds and public service reform: back to the future of New Public Management?" *International Public Management Journal* 1–20.

Giacomantonio, Chris. 2017. "Grant-Maximizing but Not Money-Making: A Simple Decision-Tree Analysis for Social Impact Bonds." *Journal*

of Social Entrepreneurship 8(1): 47–66. https://doi.org/10.1080/19420676.2016.1271348.

Girth, Amanda M. 2017. "Incentives in Third-Party Governance: Management Practices and Accountability Implications." *Public Administration Review* 77(3): 433–44. https://doi.org/10.1111/puar.12645.

Glaser, Barney G. 1965. "The Constant Comparative Method of Qualitative Analysis." *Social Problems* 12(4): 436–45. https://doi.org/10.2307/798843.

Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine de Gruyter.

GO Lab. 2020. Impact Bond Dataset https://golab.bsg.ox.ac.uk/knowledge-bank/indigo/impact-bond-dataset-v2/.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1): 25–46.

Hall, Peter A., and David Soskice, eds. 2001. *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford, New York: Oxford University Press.

Heckman, James J., Carolyn Heinrich, and Jeffrey Smith. 2002. "The Performance of Performance Standards." *The Journal of Human Resources* 37(4): 778–811. https://doi.org/10.2307/3069617.

Heinrich, Carolyn J. 2007. "False or Fitting Recognition? The Use of High Performance Bonuses in Motivating Organizational Achievements." *Journal of Policy Analysis and Management* 26(2): 281–304. https://doi.org/10.1002/pam.20244.

Heinrich, Carolyn J., and Sarah E. Kabourek. 2019. "Pay-for-Success Development in the United States: Feasible or Failing to Launch?" *Public Administration Review* 79(6): 867–79. https://doi.org/10.1111/puar.13099.

Heinrich, Carolyn J., and Gerald Marschke. 2010. "Incentives and their Dynamics in Public Sector Performance Management Systems." *Journal of Policy Analysis and Management* 29(1): 183–208. https://doi.org/10.1002/pam.20484.

Hevenstone, Debra, Alec Fraser, Lukas Hobi, and Gemma Geuke. 2023. "Why Is Impact Measurement Abandoned in Practice? Evidence Use in Evaluation and Contracting for Five European Social Impact Bonds." *Evaluation* 29(1): 91–109.

Jackson, Edward T. 2013. "Evaluating Social Impact Bonds: Questions, Challenges, Innovations, and Possibilities in Measuring Outcomes in Impact Investing." *Community Development* 44(5): 608–16. https://doi.org/10.1080/15575330.2013.854258.

Jahn, Elke J., and Wolfgang Ochel. 2007. "Contracting-out Employment Services: Temporary Agency Work in Germany." *Journal of European Social Policy* 17(2): 125–38. https://doi.org/10.1177/0958928707075196.

Joy, Meghan, and John Shields. 2013. "Social Impact Bonds: The Next Phase of Third Sector Marketization?" *Canadian Journal of Nonprofit and Social Economy Research* 4(2): 39–55. https://doi.org/10.22230/cjnser.2013v4n2a148.

Koning, Pierre, and Carolyn J. Heinrich. 2013. "Cream-Skimming, Parking and Other Intended and Unintended Effects of High-Powered, Performance-Based Contracts." *Journal of Policy Analysis and Management* 32(3): 461–83. https://doi.org/10.1002/pam.21695.

Lalive, Rafael, Michael Morlok, and Josef Zweimüller. 2011. *Applying for Jobs: Does ALMP Participation Help?*" Working Paper No. 19. Zurich Switzerland: University of Zurich. info:doi/10.5167/uzh-51523.

Langley, Ann. 1999. "Strategies for Theorizing from Process Data." *Academy of Management Review* 24(4): 691–710. https://doi.org/10.5465/amr.1999.2553248.

Larsen, Flemming, and Sharon Wright. 2014. "Interpreting the Marketization of Employment Services in Great Britain and Denmark." *Journal of European Social Policy* 24(5): 455–69. https://doi.org/10.1177/0958928714543903.

Lechner, Michael, and Conny Wunsch. 2009. "Are Training Programs more Effective when Unemployment Is High?" *Journal of Labor Economics* 27(4): 653–92. https://doi.org/10.1086/644976.

Leventhal, R. 2012. "Effecting Progress: Using Social Impact Bonds to Finance Social Services." *New York University Journal of Law & Business* 9: 511–34.

Liebman, Jeffrey. 2011. Social Impact Bonds https://www.hks.harvard.edu/publications/social-impact-bonds.

Liebman, Jeffrey, and Avi Feller. 2014. The Economics and Econometrics of Social Impact Bonds.

Lowe, Toby, Jonathan Kimmitt, Rob Wilson, Mike Martin, and Jane Gibbon. 2019. *The Institutional Work of Creating and Implementing Social Impact Bonds*, Vol 47 353–69. Bristol UK: Policy Press. https://doi.org/10.1332/030557318X15333032765154.

Lu, Jiahuan. 2016. "The Performance of Performance-Based Contracting in Human Services: A Quasi-Experiment." *Journal of Public Administration Research and Theory* 26(2): 277–93. https://doi.org/10.1093/jopart/muv002.

