



**Informe Metodológico para la Medición de
Impacto del Programa de Rescate de Espacios Públicos 2011**

Raymundo M. Campos-Vázquez

Centro de Estudios Económicos

Índice

| | |
|-----------------------------------------------------------|-----------|
| I. Introducción..... | 3 |
| II. Revisión de la literatura y Marco Teórico..... | 4 |
| III. Objetivos..... | 9 |
| IV. Datos | 10 |
| V. Indicadores de impacto | 11 |
| VI. Metodología..... | 15 |
| VII. Pasos Futuros | 21 |
| Referencias | 22 |

I. Introducción

La evaluación de los programas sociales es una herramienta indispensable para la rendición de cuentas sobre el cumplimiento de los objetivos del programa, su adecuada operación, y su impacto en el bienestar de la población. En este sentido la evaluación también contribuye a:

- Retroalimentar el diseño de los programas, permitiendo su mejoramiento, readecuación o reorientación.
- Rendir cuentas sobre el uso de recursos públicos y fomentar una cultura de resultados en la gestión pública.
- Incrementar la efectividad de la política social.
- Diseminar el conocimiento sobre el impacto de distintas herramientas de desarrollo social y promover el debate informado sobre políticas públicas.

La necesidad de obtener evidencia sobre la eficiencia en el diseño y operación de los programas sociales o establecer información confiable que dé un marco para la reorientación de los mismos con el fin de lograr los resultados deseados, ha hecho de la evaluación un instrumento de gran importancia. La evaluación capaz de generar evidencia de manera rigurosa sobre los efectos de las intervenciones es la evaluación de impacto; esto se logra al utilizar técnicas estadísticas avanzadas que hacen posible conocer de manera confiable los cambios que se producen en la población objetivo, es decir se puede determinar si los apoyos entregados logran los objetivos planteados.

El presente documento plantea la estrategia que se seguirá para la evaluación del Programa de Rescate de Espacios Públicos (PREP); dicha estrategia se justificará mediante la evidencia científica internacional. Asimismo, se especificarán tanto los datos a utilizar, como los indicadores principales a evaluar para conseguir los objetivos propios del programa.

En esencia, el PREP es una política pública que promueve el rescate y/o intervención de los espacios públicos, esto con el fin de promover la cohesión social y prevenir conductas antisociales y de riesgo entre la población que habita en colonias y barrios que se encuentran alrededor de los mismos y que se ubican en localidades urbanas de zonas metropolitanas y ciudades de 50,000 o más habitantes, preferentemente localizados en barrios o colonias que presentan marginación e inseguridad.

El objetivo general del Programa es contribuir a mejorar la calidad de vida y la seguridad ciudadana, mediante el rescate de espacios públicos en condición de deterioro, abandono o inseguridad, para que sean utilizados preferentemente por la población en situación de pobreza de las ciudades y zonas metropolitanas.

Para lograr dicho objetivo el programa considera dos tipos de intervención, una intervención general que ocurre cuando se interviene por primera vez y una intervención de consolidación. La aportación federal varía dependiendo de acuerdo con el tipo de intervención y espacio. En la modalidad de participación ciudadana se entregan de 30 mil a 300 mil pesos. La modalidad de mejoramiento físico considera aportes de 300 mil a 3 millones de pesos. Finalmente, las aportaciones de consolidación oscilan entre 150 mil y 7 millones de pesos.

La operación del programa sigue este marco típico: los municipios identifican espacios públicos posibles a rescatar. El gobierno estatal posteriormente informa a SEDESOL de los espacios preseleccionados para ser intervenidos, posterior a esto se decide quienes serán los beneficiarios del programa. Algunos criterios para su selección revisan la densidad poblacional, índices de violencia y posibilidades de realizar múltiples actividades en los espacios. Al final, el monto total de apoyo es la suma de los recursos de la SEDESOL, la aportación de los gobiernos locales y, en su caso, la de los beneficiarios o de otras instancias reconocidas. La aportación federal máxima podrá ser de hasta 64% del costo total del proyecto integral.

El orden del documento es el que sigue. En la sección II revisamos literatura previa con programas similares a nivel internacional así como el marco teórico. En la sección III mencionamos los objetivos de la evaluación. En la sección IV analizamos los datos disponibles y a utilizar en la evaluación. En la sección V discutimos las variables de resultados que se utilizarán en la evaluación. En la sección VI mencionamos la metodología del estudio a seguir, y finalmente discutimos los futuros pasos a seguir.

II. Revisión de la literatura y Marco Teórico

Una perfecta evaluación de cualquier programa contendría información sobre lo que pasaría si el grupo beneficiado por el programa no lo hubiese recibido y se observase el efecto del programa en el mismo grupo con tratamiento y sin tratamiento. Obviamente, esta información es imposible de obtener, lo que ocasiona que los investigadores busquen medios estadísticamente creíbles para formar un grupo de comparación adecuado. Rosenbaum y Rubin (1983) mencionan el error de realizar una simple comparación de distintas poblaciones, es decir, comparar los resultados de los

beneficiarios de algún programa contra los no beneficiarios del mismo. Estas comparaciones son equívocas pues los no beneficiarios de algún programa podrían no ser elegibles para el mismo y, por lo tanto, poseer características totalmente diferentes.

