Evaluation of the Corporation for Supportive Housing's
Social Innovation Fund Initiative

Final Report
August 2017

Prepared by the
New York University Evaluation Team:
Beth Weitzman, PhD
Tod Mijanovich, PhD
Carolyn Berry, PhD
Maggie Giorgio, PhD MPH
Maggie Paul, MS

Acknowledgements

This publication represents a collaborative effort by researchers at New York University Steinhardt School of Culture, Education and Human Development, College of Global Public Health, and School of Medicine.

This work was funded by the Corporation for Supportive Housing, award #12-042-C

Principal Authors:
Beth C. Weitzman, PhD
Tod Mijanovich, PhD
Carolyn Berry, PhD
Maggie Giorgio, PhD MPH
Maggie Paul, MS

Table of Contents

EXECUTIVE SUMMARY

The Corporation for Supportive Housing (CSH) with support from the Corporation for National Community Service (CNCS) has funded a team of researchers at New York University (NYU) to test a multi-site program model that links a subset of 'super-utilizers' who experience homelessness to affordable supportive housing with medically-focused case management services. This demonstration project is distinguished from the typical model of supportive housing in two ways: 1) the identification of high medical cost homeless individuals through the use of empirical data; and 2) the explicit integration of primary health care into the mix of services clients receive once housed. Once patients were targeted and enrolled, full service delivery thus required four key components: permanent housing; ongoing supportive case management; appropriate mix of mental health and substance abuse treatment services; and primary health care. The demonstration was implemented in four sites, including Connecticut (CT), Los Angeles (LA), Washtenaw County, Michigan (MI), and Francisco (SF). Implementation activities in each site were guided by the theory that when high cost, high risk homeless individuals are identified, housed and receive the case management they need to stabilize and access the other services they need, their physical and mental illnesses will be stabilized such that they will be healthier and incur less cost to the health care system. The evaluation of this demonstration project was also guided by this theory.

While prior evaluations have described the effect of supportive housing on health care utilization, findings are mixed and the studies were often limited to assessing the impact of a program in a single site. This evaluation is distinguished from prior research in that it employs a randomized control trail (RCT) design across multiple, distinct study locations using a relatively large sample size. The evaluation included several key components used to assess program implementation and to build strong evidence of whether and to what extent the program produced impacts as whole and/or across all study sites in order to address gaps in existing research. The evaluation activities, staggered at appropriate intervals throughout the five-year grant period, included visits to all program sites, including interviews with a broad array of stakeholders, a pre/post participant survey to assess changes in self-reported measures (no comparison group), a cost effectiveness analysis, and an impact analysis. Information gathered on site visits was used to track program implementation, including site-specific facilitators and barriers to program implementation. An analysis of the participant survey data was used to assess statistically significant changes in participant responses from pre-test to post-test. The cost effectiveness analysis was performed using data gathered from sites to determine the extent to which cost-savings across multiple programs covered the cost of implementation. Finally, the impact analysis employed a RCT design in CT, MI, and SF[1], allowing for a rigorous assessment of program effects on health care utilization and cost and – to the extent that data were available – on shelter and jail utilization. LA, though included in the survey data analysis, was excluded from the impact analyses as there were no available data to support such analysis. Analyses were conducted using both intent-to-treat (ITT) and treatment-on-the-treated (TOT) approaches to assess changes in utilization as a result of participation in the program, which in this project is tenancy in supportive housing. The control group for the ITT approach was identified during random assignment (full control group used) whereas the comparison group for the TOT

---

[1] LA was not included in the impact analysis, as both the targeting strategy, which did not allow for randomization or the development of a comparison group, and the lack of administrative data precluded rigorous evaluation.

approach was identified using 1:1 pair matching of each housed individual with the control group member who was most similar in terms of their log of total cost, number of hospitalizations, and number of ED visits within a 12-month period prior to randomization.

INTRODUCTION

Program

The Corporation for Supportive Housing (CSH) with support from the Corporation for National Community Service (CNCS) has tested a multi-site program model that links a subset of 'super-utilizers' who experience homelessness to affordable supportive housing with medically-focused case management services. This demonstration project is distinguished from the typical model of supportive housing in two ways: 1) the identification of high medical cost homeless individuals through the use of empirical data; and 2) the explicit integration of primary health care into the mix of services clients receive once housed. Full service delivery thus required four key components: permanent housing; ongoing supportive case management; appropriate mix of mental health and substance abuse treatment services; and access to primary health care.

The grant period for program implementation was a five year period beginning in the summer of 2012 and concluding in 2017. The four sites selected for program implementation included the state of Connecticut (CT), Los Angeles County (LA), Washtenaw County, Michigan (MI), and San Francisco (SF). Each site was expected to enroll and place a minimum number of homeless individuals (SF = 50^2, MI = 100, CT = 150, LA=107) into supportive housing. Although the program in each site was based on the same overarching theory and core components, there were planned differences across sites, notably the type of housing model, the targeting strategy, and staffing model (Table 1).

Table 1. Planned Cross-Site Differences in CSH-SIF Program Implementation

Site	Targeting Strategy	Service Providers	Housing Model	Staffing Model
CT	Match list of highest utilizers	Five agencies initially, down to four before end of first year	All agencies placed clients in scattered site housing	Patient navigators as main service providers, with some support from contracted case managers (eventually hired in-house case managers in some sites)

---

[2] While the SF site committed to house at least 50 individuals, the congregate facility in SF (the Kelly Cullen Community – or KCC) was used to house 172 homeless high utilizers. The first 50 individuals were randomized into housing and the remaining 122 initial slots and all slots resulting from turnover were identified and placed through the San Francisco Department of Public Health's standard operating procedures for housing people on the waiting list for the Direct Access to Housing (DAH) program. Although only the first 50 individuals housed were randomized, all individuals housed in KCC during the study period were asked to take the participant survey.

MI	Match list of highest utilizers	Single agency	Scattered site	Case managers as main service providers
LA	Point-of-care using utilization triage tool	Five agencies initially, down to four before end of first year	All agencies placed clients in scattered-site housing	Case managers as main service providers
SF	Match list of highest utilizers	Single agency	Single, congregate facility	Social workers as main service providers, with in-house, on-site support from a clinical team and a money manager

Three of the four sites employed some type of cross-system data matching procedure between the homeless and health care systems (Medicaid or health/hospital system), at least for part of their caseloads, to generate lists of people who meet some definition of "homeless" and some threshold of health care utilization costs. Program staff then received the "match lists" and attempted to locate and engage potential clients. MI identified high cost patients using data from two of the three hospital systems serving the county and matching those data with shelter use data from Washtenaw County. CT used statewide Medicaid data and the CT Homelessness Management Information System (HMIS) data. For 50 of its 172 slots, SF used data from the San Francisco Health Plan (SFHP), a Medicaid health care plan, and other sources of data on to target clients; however, once individuals were randomized into the 50 spots committed to CSH-SIF, this procedure was no longer followed. LA utilized a point-of-care approach in which individuals were triaged for high utilization and homelessness and recruited into the program at their place of care. While this was a highly effective strategy for locating and engaging clients, it did not allow for a randomized study design.

Two of the sites, LA and CT, used a regional approach to program implementation where they relied on multiple sub-contracted agencies across the state or county, while the programs in SF and MI were implemented by a single organization. CT, MI and LA used scattered-site housing in which clients were placed in available units throughout the catchment area. In SF, all clients were placed into the Kelly Cullen Community (KCC), a single building dedicated to housing homeless high utilizers that was co-located with a federally qualified health center on the ground floor. Staffing models also vary by site. As sites were required by CNCS to meet a one to one match for SIF funding, the SIF funds were not designed to fund the entire service and housing model. Given the diversity of sources of match funds across all sites there were differences in the model as well. The SIF funding supported case managers in LA and MI and these case managers are the core SIF program staff. In SF, on-site social workers from two different agencies worked together with public health nurses and a money manager in a team based approach. Case management services were not explicitly funded through the grant in CT, as the grantees anticipated that other service providers in the community would fill this role through partnership arrangements. The grant funded patient navigators in CT, who were designed to concentrate on health care and were not meant to provide ongoing supportive case management services. However, as the program matured, several sites in CT hired in-house case managers to

support clients. As mentioned above, in all four sites, the program design required engagement of additional services and community partners, including hospitals and health care systems.

In addition to these planned differences, a number of other important and unplanned differences, such as local context (e.g. local attitude and norms around supportive housing for substance users) and state policy (e.g. the extent to which Medicaid covers needs of target population especially in regard to substance abuse and mental health services) shaped program implementation to a high degree. While the implications of these factors are discussed in the findings section of this report, they are described in greater detail in the series of site visit reports produced by the evaluation team throughout the grant period. These reports are included as attachments.

The overarching program theory guiding program implementation across all sites was that when high cost, high risk homeless individuals are identified, housed and receive the case management they need to stabilize and access the healthcare services they need, including integrated primary health care, their physical and mental illnesses will be stabilized and controlled such that they will lead healthier lives and incur less cost to the health care system. The evaluation of this demonstration project was also guided by this theory.

Evaluation

The evaluation timeline mirrored that of the program, developed around the original timeline, was a five-year funding period running through July 2017. While prior evaluations have described the effect of supportive housing on health care utilization, the findings from these studies are mixed and the designs are often limited to assessing the impact of a program in a single site. This evaluation is therefore distinguished from prior research in that it employs a randomized control trail (RCT) design across multiple, distinct study locations using a relatively large sample size.

This study included several key components used to assess both program implementation and impacts across sites. These evaluation activities, which were staggered at appropriate intervals throughout the five-year study period, included a series of three site visits to all program sites, a pre/post participant survey, a cost effectiveness analysis, and an impact analysis.