MacKillop, Eleanor, and James Downe. 2022. "What Counts as Evidence for Policy? An Analysis of Policy Actors' Perceptions." *Public Administration Review* 1–14. https://doi.org/10.1111/puar.13567.

Maier, Florentine, Gian Paolo Barbetta, and Franka Godina. 2018. "Paradoxes of Social Impact Bonds." *Social Policy & Administration* 52(7): 1332–53. https://doi.org/10.1111/spol.12343.

Osborne, Stephen P. 2010. *The New Public Governance?: Emerging Perspectives on the Theory and Practice of Public Governance*. Milton Park UK: Routledge & CRC Press. https://www.routledge.com/The-New-Public-Governance-Emerging-Perspectives-on-the-Theory-and-Practice/Osborne/p/book/9780415494632.

Pauly, Mark V., and Ashley Swanson. 2017. "Social Impact Bonds: New Product or New Package?" *The Journal of Law, Economics, and Organization* 33(4): 718–60. https://doi.org/10.1093/jleo/ewx012.

Pihl-Thingvad, Signe. 2016. "The Inner Workings of Performance Management in Danish job Centers: Rational Decisions or Cowboy Solutions?" *Public Performance & Management Review* 40(1): 48–70. https://doi.org/10.1080/15309576.2016.1177556.

Romzek, Barbara S., and Jocelyn M. Johnston. 2005. "State Social Services Contracting: Exploring the Determinants of Effective Contract Accountability." *Public Administration Review* 65(4): 436–49. https://doi.org/10.1111/j.1540-6210.2005.00470.x.

Sacristán López de los Mozos, Ignacio, Antonio Rodríguez Duarte, and Óscar Rodríguez Ruiz. 2016. "Resource Dependence In Non-Profit Organizations: Is it Harder to Fundraise if you Diversify your Revenue Structure?" *Voluntas: International Journal of Voluntary and Nonprofit Organizations* 27(6): 2641–65. https://doi.org/10.1007/s11266-016-9738-8.

Selviaridis, Kostas, and Finn Wynstra. (2015). Performance-based contracting: a literature review and future research directions. *International Journal of Production Research* 53(12): 3505–3540.

Schmidthuber, Lisa, Jurgen Willems, and Bernhard Krabina. 2022. "Trust in Public Performance Information: The Effect of Data Accessibility and Data Source." *Public Administration Review* 83: 279–95. https://doi.org/10.1111/puar.13603.

Tan, Stefanie, Alec Fraser, Neil McHugh, and Mildred E. Warner. 2021. "Widening Perspectives on Social Impact Bonds." *Journal of Economic Policy Reform* 24(1): 1–10. https://doi.org/10.1080/17487870.2019.1568249.

Van Slyke, D. M. 2007. "Agents or Stewards: Using Theory to Understand the Government-Nonprofit Social Service Contracting Relationship." *Journal of Public Administration Research and Theory* 17(2): 157–87. https://doi.org/10.1093/jopart/mul012.

Warner, Mildred E. 2013. "Private Finance for Public Goods: Social Impact Bonds." *Journal of Economic Policy Reform* 16(4): 303–19. https://doi.org/10.1080/17487870.2013.835727.

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York; London: Free Press; Collier Macmillan.

Yin, Robert K. 2013. "Validity and Generalization in Future Case Study Evaluations." *Evaluation* 19(3): 321–32. https://doi.org/10.1177/1356389013497081.

AUTHOR BIOGRAPHIES

Debra Hevenstone is a Professor of social work at the Bern University of Applied Sciences in Switzerland. Her research focuses on social policy including social insurance design, labor market integration programs, digitalization in the social services, and contracting. Debra also has an interest in diverse quantitative methods and interdisciplinary research. Email: debra.hevenstone@bfh.ch

Alec Fraser is a Lecturer in Government and Business at King's Business School, King's College London, UK. His research centers on the use of evidence in policymaking and practice across the public sector. He has researched the development of social impact bonds and outcomes-oriented reforms in public service delivery in the UK and internationally over recent years. Email: alec.fraser@kcl.ac.uk

Lukas Hobi is a Scientific Collaborator at the Bern University of Applied Sciences and a PhD candidate at the University of Bern. His research focus is on active labor market policy, income and asset-based poverty (dynamics), and understanding which factors facilitate an exit from poverty. Email: lukas.hobi@bfh.ch

Wojtek Przepiorka is an Associate Professor in the Department of Sociology at Utrecht University. He uses a diverse set of research methods to study, among others, social norm dynamics, the workings of reputation mechanisms, communication in strategic interactions, intergroup relations, and the interplay of moral norms and institutions. Although aimed at fundamental insights into the problem of social order, Wojtek's research offers many points of contact for practitioners interested in designing institutions for open societies. Email: w.przepiorka@uu.nl

Gemma G. M. Geuke was a Research Intern at Utrecht University and is now a junior researcher at the department of Tobacco Control at the Trimbos Institute. Her research interests include social participation and health behaviors, their underlying causes, and how public policy can influence these behaviors. Email: ggeuke@trimbos.nl

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hevenstone, Debra, Alec Fraser, Lukas Hobi, Wojtek Przepiorka, and Gemma G. M. Geuke. 2023. "The Impact of Social Impact Bond Financing." *Public Administration Review* 83(4): 930–946. https://doi.org/10.1111/puar.13631