Un ejemplo particularmente interesante se puede encontrar en el trabajo de Jalany Ravallion (1998), donde evaluaciones previas a la propia de los autores de un programa anti-pobreza en China encontraron efectos importantes en el corto plazo, aunque estos desaparecían a largo plazo. El programa otorgaba crédito para proyectos de infraestructura en algunas aldeas previamente seleccionadas en situación de pobreza. Después de 10 años se encontró que las aldeas en el programa no alcanzaron mayores beneficios que las no seleccionadas. Los autores argumentan el error anterior a que las aldeas no seleccionadas contenían población no pobre. Los autores al diseñar correctamente un grupo de comparación mediante la metodología de *Propensity Score Matching* (PSM) (utilizaremos las siglas PSM para denotar el método de pareamiento por balanceo en los puntajes de propensión) lograron observar que de no haber recibido el programa las aldeas pobres encontrarían en una situación mucho peor. Con este correcto análisis se señaló que las políticas públicas pueden tener efectivamente efectos positivos en el largo plazo.

El contar con un diseño experimental es, sin duda, el mejor escenario para la evaluación de cualquier política o programa público. No obstante, resulta sumamente común el no contar con un grupo de control perfecto para ser comparado. Dehejia y Wahba (1999) encontraron que la metodología de PSM suele ser la estimación que más se acerca a los resultados experimentales. El PSM calcula la probabilidad de ser elegible a algún programa y, mediante ésta, genera un grupo deseado para calcular el impacto de recibir algún tratamiento. Una buena guía metodológica reciente se puede encontrar en el trabajo de Caliendo y Kopeinig (2008), pues sus recomendaciones son seguidas muy de cerca para la realización de la presente evaluación.

Dado que la evaluación de impacto que se realiza en el presente trabajo considera tres importantes objetivos: la disminución de percepción de inseguridad, el aumento del uso de espacios públicos y el aumento de la cohesión social, se ha revisado literatura de programas internacionales que cubran los mismos objetivos o los más similares posibles.

El trabajo clásico de Cho (1972) mostró mediante un análisis de regresión múltiple que los programas de servicio público disminuían mucho más la delincuencia en las 50 ciudades más grandes de E.E.U.U. que los programas de control. El autor define los programas de servicio público como todos aquellos que aumenten la calidad de vida urbana, como son acceso a caminos, transporte público, áreas verdes y de recreación (los programas de control se refieren a aumentar

número de policías o sentencias severas). Los efectos encontrados se mantienen en disminución de tasas de homicidio, violación, asaltos y robos violentos.

Brandon y Farrington (2004) evaluaron el efecto de la reducción en tasas de crimen mediante la imposición de circuitos cerrados y alumbrado público en distintas zonas peligrosas en Inglaterra. Se realizaron 19 evaluaciones de circuito cerrado y 13 evaluaciones de alumbrado público. En ambos casos se redujo el total de crímenes. Sin embargo, en crímenes específicos de robo de autos y asalto a mano armada el alumbrado público redujo aún más la tasa de criminalidad que las medidas de circuito cerrado. La evaluación se condujo para 2001 y es el resultado de un experimento natural.

En 1998 a nivel nacional en los Estados Unidos se implementó un programa que brindaba actividades recreacionales a niños después de clases. James-Burdumy et al (2007) condujeron una evaluación para identificar si se obtuvieron beneficios en materia de percepción de seguridad de los niños. Para realizar la evaluación se diseñó un experimento que fue efectuado desde 1999 hasta 2001. Los resultados fueron obtenidos mediante entrevistas directas en los grupos de control y tratamiento y registros de las primarias. Los resultados de la evaluación indican que los estudiantes se sienten mucho más seguros después de clases y los padres menos preocupados.

Makinen y Tyrvaenen (2008) levantaron un estudio cualitativo en Helsinki durante 2007 para entender las preferencias por parques y centros de recreación de los finlandeses ciudadanos. Es interesante mencionar que los adultos y adolescentes declaran acudir a este tipo de lugares cuando saben que la seguridad es buena. Además, en general, se asocia a este tipo de espacios públicos como lugares donde se pueden realizar actividades físicas en contacto con otras personas, de tal forma que es posible aumentar la cohesión social.

Jim (1989) presentó una evaluación de largo plazo (1978-1988) en Hong Kong para un programa nacional de provisión de parques. Debido a la difícil geografía del país se decidió establecer un programa donde se crearían constantemente parques nacionales, sin embargo la creación de estos sería a través de la demanda social. La demanda de estos parques ha crecido a través de los años y su uso ha crecido exponencialmente en los últimos dos años, según las técnicas de regresión con controles usadas por el autor.

Un resultado muy similar se observa en un estudio reciente de Choumert (2010). El autor evalúa las preferencias del votante medio en Francia por los espacios con áreas verdes durante el periodo de 2006-2007. Usando regresión con controles y regresión logística el autor puede identificar que la provisión de espacios de áreas verdes es menor que lo que desea el votante

promedio. Además, considera que sería viable el iniciar planes mixtos de cooperación gobierno-ciudadanía para proveer este bien público.

Los efectos de la proximidad de parques, plazas y áreas verdes fueron estudiados por Cohen et al (2006). Los autores realizaron un estudio sobre actividad física en distintas ciudades de Estados Unidos con población adolescente femenina. Mediante el seguimiento de reportes de ejercicio en un año (2005) y con información adicional para controlar (edad, raza, educación, etc.) los autores argumentan el aumento de 17 minutos al día en actividad física en la adolescente promedio causado por tener un parque una milla más cerca.