> **Site visits** were divided into three rounds, with evaluators visiting all sites during each cycle. These visits occurred in the spring of 2013, 2015, and 2016, roughly corresponding to early/pre, intermediate and mature program phases. Site visits were conducted by two members of the evaluation team who interviewed key stakeholders in each site, typically speaking with between 75 and 90 individuals across all four sites in each round. These site visits were the primary source of information used to define the program in each site, and findings from the site visits were used to interpret cross-site differences in survey, cost-effectiveness, and impact analyses. Relevant attachments include: Round 1 site visit memo, round 2 site visit memo, round 3 site visit memo

> **Participant surveys** were the main source of data on key measures that were not available through secondary data. Survey topics included quality of life (including health-

related quality of life), substance use, health care access, mental health and wellbeing, history of incarceration, history of homelessness, and, at post-test, reflections on their experience with the CSH-SIF program. Both pre-and post-surveys were administered to clients across all sites in order to capture change in key metrics over time after client engagement with the program. Survey administration was only attempted among individuals receiving services and therefore there is no comparison group for this set of analyses. The baseline participant survey was launched among the program group shortly after the beginning of the study period, as SF was ready to begin enrolling clients at the start of the program period. The post survey was administered to clients who had been enrolled in the program for at least one year. Relevant attachments include: Final survey report

Tracking sheet data were collected for program administrative purposes. The evaluation team used these data for the sole purpose of tracking entry and exit into the program and housing.

Cost effectiveness analyses were used to compare the cost of implementing the program to the costs off-set from reductions in health care utilization as well as shelter and jail utilization. The NYU evaluation team worked with program teams in each site to determine the annual, per-capita cost of running the program, including all support services as well as rent for housing. As many of the sites experienced substantial and often unexpected start-up costs, all sites chose either year 3 or year 4 of the grant cycle to represent the stable, mature program. These costs were then compared to cost savings associated with identified program impacts in regard to healthcare, shelter use, and jail time.

Impact analyses were used to assess whether and to what extent the demonstration had an effect on health care utilization as well as shelter stays and jail utilization (where data were available). It is worth noting up front that, of the evaluation components, the impact analyses presented the greatest challenges to the evaluation team due to the sites' inability to independently provide the required administrative data. We therefore had to alter our scope of work to assist sites with data acquisition and never acquired appropriate data for LA (excluded from impact analysis). For the three sites in which data was available, we used two approaches – treatment on treated (TOT) and intent to treat (ITT) with propensity score matching – to assess impacts. Relevant attachments include: CSH-SIF face-to-face meeting presentation

Together, these components provided a rich description of the CSH-SIF demonstration that includes everything from a nuanced description of the differences in program implementation to a discussion of overarching findings from the impact analyses. In addition to the attachments referenced above, several reports, papers, and presentations have included findings from multiple components. Many of these are referenced in and attached to this report. While this report provides sufficient detail to understand major evaluation activities, please review all attachments for a full picture of the evaluation of the CSH-SIF demonstration.

METHODS AND ANALYTIC APPROACH

Descriptive Statistics for Survey Sample

The sample used in the analysis of survey data included 238 individuals spread across all four sites who took both a pre- and post-survey. Overall, the average survey response rate for both pre- and post-surveys among eligible participants was 47% with the lowest rate in LA (34%) and the highest in CT (55%). Descriptive statistics for demographics characteristics and housing-sensitive indicators are provided in table 2. As shown in table 2, there were no statistically significant differences in key baseline characteristics between the group with pre and post surveys and those without.

Table 2. Differences in demographic characteristics and housing-sensitive indicators by survey status among housed participants who completed surveys

	All surveys (n=490)	Pre- or post-survey only (n=252)	Both surveys (n=238)	p-value
Age				
Mean(SD)	49.0 (10.7)	49.3 (11.3)	48.6 (10.0)	0.533
range	14-83	14-74	19-83	
Gender				
Male	67.8% (329)	71.8% (178)	63.7% (151)	
Female	30.5% (148)	27.0% (67)	34.2% (81)	0.150
Trans/Queer	1.7% (8)	1.2% (3)	2.1% (5)	
Race				
Black	40.1% (193)	37.7% (92)	42.6% (101)	
Hispanic	12.5% (60)	13.9% (34)	11.0% (26)	0.099
White	41.2% (198)	39.8% (97)	42.6% (101)	
Other	6.2% (30)	8.6% (21)	3.8% (9)	
Months on the street				
Mean(SD)	59.6 (77.4)	61.2 (79.8)	58.1 (75.2)	0.678
range	0-432	0-420	0-432	
Ever in foster care	13.2% (60)	14.5% (33)	11.9% (27)	0.406
Reported sleeping on the street/in a shelter most often during the 12 months prior to being housed	51.0% (237)	50.2% (115)	51.7% (122)	0.750
Number of lifetime arrests				
Mean(SD)	11.4 (21.7)	12.1 (22.3)	10.7 (21.1)	0.529
range	0-200	0-200	0-200	
Ever been treated for drug or alcohol use	60.9% (285)	60.9% (142)	60.8% (143)	0.984
LGBT	10.8% (46)	9.6% (21)	10.8% (25)	0.677
Site				
CT	30.6% (150)	25.8% (65)	35.7% (85)	
LA	20.0% (98)	26.2% (66)	13.5% (32)	0.002
MI	17.1% (84)	15.9% (40)	18.5% (44)	
SF	32.2% (158)	32.1% (81)	32.3% (77)	

Among those who were enrolled and housed, survey data indicates that participants experienced significant health care needs at baseline (Table 3). This finding is reinforced by the baseline data from the impact analysis presented in the following section. As illustrated by Table 3, among those who were surveyed in each site, well over half of the respondents indicated that they were in fair or poor health and had substantial difficulty with activities of daily living, such as dressing or bathing and walking or climbing stairs. Further, the proportion of respondents who indicated that they experienced medical problems every day in the past month ranged from 28% in CT to 41% in LA, 36% in MI, and 45% in SF. Though not reported in this table, the survey also asked about clients' usual source of care. While the vast majority of respondents across sites indicated that they had a usual source of care in the year prior to enrollment, only 15% of those in SF indicated the ED to be their usual source of care, as compared to 30% in CT, 52% in MI, and 39% in LA.

Table 3. Health related quality of life indicators at baseline, overall and by site

	Overall (n=384)	CT (n=131)	LA (n=66)	MI (n=62)	SF (n=125)
Self-rated Health Status*					
Excellent/very good	11% (41)	8% (11)	13% (8)	8% (5)	14% (17)
Good	26% (99)	38% (50)	7% (4)	21% (13)	26% (32)
Fair/poor	63% (237)	53% (69)	80% (48)	71% (44)	61% (76)
Any serious medical problems*	86% (319)	85% (109)	96% (55)	77% (48)	86% (107)
Days in the past month with a medical problem*					
No days	24% (87)	29% (36)	10% (6)	27% (16)	24% (29)
Some days	39% (140)	43% (53)	48% (28)	37% (22)	31% (37)
Everyday	37% (134)	28% (34)	41% (24)	36% (21)	45% (55)
Difficulty walking or climbing stairs*	61% (230)	53% (69)	69% (42)	53% (33)	69% (86)
Difficulty dressing or bathing*	26% (98)	15% (19)	44% (25)	20% (12)	34% (42)

*Significant difference by site at p < 0.05

Descriptive Statistics for Impact Analysis Sample

Descriptive statistics for the randomized treatment and control groups are provided in Table 4. Across the 3 sites included in the impact analysis, a total of 1,512 individuals were randomized to either the treatment or control groups (CT = 907, MI = 407, SF = 199). We randomized more individuals relative to program target numbers, as we expected that staff would not be able to find everybody on the list and, in fact, locating individuals was even more difficult than

hypothesized. (Due to problems with identifying the full sample in subsequent Medicaid datasets, the analytic sample for CT=809). As seen in Table 4, the randomization process resulted in closely matched samples in each site.

Table 4. Baseline characteristics for treatment and control groups by site[1]

	Connecticut*			Washtenaw County, MI**			San Francisco		
	Treatment (n=430)	Control (n=379)	p-value	Treatment (n=242)	Control (n=156)	p-value	Treatment (n=95)	Control (n=94)	p-value
Age Mean (SD)	46.0 (10.9)	46.8 (11.5)	0.404	N/A	N/A	N/A	50.5 (9.9)	50.5 (10.6)	0.951
Gender Male Female	68% 32%	72% 28%	0.336	N/A	N/A	N/A	60% 40%	63% 37%	0.591
Race White Black Hispanic Other	53% 25% 21% 1%	54% 24% 21% 1%	0.978	N/A	N/A	N/A	41% 41% 5% 12%	40% 37% 11% 13%	0.691
Medical hospitalizations Mean (SD)	3.4 (3.6)	3.0 (3.0)	0.112	2.0 (3.0)	1.4 (2.0)	0.025	2.1 (2.4)	2.0 (2.6)	0.684
Total hospital days Mean (SD)	14.0 (26.4)	16.2 (35.1)	0.248	11.4 (23.4)	6.1 (10.8)	0.009	12.8 (25.2)	11.4 (24.9)	0.704
ED visits Mean (SD)	8.9 (11.6)	8.4 (10.9)	0.419	10.3 (9.9)	9.5 (8.1)	0.367	7.6 (8.2)	10.2 (13.3)	0.100
Psychiatric hospitalizations Mean (SD)	0.6 (1.1)	0.5 (0.9)	0.115	0.4 (1.0)	0.6 (1.3)	0.226	0.3 (0.7)	0.4 (1.0)	0.552
Outpatient visits Mean (SD)	40.4 (29.3)	39.7 (34.8)	0.770	6.6 (10.6)	7.0 (12.8)	0.749	8.6 (6.1)	6.6 (5.3)	0.152
Costs of care (dollars) Mean (SD)	61,185 (38,599)	58,272 (54,789)	0.261	29,086 (45,012)	29,436 (104,006)	0.963	30,432 (n/a)	29,616 (n/a)	n/a