El aumento de ejercicio en todas las edades se estudió en la ciudad de Chicago por Fan et al (2011). Su estudio, de carácter altamente cuantitativo, corrobora lo previamente mencionado y además argumenta que el tener acceso a áreas verdes disminuye el estrés y aumenta la cohesión social. La investigación utilizó una encuesta de 2002 y 2003, y el método de variables instrumentales.

Lokshin y Ruslan (2003) diseñaron distintas evaluaciones de impacto para diferentes programas de rehabilitación de infraestructura en las áreas rurales de Georgia. Los autores usaron datos provenientes de encuestas politemáticas a los hogares. Se encontraron 549 proyectos distintos de infraestructura entre los cuales destacan escuelas, caminos y sistemas de agua. Para realizar la evaluación los autores utilizan el PSM. En las aldeas con programas activos el número de estudiantes se elevó un 13% más que en las que no contaban con el programa. Además, se redujo en 5.7% la falta de estudiantes por mejoras en las escuelas, así como se redujo el tiempo de llegar al distrito más cercano hasta en 35 minutos. El número de actividades agrícolas aumentó y se redujo el tiempo de auxilio de las ambulancias hasta los 23 minutos. También se registraron menos problemas de enfermedades por agua contaminada.

En Florida se realizó un diseño semi-experimental para evaluar los resultados de un programa que se dedica a controlar el comportamiento riesgoso de los adolescentes (problemas con la autoridad, manejo de la ira, etc.). El programa, evaluado por Stewart y Rowe (2011), invitó a participar a jóvenes en 2007-2009, donde se recuperaron encuestas específicas del programa. Se utilizaron los censos para descripciones del vecindario con 10 variables que medían el vecindario (ventajoso, desventajoso e inhabitable). La evaluación se realizó con PSM para comparar áreas tratadas y no tratadas. Como resultado interesante, destaca que al padecer un vecindario de ausencia de servicios como parques, gimnasios y clubes sociales hace que se disminuyan los efectos del programa sobre el comportamiento riesgoso de los jóvenes.

Pradhan y Rawlings (2002) analizaron el caso de gasto social en Nicaragua para educación, salud, drenaje y alcantarillado. Ellos utilizaron una estrategia de identificación de PSM en características observables, además de proximidad geográfica, para una muestra de hogares beneficiarios y no beneficiarios. Encontraron que la inversión en educación tiene un efecto positivo en la asistencia escolar. Por otro lado, inversiones en salud y alcantarillado causan una mejora en el servicio pero no tiene efectos significativos en la salud.

Además de aumentar el uso de los espacios públicos y convertirlos en lugares seguros, la presente evaluación está interesada en saber si se logra mejorar la cohesión social. Gibbs (1974) explica que la recreación en espacios públicos es un factor crucial para aumentar la convivencia de los individuos. En el trabajo de Kázmierczak (2011) se encuentra un ejemplo muy interesante sobre esto. En ella se analizaron 3 vecindarios en Manchester durante 2011, cada uno con distintos niveles de etnicidad y pobreza, y se descubrió que la presencia de parques facilita la convivencia de distintas clases sociales o razas. Para llegar a este resultado se empleó la técnica de variables instrumentales.

Durante 2010 se realizó una evaluación de impacto del programa brasileño Vencer. El trabajo de Spaaij (2012) analiza cómo la creación de programas deportivos y restauración de canchas aumentaron la cohesión social entre jóvenes con desventajas. Para lograr dicha evaluación se condujeron entrevistas cualitativas en el área de Rio de Janeiro (específicamente en las favelas) y se utilizó información relevante del instituto de población, para después realizar regresiones con controles. Los resultados apuntan a un incremento en la cohesión, reducción de violencia intrajuvenil y un aumento del capital social.

Esta sección se limita a una revisión de la literatura de evaluaciones de impacto similares al PREP, no obstante, existe amplia evidencia de evaluaciones de impacto muy transparente y clara que se tomará en cuenta. Por ejemplo, se siguen las recomendaciones de Van de Walle (2008), mismas que han sido aplicadas con éxito para la evaluación de programas de mejora de caminos rurales. Además se cuenta con un resumen excelente de distintas evaluaciones para economías desarrolladas (ver OECD 2004, 2008). Finalmente para profundizar en el tema de la implementación correcta del PSM en evaluaciones donde se mejore la infraestructura, se revisaron los artículos de Todd (2006) y Kandikov y Rnkow (2009).

III. Objetivos

La evaluación debe ser capaz de generar evidencia de manera rigurosa sobre los efectos de las intervenciones; esto se logra al utilizar técnicas estadísticas avanzadas que hacen posible conocer, de manera confiable, los cambios que se produce en la población objetivo. Es decir, se puede determinar si los apoyos entregados logran los objetivos planteados.

En este sentido y, con base en evaluaciones externas anteriores, se ha manifestado la necesidad de una evaluación de impacto para el PREP, que dé cuenta sobre sus efectos en la población objetivo (ver la Evaluación Específica de Desempeño 2009-2010 y 2010-2011)¹. Además, tomando en consideración que el programa desde 2009 elaboró un instrumento para monitorear sus resultados, mediante el cual se ha logrado la obtención de información estadísticamente representativa (antes y después de la intervención), se poseen elementos para llevar a cabo un análisis sobre los impactos generados al rescatar espacios públicos.