*Data for age, gender, and race for CT is based on an earlier analysis with an N=526. We do not expect these numbers to change dramatically with the addition of the remaining cases.
*Note that data on race was not available in the MI dataset

With the exception of medical hospitalizations and hospital days in MI, there were no statistically significant differences in demographic characteristics or healthcare utilization in the year prior to randomization between the treatment and control groups for any of the sites. There were also no statistically significant differences between random assignment groups in average costs for CT or MI in the year prior to randomization between the treatment and control groups. There was, however, a large difference between average costs for each site; MI sample members accrued average costs of approximately $30,000 in the year prior to the intervention, SF about $30,000, and a little more than $60,000 in CT. We expect that this difference reflects the fact that our CT cost data captures all (paid) health system costs regardless of provider or category of utilization (which include all primary and specialty, medications, etc.), whereas our MI cost data

captures utilization at two hospitals only. This difference may also reflect higher illness severity levels in the CT sample. In thinking about these numbers, it is important to remember that the size of the homeless population varies dramatically across these locales. The severity of illness was higher in CT which may be due to the fact that the program in MI was pulling from a smaller pool of eligible individuals, thereby representing a larger share of the homeless population more generally. In SF, where there is a strong safety net, it may be that individuals had a high level of access to care prior to engagement with the CSH-SIF program.

Table 5 displays the burden of specific types of health conditions among the full randomized sample in CT, MI, and SF during the baseline period. The percentages in the table represent the proportion of those randomized in each site who had at least one visit in the baseline period with the selected outcome listed as the primary diagnosis. For example, 41% of randomized individuals in CT had at least 1 visit during the baseline period where alcohol use was listed as the primary diagnosis, compared to 36% in MI and 23% in SF.

Table 5 reveals the deep medical needs of the CSH-SIF study population. The vast majority of those randomized had at minimum of 1 chronic disease diagnosis at baseline (81%-91%). Unsurprisingly, mental health and substance use issues were also common. Approximately half of the individuals targeted for randomization in MI and SF had a mental health visit in the baseline period, compared to 88% in CT. A similar trend was seen for any substance use (alcohol or illicit drug) across the 3 sites (72% in CT, 44% in MI, 42% in SF). While 6 in 10 randomized individuals in CT were "triple diagnosed" (substance use, mental health, and chronic disease), this percentage was lower in MI (18%) and SF (28%).

Differences across the sites also emerged when looking at the burden of chronic disease as measured in this way. While approximately 60% of those randomized in CT and SF were diagnoses with 3 or more different types of chronic disease at baseline, this percentage drops to 28% in MI. Further, we observe that MI has consistently lower rates of individual chronic conditions. These findings are consistent with the far lower average costs of health care at baseline in MI.

Table 5. Alcohol, drug, mental health, and chronic disease diagnoses at baseline, by site and random assignment

	CT (N=809)	MI (N=405)	SF (N=199)
Alcohol use	41%	36%	23%
Illicit drug use	59%	13%	27%
Any substance use (alcohol or drug)	72%	44%	42%
Mental health	88%	49%	53%
Any chronic disease	91%	81%	91%
Alcohol/drug and mental health	65%	22%	29%
Alcohol/drug, mental health, and any chronic disease	60%	18%	28%
Only 1 chronic disease	14%	27%	10%
2 chronic diseases	21%	26%	22%
3+ chronic diseases	57%	28%	60%
Cancer	6%	3%	3%
Liver	19%	6%	16%

Asthma	11%	7%	6%
Other lung	58%	34%	59%
Cardio	51%	33%	57%
Renal	6%	3%	6%
Diabetes	20%	12%	10%
HIV	10%	2%	16%
Arthritis	52%	33%	63%

Time to housing for those in the treatment group varied between sites. Less than half of the treatment group had been successfully located and engaged by program staff six months after randomization across the three sites (Table 6). A low "take-up" rate was expected. As previously mentioned, we anticipated that it would be difficult for teams to locate and engage homeless individuals off of a list and therefore the number of individuals randomized to the treatment group far exceeded program capacity. Indeed, this proved more difficult and took longer than anticipated. There was larger variation at the six month point in terms of actual housing, with 50% (n=47) of the treatment group in SF housed, compared to 15% (n=71) in CT and 17% (n=41) in MI. Average time to housing from random assignment was 4.2 months for SF, 7.0 months for CT, and 9.1 months for MI. By the end of the follow-up period, (18 months post randomization), approximately one third of clients randomized to the treatment group were located and housed in CT (31%, n=148) and MI (33%, n=79), compared to half of the treatment group in SF (50%, n=47). Of note, by the end of the program period, all program slots were filled though not necessarily in time to be included in the evaluation.

Table 6. Treatment group's identification and housing uptake at 6 months and 18 months after random assignment

	Connecticut (n=481)		Washtenaw County, MI (n=242)		San Francisco (n=95)	
	6 months	18 months	6 Months	18 months	6 months	18 months
Percentage located and successfully engaged by program staff	30.8% (n=148)	35.6% (n=171)	40.8% (n=99)	66.1% (n=160)	49.5% (n=47)	49.5% (n=47)
Percentage housed	14.8% (n=71)	30.8% (n=148)	16.9% (n=41)	32.6% (n=79)	49.5% (n=47)	49.5% (n=47)

Data Collection and Measurement

Program Tracking Sheet
Each site completed a site level client tracking sheet which they submitted to CSH on a quarterly basis. This was only used for the purpose of tracking entries and exits from housing.

Participant Survey
Survey data was collected from program participants only. Though we thoroughly explored the possibility, either the sites nor the evaluation team had the resources that would have been required to survey control group members. In CT, MI, and SF, this would have required locating, consenting, and surveying homeless individuals off of the match lists. Furthermore, because we were most interested in the pre/post analysis to assess statistically significant changes over time, we would have had to go through the process of locating and surveying this group of individuals – many of whom likely would not have had reliable contact information – again one year after the first survey. Ultimately, though it would have been useful, this did not seem feasible. At baseline, the survey was administered to program participants by front line program staff at the time of housing. We were not concerned about biased responses because, at this early stage, participants had not yet established a relationship with front line program staff. However, this was not the case with the follow up survey, which directly asked about participants' views of the program including staff members. Therefore, at follow-up, sites utilized individuals outside of the program (e.g. SF used graduate student interns) to administer the survey.

Survey measures included a broad range of variables intended to capture information on measures that we knew we could not get from the administrative data. Topic areas included socio-demographic variables, history of homelessness and engagement with the criminal justice system, health-related behaviors such as smoking and alcohol and drug use, quality of life, mental health, overall health status, and access to medical care.

Cost Effectiveness

During year 4 of the demonstration project, the evaluation team worked with sites to acquire and assess the detailed operating costs of the established program. We requested line item budgets that represented the actual operating costs, including but not restricted to the expenses covered through CSH-SIF grant funding, match sources, as well as other leveraged resources. Our team initially approached the directors of four service provider agencies (Avalon Housing, TNDC, Columbus House in CT and Housing Works in LA), who subsequently designated individuals - the CSH-SIF project manager(s) in SF, MI, and LA and the CFO in CT - to work directly with our team to fulfill the request. In both CT and LA, rather than collecting budgets on all the implementing sites, the NYU team requested budgets from the regional site that best reflected the services to be provided through the program model. The evaluation team reviewed initial budgets and then worked with the contacts at each site to further refine the information, asking clarifying questions and requesting missing information as needed. For example, many sites did not initially include an estimate of overhead costs or administrative staff time and therefore had to estimate these costs in subsequent drafts of the budget. Once a comprehensive budget was in hand, the evaluation team worked to identify common spending buckets across sites, ultimately including: personnel, fringe, client needs/activities, case manager travel expenses (this also includes patient navigator travel expenses where applicable), staff supplies, overhead, other direct costs, and housing related costs. These common categories emerged through careful review of the site-specific budgets and include all of the detailed line item expenses provided by sites. Once estimated, program costs were compared to estimated cost savings from changes in shelter and jail time by site to the extent that data were available.

Impact Analysis

The administrative data used for the impact analysis was collected on treatment and control groups, and data sources varied by site. In each of the three sites included in this study, administrative health care data was collected for all individuals in the treatment and control groups for the year prior to and after random assignment. Administrative data on days in jail were collected from the Washtenaw County Sheriff's Office (MI), and the Course Case Management System (CCMS) database in SF. Homeless Management Information System (HMIS) data on emergency shelter utilization were collected from the Washtenaw County Office of Community and Economic Development (MI), and Nutmeg Consulting (CT). The CT Department of Social Services provided Medicaid claims records of health care utilization. Administrative data from MI was obtained from the two major hospitals serving the area (University of MI and Saint Joseph's), which comprised hospital, emergency department, and outpatient department billing records. Finally, health care utilization data for SF came from the San Francisco Health Plan, which is a Medi-Cal (California Medicaid) managed care provider.

The main outcomes for the impact analyses were measured during the baseline period (the 12 months prior to random assignment) and during the follow-up period (months 6 to 18 after random assignment). The six-month lag for the follow-up period allowed for a period of outreach, recruitment, housing placement, and for the intervention to begin to take effect (i.e. for clients to be housed and to begin receiving case management services). It also allowed the MI and CT sites to "catch up" in their housing participation rates to the much faster housing placements that occurred in SF.