Por tanto, la presente evaluación de impacto tiene los siguientes objetivos:

3.1 Objetivo general

El objetivo de la evaluación es determinar el impacto atribuible a las intervenciones del Programa de Rescate de Espacios Públicos sobre la percepción de inseguridad y cohesión social en torno a los espacios públicos.

3.2 Objetivos específicos

- A. Establecer si las intervenciones del Programa de Rescate de Espacios Públicos tienen un impacto en la disminución de la percepción de inseguridad en los espacios públicos en zonas urbano-marginadas.
- B. Delimitar si las intervenciones del Programa de Rescate de Espacios Públicos aumentan el uso de los espacios públicos trabajados.
- C. Determinar si las intervenciones del Programa de Rescate de Espacios Públicos incrementan la cohesión social entre los habitantes circuncidantes a los espacios públicos intervenidos.

¹ Evaluación Específica de Desempeño del Programa de Rescate de Espacios Públicos 2009 – 2010
<http://www.coneval.gob.mx/cmsconeval/rw/pages/evaluacion/evaluaciones/especificas/SEDESOL.es.doy>
Evaluación Específica de Desempeño del Programa de Rescate de Espacios Públicos 2010 – 2011
http://www.coneval.gob.mx/cmsconeval/rw/pages/evaluacion/evaluaciones/especificas/edd_2010_2011/sedesol.es.do

La pregunta clave que se buscó responder desde el inicio de la implementación es: ¿Qué tan efectivo es el Programa de Rescate de Espacios Públicos en el logro de sus objetivos?

IV. Datos

El Programa de Rescate de Espacios Públicos desde 2009 ha desarrollado un instrumento riguroso con validez externa e interna para conocer los resultados de sus intervenciones. Este instrumento es una encuesta que permite comparar la situación de espacios antes de ser intervenidos por el programa, con la de espacios que fueron “rescatados” (intervenidos) en ese año, y establecer así algunas relaciones en los resultados.

Dicho instrumento se denomina, “Encuesta Nacional sobre Percepción de Inseguridad, Conductas de Riesgo y Participación Social en Espacios Públicos”, y es el resultado de un esfuerzo conjunto de la Dirección General de Equipamiento e Infraestructura en Zonas Urbano-Marginadas (DGEIZUM) y la Dirección General de Evaluación y Monitoreo de los Programas Sociales (DGEMPS) de la SEDESOL.

Dicha encuesta se realiza de manera anual desde 2009², y permite que el Programa de Rescate de Espacios Públicos cuente con indicadores para medir sus resultados. También es importante indicar que la encuesta se levanta en el primer semestre del ejercicio fiscal, con esto se consigue que en los espacios públicos que se pretenden intervenir aún no se realicen acciones en ese año.

Otra base de datos interna corresponde a los registros administrativos, donde con ella se puede conocer información financiera sobre las aportaciones federales, estatales y, en caso de que aplique, la de los beneficiarios. También se podrán utilizar los resultados del Censo de Población y Vivienda 2010(que incluye alrededor de 190 variables a nivel de manzana), las encuestas nacionales sobre inseguridad 2009 (ENSI-6), las de victimización y percepción sobre la seguridad pública 2011 (ENVIPE) y los datos del Censo de Población 2005.

El análisis que pretendemos realizar es a nivel espacio público, por lo que en ningún momento se utilizarán variables individuales para estimar impactos. Las variables de resultados se obtendrán de “Encuesta Nacional sobre Percepción de Inseguridad, Conductas de Riesgo y Participación Social en Espacios Públicos”. Las variables de control se obtendrán de registros administrativos y del Censo y Censo de Población y Vivienda 2005 y 2010 respectivamente.

²En 2008 se hizo el primer ejercicio, pero no con un instrumento estandarizado.

V. Indicadores de impacto

Como se ha mencionado existen 3 distintos objetivos a evaluar, por lo que en esta sección se pretende describir sus indicadores de impacto específicos. Los indicadores de impacto son las variables que indicarán el éxito del programa y qué tanto cumplen con los objetivos del mismo. La Tabla 1 contiene la información relevante con respecto al cumplimiento del objetivo relacionado con la disminución de la percepción de inseguridad (Objetivo específico A, sección 3.2).

Es importante mencionar que existen diferencias entre la encuesta de 2009 y las encuestas realizadas en 2010 y 2011, este cambio en la información imposibilita la comparación perfecta de distintos indicadores y, por lo tanto, sólo se evaluará 2010 y 2011 cuando llegase a faltar información de 2009.

La primera comparación de percepción de seguridad de la Tabla 1 cumple con el objetivo, sin embargo la respuesta es limitada ya que el encuestado se encuentra forzado a encontrarse en algún extremo o punto medio. Sin embargo la pregunta es comparable en los 3 años. La segunda comparación le da la flexibilidad al encuestado de disponer de calificar del 1 al 10 la seguridad, otorgando mayor información. No obstante, esta información se encuentra disponible sólo para el periodo de dos años de 2010-2011.

| Tabla 1 Cambio en el porcentaje de hogares que manifiestan percepción de inseguridad en los espacios públicos intervenidos. | | |
|----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| 2009 | 2010 | 2011 |
| P16¿Cómo se siente en el(Espacio Público)? | B4¿Cómo se siente en el(Espacio Público)? | B4¿Cómo se siente en el(Espacio Público)? |
| Totalmente seguro Medianamente seguro Totalmente Inseguro NS/NR | Totalmente seguro Medianamente seguro Totalmente Inseguro NS/NR | Totalmente seguro Medianamente seguro Totalmente Inseguro NS/NR |
| No existe pregunta para comparar | B10 Dígame qué tan seguro(a) se siente en el (Espacio Público) | B10 Dígame qué tan seguro(a) se siente en el (Espacio Público) |
| No existe pregunta similar para comparar | Posible resultado de 1 a 10, donde 1 se considera muy bajo y 10 se considera muy alto | Posible resultado de 1 a 10, donde 1 se considera muy bajo y 10 se considera muy alto |

| Cambio en el porcentaje de hogares que consideran buenas las condiciones de seguridad en el espacio público intervenido | | |
|-------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| P23En general, ¿cómo considera usted las condiciones de seguridad en (Espacio Público)? | B5En general, ¿cómo considera usted las condiciones de seguridad en (Espacio Público)? | B5En general, ¿cómo considera usted las condiciones de seguridad en (Espacio Público)? |
| Muy buenas | Muy buenas | Muy buenas |
| Buenas | Buenas | Buenas |
| Malas | Malas | Malas |
| Muy Malas | Muy Malas | Muy Malas |
| NS/NR | NS/NR | NS/NR |

Respecto a la percepción de las condiciones de seguridad sólo se cuenta con un indicador y, además, este resulta comparable durante los tres años. Existen otras preguntas de seguridad en la encuesta que posiblemente podrían ser utilizadas, no obstante sólo aplican para 2010 y 2011. Estas preguntas competen información sobre el conocimiento del entrevistado de actos delictivos dentro del espacio público (robos, gente lastimada, venta de drogas y ascoso femenino). El estudiar esas variables podría arrojar información interesante sobre el tipo de percepción de actos ilegales mejorados por el programa.

La forma de medición que se propone es la siguiente. Para la pregunta de percepción de seguridad en el espacio público proponemos medir como éxito si el hogar respondió Totalmente seguro o Medianamente seguro. Aunque dependiendo de la variación de las respuestas se podrá ajustar este indicador. Para la pregunta sobre las condiciones de seguridad se considera como éxito si el hogar contestó Muy Buenas o Buenas. En general, para las siguientes preguntas se propone definir un éxito si se contestó a las primeras dos opciones de respuesta. Aunque puede haber variaciones en esta codificación dependiendo de los tabulados que se observen en las bases de datos.

La Tabla 2 discute los indicadores que se utilizarán para el aumento del uso de espacios públicos (Objetivo específico B, sección 3.2). Desgraciadamente la forma en que se realizaron las encuestas a través de los años resulta en una comparación no perfecta entre 2009 y 2010, 2011. Este aspecto ocurre cuando se trata de comparar qué tipo de actividades se realizó en el espacio público puesto que el formato de la pregunta cambió, no obstante se tratará de recuperar la información.

| Tabla 2 Cambio en el porcentaje de hogares que asistena los espacios públicos intervenidos. | | |
|------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2009 | 2010 | 2011 |
| P17 Durante este año, ¿usted o alguna persona de su hogar asistieron o asisten a (Espacio Público)? | B3. Durante este año, ¿usted o alguna persona de su hogar asistieron a (Espacio Público)? | B3. Durante este año, ¿usted o alguna persona de su hogar asistieron a (Espacio Público)? |
| Sí No NS/NR | Sí No NS/NR | Sí No NS/NR |
| Cambio en el porcentaje de hogares satisfechos con las condiciones físicas de los espacios públicos | | |
| P22En general, ¿Cómo considera usted las condiciones de físicas del(Espacio Público)? | B5En general, ¿cómo considera usted las condiciones físicas del (Espacio Público)? | B5En general, ¿cómo considera usted las condiciones físicas del (Espacio Público)? |
| Muy buenas Buenas Malas Muy Malas NS/NR | Muy buenas Buenas Malas Muy Malas NS/NR | Muy buenas Buenas Malas Muy Malas NS/NR |
| Cambio en el porcentaje de hogares que participaron en actividades deportivas, culturales o recreativas del espacio público intervenido | | |
| P 18¿Qué persona de su hogar asiste y en qué participa? Mencione todos los que asistan. | B8Durante este año, ¿usted o alguna persona de su hogar participó en actividades deportivas, culturales, cívicas o recreativas en (Espacio Público)? | B8Durante este año, ¿usted o alguna persona de su hogar participó en actividades deportivas, culturales, cívicas o recreativas en (Espacio Público)? |
| Se tiene que capturar a cada integrante y especificar la actividad | Sí No NS/NR | Sí No NS/NR |

La Tabla 2 muestra tres preguntas de interés. La primera se refiere al cambio en los hogares que asisten a los espacios públicos intervenidos. La segunda pregunta de interés se refiere a si los hogares están satisfechos con las condiciones físicas de los espacios públicos. La tercera pregunta se refiere a si se participó en actividades que se realizan en el espacio público. Lamentablemente, esta última pregunta no es estrictamente comparable entre encuestas. Por lo que se analizará la posibilidad de comparación a través de los años, o bien restringir la comparación para el periodo 2010-2011.