Outcomes included the following:

- Number of medical hospitalizations.
- Number of hospitalizations related to psychological/psychiatric/behavioral factors. These include any hospitalizations with principal diagnoses of psychiatric disease, or drug or alcohol abuse.
- Number of days in hospital.
- Number of emergency department (ED) visits (not including ED visits on the day of a hospital admission, since these ED visits are inconsistently billed).
- Number of non-ED outpatient (OPD) visits.
- Costs of care. In CT, "costs" were total payments by Medicaid. In MI, "costs" were imputed from charges using 0.3 as the cost-to-charge ratio (i.e., costs were considered to be 30% of charges). This ratio is slightly less than each hospital's 2015 CMS-reported Medicare cost-to-charge ratio (which ranged from 0.33 to 0.36), in order to account for the somewhat less generous payments made by Medicaid. Since payment or charge data were unavailable in SF, for some analyses we imputed costs by assigning 2011-12 San Francisco Department of Public Health Medicare reimbursement rates to medical hospitalizations ($2072/day), psychiatric hospitalizations ($627/day), ED visits ($482/visit), and outpatient visits ($199/visit).
- Days in jail
- Nights in emergency shelter

In order to limit the influence of outliers, outcome measures were top-coded at the 95th percentile.

We also used data on medical utilization to create subcategories of clients and care. Supportive housing is generally thought to reduce health care costs and utilization by reducing acute incidents of disease and by allowing chronic disease to be better managed through more stable living and improved access to primary care. Likewise the initial thought was also that some diseases and their care are much less likely to be affected by a change in housing status. Further, the first year of stability in supportive housing may bring about case finding of many serious conditions, previously neglected or undiagnosed. Therefore, using information gathered from our site visits and in consultation with two physicians directly involved in the care and oversight of services for homeless and other vulnerable populations, we identified categories of disease conditions that might be differently affected by supportive housing within our study's time frame. These subgroups include:

- Theorized "Housing-sensitive" conditions. The incidence of, and thus utilization for, these conditions was hypothesized to decrease with the receipt of housing. These conditions include dermatological conditions, physical trauma, injuries, poisonings, and TB.
- Theorized "Care management sensitive" conditions. These are conditions for which emergent utilization (inpatient hospitalization and ED) might be reduced due to the receipt of supportive housing and better care management, even as non-emergent (i.e., OPD) care might increase. Similar to "ambulatory care sensitive" conditions, these include diabetes, asthma, and mental and behavioral health conditions including alcohol and substance abuse.
- "Silent" conditions. These are conditions that can remain relatively asymptomatic for extended periods of time, so we might expect to see that the support and improved care in a supportive housing setting would lead to new diagnoses of these conditions in a recently housed population. Treatment of many such conditions can be costly. These include cancer, renal failure and kidney disease, diabetes, hepatitis and liver disease, cardiovascular disease and hypertension.
- Serious chronic conditions. These are serious conditions that, if already diagnosed in the homeless, are likely to continue driving significant utilization after individuals were housed. These include cancer, dialysis treatment, and neurological conditions like Multiple Sclerosis and Parkinson's. Reduced utilization for such conditions can come only through changes in the pattern of utilization rather than reduced incidence or medical need.

The above conditions were identified by categorizing the principal diagnoses from all utilization events using AHRQ Clinical Classification Software (CCS) categories.

Key Research Questions

While the data collected throughout the course of this evaluation are suitable to address many more questions, the focus of the evaluation was on answering a small number of key questions:

- Is it possible to effectively target and engage the kinds of homeless high utilizers for whom this program was intended and provide them with the type of supportive housing that was thought likely to be effective?
- How does the program vary by site, and what factors influence (both facilitate and impede) program implementation?
- If housing incorporated supports oriented specifically to the needs of homeless high utilizers of the health care system, would we see impacts on health care utilization – that is reductions beyond what likely would have happened <u>even without</u> the program?
- Would these impacts prove sufficient to cover the costs of the program?

Although these questions are relatively straightforward, the answers are quite complex, as detailed in the findings and discussion sections.

Counterfactual Condition

The evaluation team utilized comparison groups for the impact analyses. Those identified for randomization into treatment and control groups were to represent the most medically costly of homeless individuals in their locale. Although the specific data-driven approach of targeting clients differed by sites, each was highly effective in identifying individuals who were homeless and high cost users of health care. Individuals were systematically randomized into treatment and control groups by first sorting individuals in order of total costs or ED utilization in the prior year, then randomly selecting either even or odd rows in the sorted list to be assigned to each group. This method served to create comparable treatment and control groups. Of note, in SF, the individuals included in the RCT represent somewhat less than half of those served by the program. An additional 122 individuals considered medically vulnerable were also included in the program but not in the RCT. Impact analyses are restricted to those in the RCT but the survey data that are presented were obtained from both groups. An *a priori* power analysis for cost impacts was conducted early in the evaluation planning phase. Assuming program and control group n's of 549 each, average costs of $25K in the baseline year with a standard deviation of $25K, a difference-in-differences design with a pre/post correlation of 0.7, a two-sided test with alpha=0.05, and power of 85%, we estimated the minimum detectable effect as a $3,500 change in cost relative to the control group.

Due to the expected difficulties of locating those who had been targeted for treatment and the strategic oversampling to address this challenge, more than half of the treatment group were never enrolled in the program and were unlikely to have received housing in this relatively short time period or with the same degree of medically-oriented support. This less-than-full participation rate limits the power of our analyses by attenuating intent-to-treat impacts toward zero, such that even large declines in utilization among those who received the treatment may not be sufficient to reach statistically significant impact. Therefore, in addition to our RCT analyses of impacts, average treatment effects on the treated (TOT) were also estimated, using 1:1 pair matching of each housed individual with the control group member who was most similar in terms of their log of total cost, number of hospitalizations, and number of ED visits within a 12-

month period prior to randomization.[3] This alternative analytic strategy is intended to help address the limited participation rate. In these analyses, the follow-up period matches the first 12 months housed, rather than months 6-18 after randomization.

FINDINGS

Site Visits

Each of the four participating sites successfully used a data-driven approach to engage a client population who had been homeless and had high levels of health care utilization in the year prior to being targeted for the program. That said, the three sites that used a match list approach for identifying potential clients found the approach to be slow, cumbersome, and resource intense. Using administrative data to identify a list of those at highest cost, assigning them to treatment or control groups, and then finding and engaging those on the treatment list was effective in both getting the "right" people into the program and in allowing for a robust evaluation. However, this approach appears difficult, and perhaps impossible, to sustain without dramatically increased capacity and resources at the site level and such an effort may not be justified and may not be a desirable strategy for all service providers.). However, attention should be made to ensure that those who are, indeed, both homeless and high utilizers are served; the use of clear guidelines is critical in making these determinations.

While each of the sites were able to fill and place the allocated program slots this took far more time than originally anticipated in all sites but for San Francisco. The success of the program across all sites rests, in no small part, on the availability of appropriate and affordable housing units, whether in specially designed facilities (as in SF) or in the private scattered site housing markets. Further, without housing vouchers, placement in the private market is impossible. Breaks in the availability of vouchers (for example, due to the federal sequestration) resulted in huge delays in placement, especially in MI and LA.

It is important to remember that federal policies are further shaped by state policies and are implemented in a local context. Medicaid policies, which vary greatly from state to state, are of considerable importance to understanding variations in this program. For example, In Washtenaw County the overwhelmingly main source of mental health services is the county mental health system; private providers who take Medicaid are extremely rare in the county. But people with a primary diagnosis of substance abuse are excluded from the county mental health

---

[3] Matches were required to be within .2 standard deviations of each of these three quantities, and the selected match was the control group member with the smallest sum of differences in these three standard deviation scores. ZIP models were then estimated for the impacts of supportive housing on the number of hospitalizations and the number of ED visits within 12 months of the date of move-in. OLS models were fitted for impacts on total costs for that same period.

system and must receive services through one of the two substance abuse programs in the county. Respondents felt that there was, moreover, a bias towards making substance abuse the primary diagnosis when there were co-occurring diagnoses of mental illness and substance abuse. Further, the local context shaped the nature of the population served, the kinds of health problems they faced, the housing that was available to these clients and the health services available in the surrounding community. This was not the case in the other sites. Implementing this program was, at first, challenging to MI because they had no prior experience with housing first models. Again, other sites had a "leg up" in this regard. We should expect a good deal of variation across states and localities in adopting or, more accurately, adapting the SIF model.

Participant Survey

Without a comparison group, we don't really know whether and to what degree the observed changes among participants would have occurred without supportive housing; however, the survey data that each site collected as part of the demonstration captured clients' perspectives before housing and after a period of one year in the program. ***The differences in their responses from pre-test to post-test strongly suggest that providing such housing to a medically-frail population greatly improves the quality of their lives and their access to care, even as their significant health problems and conditions remain.***

The data presented represents responses of only the subgroup of the clients for which both a baseline and follow-up survey were available. Yet, as previously discussed, we are reasonably confident that these respondents are a good reflection of all the clients served since those with baseline and follow up surveys look quite similar on a number of key characteristics to the entire client group and to those who only have baseline interviews.

After a year of housing, many clients feel much better about their lives, with noticeable and statistically significant improvements among many of the variables capturing quality of life and mental health (Table 7).