La Tabla 3 muestra las preguntas con las cuales se evaluará la cohesión social, la comparación es perfecta entre los años (objetivo específico C, sección 3.2).

| Tabla 3 Cambio en el porcentaje de hogares que consideran muy unida la relación entre vecinos | | |
|----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| 2009 | 2010 | 2011 |
| P13 ¿Cómo clasificaría la relación entre los vecinos de esta colonia/barrio? | B1 ¿Cómo clasificaría la relación entre los vecinos de esta colonia/barrio? | B1 ¿Cómo clasificaría la relación entre los vecinos de esta colonia/barrio? |
| Muy unida | Muy unida | Muy unida |
| Unida | Unida | Unida |
| Indiferente | Indiferente | Indiferente |
| Algunas disputas y conflictos | Algunas disputas y conflictos | Algunas disputas y conflictos |
| Gran conflicto y violencia | Gran conflicto y violencia | Gran conflicto y violencia |
| NS/NT | NS/NT | NS/NT |

Se propone clasificar como éxito de cohesión social si el hogar responde a la pregunta en la Tabla 3 como Muy Unida o Unida.

Adicionalmente se requiere crear un índice de capital social para lograr el objetivo anterior. En 2010 y 2011 se utilizaron las preguntas indicadas en la Tabla 4 (objetivo específico C, sección 3.2). Se sumarán la respuesta de cada una de ellas y se dividirá entre 80. Las personas pueden responder cada una de estas preguntas dentro de un rango del 1 al 10, donde 1 significa totalmente en desacuerdo y 10 totalmente de acuerdo.

Tabla 4
Construcción de un índice de capital social comparable
entre 2010-2011

- D1.- Las personas que viven en su colonia están dispuestas a ayudar a sus vecinos
- D2.- Los vecinos de su colonia se llevan bien en general
- D3.- Se puede confiar en los vecinos de su colonia
- D4.- Los vecinos de su colonia comparten los mismos valores morales
- D5.- Me gusta participar en convivios con su vecinos
- D6.- Le pediría consejo o ayuda a mis vecinos
- D7.- Los vecinos se organizan para realizar actividades en beneficio de la colonia
- D8.- La relación entre los vecinos de mi colonia es buena en General

Se puede construir un índice de capital social con la encuesta de 2009, sin embargo la información que se capture en ese año será muy diferente a la capturada en 2010-2011. En específico, se puede utilizar las preguntas referentes a acción colectiva y cooperación (p11-p12).

VI. Metodología

La metodología que se pretende seguir es el de pareamiento mediante puntajes de propensión a nivel espacio público (por simplicidad y comparación con la literatura en el tema llamamos a la metodología PSM por sus siglas en inglés Propensity Score Matching). Esta metodología ha sido ampliamente utilizada en evaluaciones de impacto (ver, por ejemplo Rosenbaum y Rubin, 1983, así como la evaluación de impacto del Programa Hábitat disponible en la página de la SEDESOL).

El reto del estudio es encontrar un grupo de control similar al grupo de tratamiento, lo cual equivale a espacios públicos intervenidos. De tal manera que una comparación entre el grupo de tratamiento y el de control nos lleve a conclusiones sobre la efectividad del programa. Una ventaja de la Encuesta Nacional sobre Percepción de Inseguridad, Conductas de Riesgo y Participación Social en Espacios Públicos es que tiene un panel rotativo y que realiza entrevistas a hogares en espacios públicos antes de la intervención. Así mismo, se tiene información para todas las manzanas intervenidas utilizando el Censo de 2005 y Censo de 2010. Por tanto, se pueden calcular

estadísticas con base en la información a nivel manzana de los censos antes del tratamiento para todas las áreas de los espacios públicos.

Una vez que se tiene información antes del tratamiento para todos los espacios públicos, el objetivo es encontrar un grupo de control adecuado para los espacios tratados. Lo que se plantea aquí es utilizar como grupo de control a los espacios públicos que serán tratados en algún momento en el tiempo. Además, un punto adicional es que dadas las limitaciones de datos a través del tiempo, no es posible calcular estimaciones de diferencias en diferencias. Es decir, el grupo de control (espacio público no tratado) no es observado por uno o dos años antes de la intervención. Por tanto, la estimación de impacto sería con una diferencia únicamente entre el grupo de tratamiento y control. Para lograr un balance en características observables, y la estimación sea más robusta, se plantea estimar esa diferencia ajustada con el método de apareamiento mediante puntajes de propensión.

En particular se desearía estimar una ecuación como la siguiente:

$$Efecto = E[Y | T = 1, X] - E[Y | T = 0, X]$$

El problema fundamental de la inferencia causal es encontrar un grupo de control adecuado. El efecto (conocido como el “Efecto de Tratamiento sobre los Tratados”) es calculado como un promedio de la variable de interés en 2010 o 2011 para el grupo de tratamiento y para el grupo de control. No hay problema alguno para encontrar el promedio para el grupo de tratamiento. El grupo de control es un poco más complicado puesto que tiene que ser un grupo con características similares al grupo de tratamiento.

Esta estrategia de balanceo por puntajes de propensión fue desarrollada por Rosenbaum y Rubin (1983) (PSM). Los supuestos principales implican que una vez que controlamos por el puntaje de propensión, el grupo de tratamiento y de control son similares (o bien que no existen sesgos por variables no observables), y que se puede lograr el balanceo de características entre grupos de control y tratamiento. Por tanto, en esta parte explicamos los supuestos y el método de implementación de la estrategia PSM para identificar efectos causales. Primero mencionamos los supuestos generales y específicos del método, y después discutimos las formas de implementar la estrategia. Discusiones generales de este método pueden encontrarse en Caliendo y Kopeinig (2008), Dehejia y Wahba (1999, 2002), Smith y Todd (2002), y Todd (2008).

La estrategia del PSM asume selección en observables. Es decir, que no existen sesgos de selección dado que se controla por diferencias en características observables (X) de los grupos de

tratamiento y control. La estrategia no es explícita en qué variables sirven para controlar por sesgo de selección. Sin embargo, Rubin y Thomas (1996) mencionan que se deben de incluir todas las variables que sean posibles *antes* del tratamiento y que *pudieran determinar* el tratamiento. La identificación del efecto causal sobre los tratados (*ATT*) para selección en variables observables requiere de ciertos supuestos:

Supuesto 1. Selección en características observables: $Y_0 \perp D|X$. Esto es que una vez que controlamos por X , no hay sesgo de selección. Donde \perp se refiere a independencia, el supuesto se refiere a que los resultados potenciales son independientes del tratamiento una vez que condicionamos por las variables observadas. Esto implica que todas las variables que influyen en el tratamiento y los resultados potenciales son observadas por el investigador. Esto es un supuesto muy fuerte y tiene que ser justificado por el investigador.

Supuesto 2. Soporte común: $0 < \Pr(D = 1|X) < 1$, aunque lo único que se necesita es que $\Pr(D = 1|X) < 1$. Este supuesto asegura que para municipios con los mismos valores de X se tiene una probabilidad positiva y no perfecta para individuos en el tratamiento y control. También implica que los municipios tienen que ser comparables dado $\Pr(D = 1|X)$.

Si tuviéramos un número limitado de variables discretas, se podría hacer un apareamiento o matching directo sobre esas variables. Lamentablemente, en la evaluación no se pueden seleccionar espacios públicos con las mismas X todo el tiempo. Esto se debe al problema de multidimensionalidad. Es decir, hacer el matching perfecto para muchas variables es prácticamente imposible. Rosenbaum y Rubin (1983) sugirieron utilizar los llamados balances por puntajes de propensión. La probabilidad de participar en el programa dado un conjunto de variables X puede ser estimado como $\Pr(D=1|X) = P(X)$ y a eso se le conoce como el *propensity score* o puntaje de propensión.

El puntaje de propensión es definido como

$$P(X) = \Pr(T = 1|X)$$

donde T es la variable de tratamiento, X es el conjunto de variables explicativas, y $\Pr()$ es una función de probabilidad, que puede ser el modelo logístico o el normal. Dentro del conjunto de variables X se encuentran características observables del año 2000 y 2005.

El efecto del programa puede ser encontrado entonces como

$$Efecto = E[Y|T = 1, p(X)] - E_{p(X), T=1}\{E[Y|T = 0, p(X)]\}$$

Es decir, la diferencia ahora es que el grupo de control tendrá una nueva distribución para que se parezca lo más posible al grupo de tratamiento. Esa nueva distribución es obtenida con nuevos pesos para cada observación en el control (puede que el peso sea diferente a la unidad). Asimismo, esos nuevos pesos pueden ser obtenidos por diferentes estrategias de apareamiento (matching).

Uno de los posibles problemas de esta estrategia es que las características de los espacios públicos sean diferentes entre el grupo de tratamiento y control. Esto es debido a la naturaleza misma del programa. Por tanto, es de suma importancia mostrar suficientes pruebas de robustez para verificar que las pruebas de balanceo se cumplen. De otra forma, es posible que la estimación sea sesgada y no muestre el efecto verdadero del programa. En términos de la ecuación descrita arriba, es necesario mostrar que el apareamiento balancea las características observables.

En este punto se plantea seguir la metodología de Dehejia y Wahba (1999) para seleccionar el puntaje de propensión adecuado. Además, se plantea seguir las pruebas que ellos mencionan para asegurarse que exista un balance en las características observables (pruebas por medio de agrupamiento en el puntaje de propensión, por tanto unidades en el control y tratamiento dentro de cada agrupamiento deben ser muy similares). Además de las pruebas mencionadas en Dehejia y Wahba (1999), se plantea incrementar el número de pruebas mencionadas en Caliendo y Kopeinig (2008). Estas pruebas consisten en pruebas de verosimilitud después del matching y en analizar el error estandarizado de las características observables antes y después de matching.

Aún con las pruebas descritas, la evaluación tiene retos importantes. En primer lugar, aún si se contara con otro grupo de control, se encuentra el aspecto de cómo controlar por la selección al programa. Es decir, los municipios proponen los espacios públicos a intervenir y con base en ciertas reglas se decide apoyar a esos municipios. No es claro cómo resolver el problema de selección teniendo una muestra diferente que no solicite apoyos del programa. La gran ventaja de los datos que se plantean utilizar es que la selección no debería ser un aspecto a considerar porque todos los espacios públicos fueron propuestos por las autoridades municipales y serán tratados en algún momento en el tiempo. Por este lado, estamos eliminando una fuente importante de sesgo en la estimación.

En segundo, la metodología descrita permite controlar por posibles efectos de temporalidad (espacios públicos entre años son diferentes entre sí). De esta forma, se trata de comparar espacios

públicos tan similares como sea posible antes de la intervención. Finalmente, otra ventaja de la metodología es que tiene pruebas de robustez que se pueden utilizar para analizar la sensibilidad del resultado al método específico utilizado así como analizar la comparación en características observables entre el grupo de control y tratamiento.