Table 7. Overall pre/post changes in mental health and quality of life indicators (N=238)

	Baseline	Follow-up	Pre/post change	
Bothered by medical problems in past 30 days*				
Not at all	21% (48)	36% (83)	Better	36% (83)
Slightly/moderately	31% (72)	29% (67)	Remained unbothered	13% (31)
Considerably/Extremely	48% (111)	35% (81)	Remained bothered	36% (83)
			Worse	15% (35)
Feelings towards overall life*				
Negative	30% (70)	19% (44)	Better	40% (93)
Mixed	32% (74)	23% (50)	Remained positive	26% (60)
Positive	38% (89)	60% (139)	Remained mixed/negative	18% (41)
			Worse	17% (39)

Frequency of feeling nervous, tense, worried, frustrated, or afraid in past month*				
			Better	35% (80)
Not at all	15% (35)	19% (44)	Remained infrequent	22% (52)
Once/several times a month	34% (78)	45% (105)	Remained	22% (52)
Several times a week/every day	51% (119)	36% (83)	weekly/daily	21% (48)
			Worse	
Frequency of feeling depressed or sad in the past month				
			Better	30% (70)
Not at all	13% (31)	23% (52)	Remained infrequent	30% (70)
Once/several times a month	42% (96)	42% (96)	Remained	24% (55)
Several times a week/every day	45% (104)	36% (83)	weekly/daily	16% (36)
			Worse	
Frequency of feeling lonely in the past month				
			Better	31% (72)
Not at all	22% (51)	30% (70)	Remained infrequent	23% (54)
Once/several times a month	35% (81)	30% (69)	Remained	24% (56)
Several times a week/every day	44% (102)	41% (95)	weekly/daily	22% (52)
			Worse	
Feels life is organized*	45% (100)	69% (154)	Better	30% (68)
			Remained organized	38% (86)
			Remained unorganized	25% (56)
			Worse	6% (14)
Feels life is stable*	54% (123)	80% (184)	Better	36% (83)
			Remained stable	44% (101)
			Remained unstable	10% (24)
			Worse	9% (22)

*Significant difference, baseline to follow-up, at $p < 0.05$

After a year of housing, there were statistically significant improvements across all access to care measures (Table 8).

Table 8. Overall pre/post changes in access to healthcare services (N=238)

	Baseline	Follow-up	Pre/post change	
Usual source of care*				
No usual source of care	5% (12)	8% (18)	Better	24% (55)
ED usual source of care	31% (73)	16% (37)	Continued medical	53% (125)
Medical office is usual source of care	64% (149)	77% (179)	Continued ED/no usual	10% (23)
			Worse	13% (31)
Difficulty finding a doctor in past 12 months*	23% (52)	13% (31)	New access	16% (36)
			Kept access	71% (163)
			Continued problems	7% (16)
			Lost access	6% (15)
Needed a dentist in past 12 months but unable to see one*	56% (129)	35% (81)	New access	30% (68)
			Kept access	35% (81)
			Continued no access	27% (61)
			Lost access	9% (20)

*Significant difference, baseline to follow-up, at $p < 0.05$

From pre-test to post-test, the majority of clients held constant in terms of activities of daily living – that is, there were relatively small positive changes in general but, also, few reported being worse off (Table 9).

Table 9. Overall pre/post changes in health status and related conditions (N=238)

	Baseline	Follow-up	Pre/post change	
Self-rated Health				
Excellent/Very Good	10% (24)	15% (35)	Better	21% (50)
Good	27% (65)	22% (53)	Stayed excellent/v.g./good	14% (33)
Fair/Poor	63% (149)	63% (149)	Stayed fair/poor	48% (114)
			Worse	17% (40)
Number of not good physical health days in past 30				
0	25% (56)	31% (69)	Better	32% (72)
1-14	28% (63)	29% (65)	Remained less than every day	26% (59)
15-29	11% (25)	7% (16)		
30	36% (80)	33% (74)	Remained every day	18% (41)
			Worse	23% (52)
Mean (SD)	14.0 (13.1)	12.1 (13.2)		
Any serious medical problems	85% (199)	81% (190)	Better	10% (23)
			Continued no problems	9% (21)
			Continued medical problems	75% (176)
			Worse	6% (14)
Difficulty walking/climbing stairs	62% (144)	60% (138)	Better	13% (30)
			Continued no difficulty	28% (65)
			Continued difficulties	49% (114)
			Worse	10% (24)
Difficulty dressing/bathing	25% (59)	24% (55)	Better	12% (28)
			Continued no difficulty	65% (151)
			Continued difficulties	13% (31)
			Worse	10% (24)

*Significant differences, baseline to followup, at p < 0.05

As with client health status and conditions, most people reported stable substance use over the course of their engagement with the program, with a relatively small proportion of clients stopping or starting during their time in the program (Table 10). The exception to this was for illicit, non-prescription drugs, for which use declined by nearly 10% over the course of the study period.

Table 10. Overall pre/post changes in substance use (N=238)

	Baseline	Follow-up	Pre/post change	
Used alcohol regularly in the past year	39% (91)	33% (77)	Stopped	13% (29)
			Remained non-drinker	54% (126)
			Remained drinker	27% (62)
			Started drinking	6% (15)
Used marijuana regularly in the past year	20% (47)	25% (58)	Stopped	6% (13)
			Remained non-user	69% (160)

			Remained user	15% (34)
			Started using	10% (24)
Used other illegal or non-prescribed drugs regularly in the past year*	23% (52)	14% (33)	Stopped	15% (34)
			Remained non-user	71% (162)
			Remained user	8% (18)
			Started using	7% (15)
Current smoker	67% (158)	70% (166)	Stopped	1% (3)
			Remained non-user	28% (67)
			Remained user	66% (155)
			Started using	5% (11)

*Significant difference, baseline to followup, at p < 0.05

Clients appreciated the program, reporting that the program largely met all or most of their needs and that their quality of life improved since entering the program (Table 11)

Table 11. Satisfaction with the SIF program overall

	Overall (n=329)	CT (n=102)	LA (n=54)	MI (n=62)	SF (n=111)
SIF program has met all or most of my needs*	90% (287)	96% (98)	94% (50)	95% (55)	80% (84)
Would recommend the SIF program to a friend*	95% (294)	98% (95)	94% (50)	100% (58)	89% (91)
Would come back to the SIF program if you were seeking help again*	94% (297)	97% (99)	98% (52)	100% (58)	85% (89)
My quality of life has improved since entering the program*	89% (293)	96% (98)	93% (50)	93% (57)	79% (88)

*Significant difference across sites at p < 0.05

While clients reported improvements in general wellbeing and access to care, and were highly satisfied with the CSH-SIF program, the majority of respondents reported the continued existence of serious medical conditions one year into the program. This is in line with the high number of chronic conditions present at baseline. So, while tenants may have better care their underlining chronic conditions persist. Moreover, habits around substance use improved for relatively few clients.

Impact Analysis

Program Impacts on Utilization Categories

Table 12 presents changes in health care utilization among treatment and control groups in the year following random assignment. Substantial reductions are observed among both groups. This finding is not surprising given the cyclical nature of health care utilization and it highlights the need to evaluate such programs using a control or strong comparison group. In the pooled analyses, we observed negative point impacts in the number of medical hospitalizations,

psychiatric hospitalizations, and outpatient visits for the treatment group. However, these differences were not statistically significant.

Table 12. Program impacts for the period 6-18 months after random assignment

Outcome	Site	Predicted number of events in the post-period		Program effect	
		Treatment (95% CI)	Control (95% CI)	Difference (95% CI)	p-value
Medical hospitalizations	Pooled	1.64 (1.45, 1.82)	1.62 (1.42, 1.81)	0.92 (-0.24, 0.27)	0.890
	CT	2.11 (1.83, 2.40)	2.00 (1.71, 2.29)	0.11 (-0.28, 0.50)	0.571
	MI	1.10 (0.88, 1.32)	0.97 (0.67, 1.26)	0.13 (-0.21, 0.47)	0.452
	SF	0.73 (0.47, 0.99)	1.36 (0.97, 1.74)	-0.62 (-1.08, -0.17)	0.006
Total hospital days	Pooled	7.71 (6.36, 9.06)	7.92 (6.47, 9.37)	-0.21 (-2.14, 1.73)	0.833
	CT	10.00 (7.80, 12.21)	9.11 (7.00, 11.23)	0.89 (-2.09, 3.87)	0.559
	MI	4.71 (3.65, 5.76)	4.76 (2.63, 6.89)	-0.05 (-2.36, 2.26)	0.966
	SF	4.25 (2.00, 6.49)	9.47 (5.97, 12.97)	-5.22 (-9.33, -1.12)	0.013
ED visits	Pooled	4.78 (4.38, 5.19)	4.44 (3.99, 4.88)	0.34 (-0.22, 0.91)	0.232
	CT	4.79 (4.27, 5.30)	4.23 (3.72, 4.75)	0.55 (-0.11, 1.22)	0.104
	MI	4.88 (4.16, 5.60)	4.53 (3.57, 5.49)	0.35 (-0.80, 1.50)	0.551
	SF	4.57 (3.18, 5.96)	5.13 (3.79, 6.48)	-0.56 (-2.38, 1.25)	0.418
Psychiatric hospitalizations	Pooled	0.27 (0.22, 0.32)	0.29 (0.23, 0.35)	-0.02 (-0.09, 0.06)	0.667
	CT	0.43 (0.34, 0.53)	0.44 (0.33, 0.55)	-0.01 (-0.15, 0.13)	0.899
	MI	0.12 (0.08, 0.16)	0.12 (0.07, 0.17)	0.00 (-0.06, 0.06)	0.945
	SF	0.13 (0.03, 0.22)	0.20 (0.09, 0.31)	0.08 (-0.22, 0.07)	0.308
Outpatient visits	Pooled	20.82 (19.23, 22.41)	21.11 (19.26, 22.96)	-0.56 (-4.02, 2.90)	0.782
	CT	32.57 (30.20, 34.95)	33.14 (30.26, 36.01)	-3.78 (-9.24, 1.69)	0.751
	MI	4.45 (3.73, 5.17)	3.67 (2.94, 4.39)	0.79 (-0.11, 1.68)	0.087
	SF	5.11 (3.18, 7.04)	6.50 (4.91, 8.10)	-1.39 (-3.77, 0.99)	0.108

Costs	Pooled	$36,782 ($32,775, $40,788)	$39,046 ($30,763, $47,329)	-$2,265 (-$11,465, $6,936)	0.630
	CT	$44,752 ($39,352, $50,153)	$45,610 ($37,906, $53,313)	-$858 (-$10,265, $8,550)	0.858
	MI	$20,620 ($15,401, $25,839)	$25,738 ($6,121, $45,356)	-$5,118 (-$25,418, $15,181)	0.621
	SF	$16,749	$28,405	-$11,656	n/a

At the site level, we observed statistically significant impacts for both the number of hospitalizations (-0.62, p=0.006) and total number of hospital days (-5.22 per person, p=0.013) for SF. In the other sites and in regard to other measures of utilization, the treatment group typically experienced greater reductions than those in the control group, but these differences are not statistically significant. It is worth recalling that SF was able to locate and house approximately 50% of those in the targeted treatment group, as compared to only one-third in CT and MI. We will return to this in our discussion of findings from the TOT analysis.

Program Impacts on Medical Costs

As previously noted, average costs in the year before random assignment were about $60,000 in CT and about $30,000 in MI. (Our imputed estimate of baseline costs in SF was approximately $30,000). As can be seen in Table 13, costs for both the treatment and control groups in CT and MI fell in the follow-up period. Further, while the point estimate of the overall cost impact in each site was negative, representing cost savings, neither estimate in CT or MI was statistically significant (-$5,049, p=0.331 in CT, -$2,394, p=0.422 in MI). In SF, the program impact on imputed costs in the follow-up period was -$11,656. This larger cost decrease in SF reflects the larger estimated reduction in hospitalizations in that site.

Table 13 presents findings regarding the relationship between program impacts and baseline costs in MI and CT, the only sites in which it was possible to do so. Cost impacts varied with the level of costs incurred prior to program enrollment, as evidenced by statistically significant interactions between pre-period costs and the treatment indicator in our models. Figures 1 and 2 are visual representations of this relationship. In each site, lower costs prior to enrollment were associated with positive and statistically significant cost impacts. In other words, those high utilizers who were in the lower range of cost at program entry saw net increases in health care costs after engaging in the program. In contrast, those most expensive at baseline randomized to treatment saw net decreases in cost relative to the controls; this was not, however, statistically significant. Of course, the relatively small number of cases within the 90th percentile and the relatively high degree of variance for health care costs make it difficult to achieve statistical significance. We will also explore these findings by adjusting, in future analyses, for individuals who were no longer eligible for Medicaid and/or no longer lived in the region. Such adjustments were not possible within the grant period.

Table 13. Program impacts on cost of care in CT and MI for the period 6-18 months after random assignment, at selected percentiles of pre-period costs

Site	Program impact on costs at the …	Average costs		Program effect	
		Treatment (95% CI)	Control (95% CI)	Difference (95% CI)	p-value
CT	… first quartile of pre-period costs	$33,920 ($29,746, $38,095)	$27,129 ($23,548, $30,710)	$6,791 ($1,291, $12,291)	0.016
	… third quartile of pre-period costs	$49,176 ($43,838, $54,514)	$49,171 ($43,065, $55,277)	$5 (-$8,104, $8,115)	0.999
	… 90th percentile of pre-period costs	$65,094 ($53,808, $76,381)	$77,041 ($60,853, $93,228)	-$11,946 (-$31,680, $7,788)	0.235
MI	… first quartile of pre-period costs	$10,154 ($7,901, $12,407)	$5,768 ($4,364 $7,173)	$4,386 ($1,730 $7,041)	0.001
	… third quartile of pre-period costs	$19,673 ($15,532, $23,814)	$19,442 ($13,673, $25,212)	-$231 (-$6,871, $7,332)	0.949
	… 90th percentile of pre-period costs	$36,669 ($26,002, $47,336)	$43,858 ($28,697, $59,019)	-$7,188 (-$25,726, $11,349)	0.447

Results are predicted values derived from site-specific generalized linear models (GLMs) of post-period costs regressed on pre-period costs, a program group indicator, and their interaction. GLMs utilized a gamma distribution for each site, a log link for CT, and an identity link for MI. All terms in both site-specific regressions were statistically significant at the .01 level.
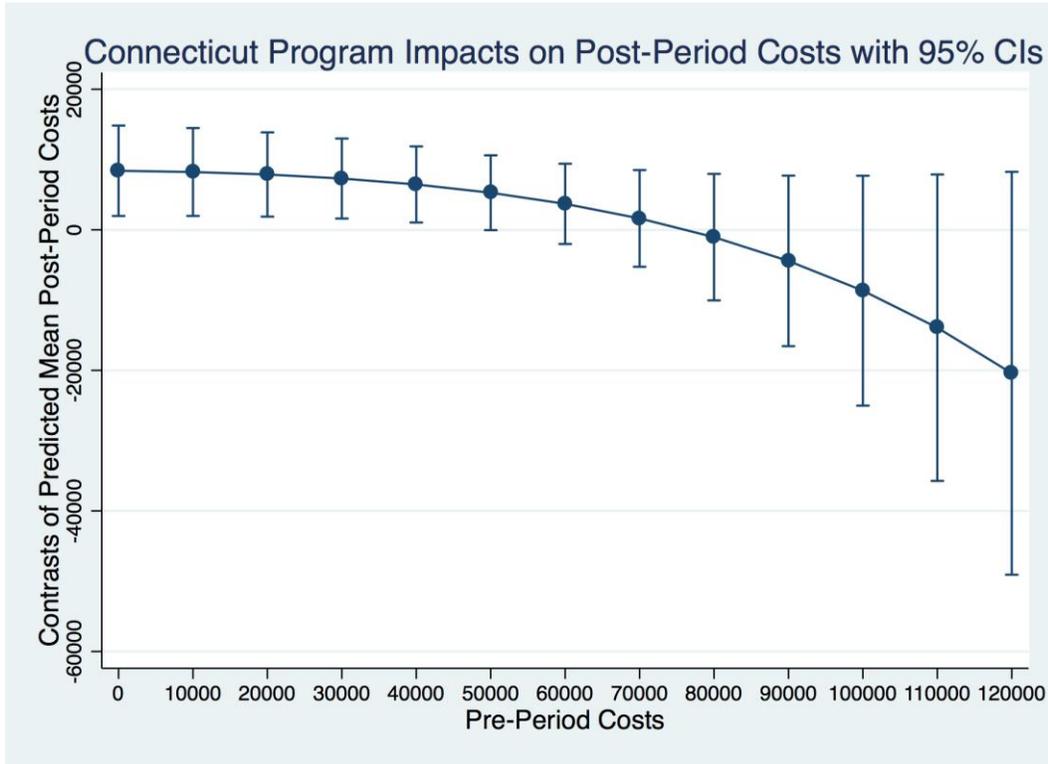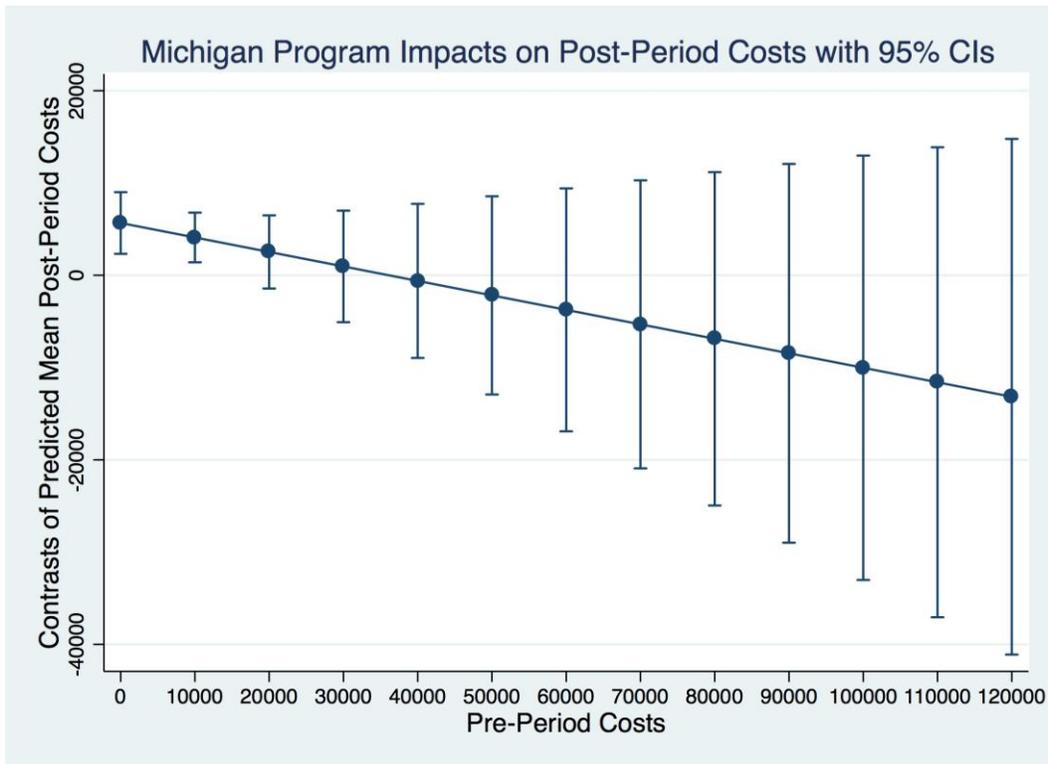
Figure 1.

Figure 2.

Program Impacts by Disease Categories

We found no statistically significant impacts in regard to the theorized "housing sensitive conditions" (e.g., trauma), which were anticipated to have lowered incidence among those housed. While the medical needs of those with serious and expensive chronic conditions, like renal failure, cannot be eliminated through the program and will remain costly regardless of housing status, better management of care might be expected to result in some shifts in utilization patterns. While only relatively small proportions of the study sample entered the program with these conditions – 7% in MI, 16% in CT, and 34% in SF -- we do find some statistically significant impacts for them. SF program group members with serious chronic conditions had reduced hospitalizations and outpatient visits (both statistically significant). CT program participants also had somewhat reduced costs (though not significantly so). MI program group members, however, had post-intervention period utilization and costs very similar to those of control group members.

Program Impacts among the Treated

Impacts of the treatment on the treated are presented in Table 14. Using the matching procedure previously described, 159 (99%) of 161 program group members who were housed early enough to observe at least 12 months of follow up were successfully matched. Very high post-matching balance was achieved: mean program group values were within .25% of the mean comparison group values for pre-period ED visits, within .36% for the log of total pre-period costs, and within .7% for the number of hospitalizations. In CT, program impacts for housed participants were a significant -.53 hospitalizations (p=0.041), a marginally significant -1.0 ED visit (p=0.071), and a non-significant -$3,539 in costs (p=0.4812) in the 12-month follow-up period. In SF, we observed a significant decrease of 0.56 medical hospitalizations (p=0.023), and a statistically significant reduction in ED visits (-2.47, p<0.001). We did not observe statistically significant changes in utilization or cost in MI.

Table 14. Impacts of the treatment on the treated

Outcome	Site	Predicted number of events in the post-period		Program effect	
		Housed (95% CI)	Comparison (95% CI)	Difference (95% CI)	p-value
Medical hospitalizations	CT	1.49 (1.12, 1.86)	2.02 (1.60, 2.45)	-0.53 (-1.05, -0.02)	0.041
	MI	0.92 (0.62, 1.21)	0.79 (0.52, 1.05)	0.13 (-0.25, 0.51)	0.495
	SF	0.55 (0.23, 0.86)	1.11 (0.75, 1.47)	-0.56 (-1.04, -0.08)	0.023
Outpatient visits	CT	27.96 (24.56, 31.36)	31.79 (27.97, 35.60)	-3.83 (-8.36, 0.70)	0.098
	MI	4.33 (3.33, 5.32)	3.84 (2.90, 4.78)	0.49 (-0.76, 1.74)	0.440
	SF	4.75 (3.44, 6.07)	5.76 (4.31, 7.21)	-1.01 (-2.88, 0.87)	0.293

ED visits	CT	4.23 (3.37, 5.09)	5.25 (4.27, 6.23)	-1.02 (-2.14, 0.09)	0.071
	MI	4.88 (3.79, 5.97)	4.95 (3.84, 6.07)	-0.07 (-1.59, 1.45)	0.927
	SF	2.41 (1.52, 3.30)	4.88 (3.55, 6.21)	-2.47 (-3.90, -1.04)	<0.001
Total Costs	CT	$36,329 ($29,555, $43,104)	$39,868 ($32,722, $47,015)	-$3,539 (-$13,386, $6,308)	0.481
	MI	$14,499 ($10,998 $18,001)	$12,184 ($9,011, $15,356)	$2,315 (-$2,197, $6,829)	0.313

Impacts on other Systems

Although a full cost-benefit analysis was beyond the scope of our study, one goal of the evaluation was to assess the degree to which impacts on health care utilization – especially hospitalizations – as well as shelter and jail utilization offset the cost of running the CSH-SIF program. We were able to estimate the impact of the program on shelter and jail utilization in SF and shelter utilization in CT (jail data not available). Due to a lack of available data, we were not able to assess cost effectiveness of the program in LA and the MI analysis was limited to program and health care and shelter costs (no jail data available). Table 15 presents findings from this analysis. While there were no statistically significant impacts on jail time in SF, we found statistically significant impacts on shelter use in SF (ITT and TOT analyses) as well as in CT (TOT only). The program did not affect shelter use in MI though, notably, there was a low level of shelter use among the target population at baseline – 55% of the randomized sample had no recorded shelter use – leaving less room for improvement relative to the other sites. This is also why the ITT estimated average number of shelter days in the post period looks so low for MI. Data on jail use for MI and CT were not available in time for this report.

Table 15. Program impacts on shelter days*

Outcome	Site	Predicted number of events in the post-period		Program effect	
		Treatment (95% CI)	Comparison (95% CI)	Difference (95% CI)	p-value
ITT	SF	7.6 (2.3, 12.9)	18.3 (10.5, 26.2)	-10.7 (-19.9, -1.4)	0.023
	CT	11.6 (8.4, 14.8)	15.7 (11.3, 20.2)	-4.2 (-9.6, 1.3)	0.136
	MI	0.9 (0.6, 1.3)	0.7 (0.3, 1.1)	0.2 (-0.3, 0.8)	0.398
TOT	SF	12.8 (-1.5, 27.1)	42.2 (18.6, 65.8)	-29.4 (-54.7, 4.1)	0.023
	CT	2.4	9.4	-7.0	0.009

		(-0.02, 4.8)	(4.7, 14.1)	(-12.3, -1.7)	
	MI	1.9 (-4.1, 7.8)	8.0 (1.3, 14.7)	-6.1 (-15.1, 2.8)	0.180

*Shelter days top-coded at 95th percentile within site and time period

Cost Effectiveness

The cost effectiveness analysis was performed to compare impacts on health care and shelter systems to the overall cost of running the program. Table 16 presents program costs by site. It is clear from this analysis that the per-capita costs of running the CSH-SIF program vary by site in each category of spending. Of course, cost of living varies substantially across these sites. The total cost of support services category captures spending on personnel, fringe, client needs and activities, staff supplies, overhead, and other direct costs. The per-capita cost of support services was quite similar overall in CT, LA, and SF, whereas this expense was substantially lower in MI. The lower cost of support services in MI was driven by low expenditures on personnel and client needs and activities compared to the other sites. Notably, MI spent only $18 per capita on other than core client needs and activities compared to several hundreds of dollars in LA and SF and more than a thousand dollars per capita in CT for needs outside of rent and core health care connection or navigation. Travel expenses for case managers also varied by site and, as expected, were quite low in SF where clients are housed in a single building.

Table 16. Per-capita Program Costs by Site

Site:	Columbus House	Housing Works	Avalon Housing	Kelly Cullen Community
Location:	New Haven, CT	Los Angeles, CA	Washtenaw County, MI	San Francisco, CA
Year:	2015	July 2015 - June 2016	2016	2015
Enrollment:	53	39	85	172
Personnel	$3,859	$4,923	$3,593	$5,212
Fringe	$1,266	$1,329	$1,074	$1,407
Client Needs and Activities	$1,278	$762	$18	$566
Case Manager Travel Expenses	$418	$154	$232	$7
Staff Supplies	$51	$38	$174	$66
Overhead	$780	$154	$52	$69
Other Direct Costs	$0	$0	$22	$60
Total Cost of Support Services	$7,653	$7,360	$5,164	$7,387
Housing Related Costs	$12,936	$14,556	$10,104	$5,205

Total Per-Capita Program Cost	$20,589	$21,916	$15,268	$12,592

These cross-site differences in costs are also consistent with our findings from the site visits. Although the broad strokes of the program model were consistent across sites, implementation varied in terms of many critical pieces, including organization and management, as well as the type of housing and nature and intensity of support services. These differences are reflected in how monies were spent in each site.

- In CT, Rental Assistance Program housing vouchers were obtained from the state and readily allowed for placement into scattered site units in the private housing market. Patient Navigators supported clients by engaging them in the program and connecting them to appropriate health and social services over time. The grant funding did not include other case management services and the Patient Navigators played this role, as well, in sites that were unable to contract such services.

- The program in LA relied heavily on case managers, who were responsible for a broad range of services with respect to clients, from initial engagement in the program, to securing housing, and connecting clients to appropriate care. As in CT and MI, clients relied on housing vouchers and were largely housed in scattered site units maintained by private landlords. In contrast to the other sites, participant eligibility was determined at the point of care at any one of several participating hospitals whereas case managers/Patient Navigators were responsible for this time-consuming task in the other sites.

- In MI, the initiative was undertaken by a consortium of non-profit and government agencies representing housing, health, and social services. Section 8, Shelter Plus Care, and Housing Choice housing vouchers were used to place clients in scattered site units in the private housing market. Service coordination and case management was achieved through regular and ongoing meeting of a multidisciplinary care team that cut across participating agencies. Case management services were intended to provide social and mental health support, as well as assistance in getting linked to health insurance and primary care; case managers typically met with each client several times per week.

- SF differed from the other sites in that clients were housed in a brand-new congregate housing facility with a federally qualified health center (FQHC) located on the ground floor built specifically for the purpose of providing supportive housing to homeless high utilizers. This facility, named the KCC supportive housing site, was a joint effort for SF's Department of Public Health and the Tenderloin Neighborhood Development Corporation. The on-site services team included KCC property management staff, one SIF project coordinator, five social workers, one nurse, one health worker, and one money manager.

While sites varied in terms of organizational experience with this kind of program prior to SIF, all sites felt that year 3 or 4 represented their established program. Prior to that, the sites were

focused on creating the program and making adaptations to "on-the-ground" conditions. This supports the notion that it takes service providers, even those with deep experience doing this kind of work with this population, a relatively long time to become fully operational after the beginning of the program implementation period. In comparing the impact of the program on health care and shelter utilization to the cost of the program, we see that the program costs, while not fully covered, are substantially offset by the cost savings stemming from reduced utilization of health care and shelter systems (Table 17). The majority of these savings come from reductions in health care utilization. Note that the cost analysis does not include outpatient visit impacts, since outpatient visit payments are reimbursed at such low rates that these impacts represent very small dollar amounts.

Table 17. Per-capita cost-effectiveness estimates in CT and SF based on TOT analysis

	CT	MI	SF
Support services	$7,653	$5,164	$7,387
Housing	$12,936	$10,104	$5,205
Total program cost	**$20,589**	**$15,268**	**$12,592**
Health care (TOT)*	-$5,487 (-.53 hosp.*$8,906)+ (-1.0 ED*$767)	No impact	-$6,682 (-.56 hosp*$8,906)+ (-2.47 ED*$767)
Shelter (TOT)	- $385 (-7.0 days x $55/night)	No impact	-$1,000 (-29.4 days x $34/night)
Total cost savings	**-$5,827**	**N/A**	**-$7,682**
Estimated net cost	**$14,762**	**$15,268**	**$4,910**

Note: health care cost data represents national averages – hospitalization cost calculated from 2014 HCUP data ($8,906) and ED cost provided by PEW ($767)

DISCUSSION

The assumptions underlying the overarching program theory did not hold in all sites. While all sites were eventually able, some only with strong supports from CSH and the NYU evaluation team, to implement a data-based targeting strategy, many sites subsequently faced challenges early on in terms of finding and engaging clients from the match lists. All sites abandoned the match list approach after the end of the study period, either because they did not have the resources to continue with that approach or because they preferred an alternate approach. So, while it seems that match-list targeting is possible within the context of a relatively resource-rich demonstration project, it is unlikely that organizations would be able to sustain such a model independent of these resources and research requirements. That said, LA used a data-driven approach to identify those who matched target criteria; doing this identification at the point of service was far more readily accomplished than by using a match list based on administrative records and, yet, ensured enrollment of those intended. Further, while obvious, it is worth mentioning that housing needs to be available to the target population once they are located and engaged. While staff were able to find housing for the large majority of those engaged, the lack of affordable housing, even when a housing subsidy is available, was and remains problematic in all sites but SF.

This study is one of the first to rigorously investigate whether supportive housing has a significant impact on healthcare utilization and cost among homeless adults who are medical high-flyers. In the strict ITT analysis, impacts were somewhat elusive. Impacts on utilization were primarily found in the SF site. Importantly, we also find significant heterogeneity among the "high flyer" homeless individuals in our sample in regard to cost. While impact on cost was found in the overall models, for those who use fewer healthcare services while homeless, the receipt of supportive housing appears to have increased utilization, suggesting significant unmet needs.

Our inability to find more robust impacts using the ITT analysis may partially be a reflection of several factors that made implementing an RCT design to address this specific research question more difficult. First, in many multi-site initiatives where data are pooled, the evaluators put great emphasis on model fidelity and comparability across sites in regard to both the program and the clients served. Yet, as is the case with so much federal policy, the delivery of uniform service to a uniform population is unrealistic; federal policies are implemented within the context of underlying conditions and relevant state and local policies. Housing conditions, housing supports, community based services, health care providers, and, most especially, Medicaid policies, varied enormously across locations. This shaped the specific services that could and could not be readily provided and who was and was not eligible for program services. Tailoring a national model to local context, as was done in this initiative, is a necessity but it makes us more cautious in how we use and interpret our analyses using pooled data.

Analyzing impact by site and subgroup significantly reduced our statistical power. While our samples may be small from the standpoint of research, however, the reality is that it was very difficult for sites to identify and serve large numbers of participants in the desired time frame. Locating and engaging those eligible was made even more difficult by the use of a data based approach to identify and target those eligible and by a data matching procedure that under girded the RCT approach. Further, in all sites but SF (which had a congregate facility), external factors such as the 2013 United States Budget Sequestration[4] made securing housing vouchers, finding safe and appropriate housing, and ensuring secure placements all made the "small" size of the intervention seem rather large and difficult. For these reasons, we believe it may be unrealistic to imagine an experiment for this population with substantially more individuals in any one setting. Finding ways to make the most of demonstrations of this size is essential if we are to bring evidence to bear on critical policy decisions.

Additionally, while we know whether or not those assigned to treatment actually received treatment (that is, they were enrolled in the SIF program), we do not know whether those in our control group received other such housing. Although information from each site suggested that relatively little "treatment contamination" occurred via housing of control group members, to the degree that this did occur, our treatment contrast was reduced, and our findings represent conservative estimates of program impacts. To this point, our quasi-experimental analysis did reveal significant impacts in CT for medical hospitalizations, OPD use, and shelter stays, even as the ITT analysis did not.

---

[4] In 2013, the United States Budget Sequestration resulted in a decrease in the number of federally funded rental assistance vouchers available to the SIF projects in both Los Angeles and MI.

This study also has limitations beyond the difficultly of investigating the impacts of supportive housing in heterogeneous settings using an RCT design. We were constrained, as have been other researchers, by available data. As already noted, our data in MI did not represent all health care utilization but, rather, data for the two largest healthcare systems in Washtenaw County. In SF, we had utilization but not cost data. These gaps are not peculiar to our evaluation. Rather, they represent the fact that administrative data, gathered locally for purposes other than research, is an imperfect match to research needs.

Further, we believe that an 18-month post-random assignment time horizon (which, in fact, typically represented about 12 months in supportive housing) may be insufficient to test the impacts of supportive housing on health care utilization and costs. Case managers reported that a substantial part of their time with clients in the first year was spent finding and settling clients into housing and then stabilizing them. Stabilization often involved focused attention on furnishing the apartments, linking to substance use services, setting boundaries with visitors, and attending to various financial issues. Only once stabilized could they turn their attention to linking clients to needed and appropriate primary and specialty health care services. Indeed, this is one important way in which SF differed from the other two sites and might explain that site's stronger impacts in the RCT; in SF, clients could more rapidly be settled into their new homes and case managers could focus on other client needs. However, preliminary analyses of our MI data (not presented here) suggest that we may not find impacts on costs even with a longer time horizon.

Finally, in examining the cost of the program relative to statistically significant reductions in health care and shelter utilization, we observe that – in CT and SF – a substantial portion of the cost of running the CSH-SIF program was offset by cost savings across these two systems. The majority of these cost savings stem from reductions if the cost incurred by health care systems, indicating that the health care sector may indeed be an appropriate source of funding for medically-oriented supportive housing that serves a high-flier population. That the overall per-capita program cost in SF is substantially lower than CT may reflect an economy of scale, as the SF program was able to serve more individuals than the other sites.

CONCLUSIONS

All in all the SIF initiative and evaluation allowed for new and significant contributions to the field on the impact supportive housing can have on health care access, utilization and costs for a high-utilizer populations.   Based on our analysis, the following key takeaways can be gleaned from the evaluation:

- It is possible to develop and deliver a medically-oriented supportive housing program targeted at homeless individuals who are high utilizers of health care using a data-driven approach, but it is very difficult in practice.
- Program implementation and capacity for impact are both heavily influenced by local context and state policy.
- This program can reduce utilization of shelters and costly health care in *some* populations, and these reductions can substantially offset program costs.

- While, on average, the program was associated with reduced costs and utilization, in some sites, and improvements in self-reported quality of life and access to care across sites, many participants were still experiencing deep and complex health problems one year into the program.

The relationships between housing, health, and health care utilization is shaped by both individual and community conditions such as different and changing medical needs and conditions and services available in the community, and placing individuals who are homeless and high utilizers of the health care system in supportive housing may not be accompanied by immediate or dramatic drops in costly utilization within a 12 month time period. Yet, we think the evidence we present gives reasons for some optimism and further exploration.
While not necessarily recommended in practice, our findings suggest that it may not be possible to achieve reductions in costly utilization within a 12 month window among all groups in this very sick population. This is in line with suggestions in some prior work, that supportive housing may do little to remedy those at the extremes, where health care needs have become deep and complex and not likely remedied through housing and improved primary care access. (Weinberger, Oddone, & Henderson, 1996; Williams, 2015) Indeed, in our site visits, case managers and other staff have frequently expressed their belief many of the clients are too sick to fully benefit from the program model.

Findings from our analysis of medical conditions across program sites provide further insight into cross-site differences in impacts. MI is a much smaller community compared to SF and the many sites in CT and, as such, has relatively fewer homeless people. The CSH-SIF initiative therefore took a deeper "bite" into the homeless population – that is, it appears that the MI target population included individuals who were relatively healthy compared to individuals in CT and SF given that there was a smaller number of very sick people. Therefore, many people included as "high users" were sick but, overall, less so than those in SF and CT. This is evidenced by the average cost at baseline and is reinforced by the percentage with 2 or more chronic conditions at baseline, as well as the presence of some of the most serious conditions. Impacts on health care costs may therefore be less likely under these circumstances.

Our main impact findings – reductions in utilization among the SF population – suggest to us that, indeed, savings from supportive housing could be achieved under the right conditions for the right people. But our findings also suggest there is much still-needed care among those who are homeless and frequent users of health care, since our least costly program participants experienced higher costs relative to those who were untreated. The hope of cost savings should not obscure the fact that this type of intervention may help people with deep medical needs receive needed and more appropriate care to meet those needs.