Ahora se explicará con más detalle el aspecto metodológico. En la Tabla 5 se muestra una tabla donde se denota explícitamente los grupos de tratamiento y control.

Tabla 5 Diseño rotativo de la Encuesta—Propuesta de grupos de tratamiento y control

| Año de Inicio | Año de Levantamiento | | |
|---------------|----------------------|----------------------|----------------------|
| | 2009 | 2010 | 2011 |
| 2009 | Línea base 2009 | Intervenidos en 2009 | |
| | | (Tratamiento) | |
| 2010 | | Línea base 2010 | Intervenidos en 2010 |
| | | (Control) | (Tratamiento) |
| 2011 | | Línea base 2011 | |
| | | (Control) | |

Es decir, los espacios públicos tratados en 2009 que se observan en la encuesta de 2010 se comparan con los espacios públicos no tratados en la encuesta de 2010, pero que eventualmente en ese año se intervendrán. Esto ocurre similarmente para el año 2011. Se plantea juntar la información de todos los espacios públicos intervenidos de 2009-2010 para obtener el impacto promedio del programa en ese periodo de tiempo.

Como se mencionaba anteriormente, un problema de esa comparación es que es posible que los espacios públicos (aunque todos sean tratados eventualmente) sean diferentes año con año. Para tratar de resolver por esta problemática se plantea utilizar la metodología de pareamiento de puntajes de propensión. Si utilizamos la información de manzanas del Censo de 2005 (o bien del Censo de 2010), podemos comparar espacios públicos que sean tan similares como sea posible antes de que todos los espacios públicos sean intervenidos. De esta forma se calcularía la diferencia en las variables de interés entre el grupo de tratamiento y control ajustando por diferencias en características observables por el pareamiento mediante puntajes de propensión.

Una limitante del enfoque anterior es que no es posible obtener estimadores de diferencias en diferencias. Esto se debe a que no se tiene información del espacio público para un periodo de 1-2 años antes de ser tratado. Por lo que sólo se puede utilizar la información de justo antes de ser

tratado. Esto puede ser una limitante en el estudio, ya que si las tendencias entre el tratamiento y control son diferentes (antes del tratamiento) podríamos argumentar que parte del efecto estimado se debe a otras causas y no al programa.

Dado esa limitante, es sumamente importante que el método de pareamiento por puntajes de propensión sea robusto. Se plantea la necesidad de estimar diferentes estadísticos para comprobar que los grupos de tratamiento y control son comparables a través del tiempo. Asimismo, la literatura de pareamiento señala que es importante que el método sea robusto al método de pareamiento. Por tanto, en la evaluación se mostrará evidencia que los grupos de control y tratamiento son comparables con los estadísticos de sesgo estandarizado, significancia conjunta, estratificación y pruebas de balance. De la misma forma, se mostrará la robustez al método de pareamiento. A continuación se explican estos métodos.

Si los grupos de tratamiento y control no son comparables en términos de características observables, entonces es posible que el efecto estimado esté contaminado, y no sea un efecto causal. Por esta razón es sumamente importante mostrar que ambos grupos son comparables. El estadístico de sesgo estandarizado se calcula como:

$$SB = 100 \times \frac{(\bar{X}_T - \bar{X}_C)}{\sqrt{0.5(Var(X_T) + Var(X_C))}}$$

Y se compara antes y después del pareamiento para cada una de las variables. Se considera un buen pareamiento si en la mediana de la distribución el SB es menor a 5%. Otro método comúnmente usado es el de significancia conjunta que consiste en estimar el puntaje de propensión ($\Pr(T = 1|X) = F(\alpha + \beta X)$) antes y después del pareamiento, y mostrar el estadístico F de $H_0: \beta = 0$. Si el pareamiento está haciendo un buen trabajo, entonces el estadístico F después del pareamiento debería ser cercano a 0 de tal forma que no pueda rechazarse la hipótesis nula. El no rechazo implica que después del pareamiento las características observables no pueden predecir el tratamiento. Este aspecto es fundamental para poder argumentar causalidad del programa. Las pruebas de estratificación y balance son simplemente pruebas de diferencias en medias para toda la distribución o partes específicas de la distribución. Esto asegura que las características observables son similares entre tratamiento y control en distintos puntos de la distribución del puntaje.

Finalmente, el estimador del efecto del programa se calcula como:

$$E_{P(X)|T=1}\{E[Y|T = 1, P(X)] - E[Y|T = 0, P(X)]\}$$

Es decir, se calcula el valor promedio de la variable de interés para los grupos de tratamiento y de control para valores similares en el puntaje de propensión (pareamiento). Existen

diversos métodos en la literatura sobre cómo hacer el pareamiento: como método no paramétrico, método de vecinos cercanos y método de radio. En la evaluación se presentarán diversos métodos de pareamiento para demostrar que no existen sesgos por el método utilizado.

VII. Pasos Futuros

Una vez recibidos los datos se procederá a la limpieza y verificación de los mismos. Se obtendrán estadísticas descriptivas y se generarán los índices e indicadores pertinentes. Se procederá a realizar la metodología de PSM para realizar la evaluación. Se realizarán las pruebas estadísticas correspondientes, de tal forma que se cuente con información transparente. Se realizará un documento en el cual se muestren los resultados obtenidos de la evaluación, así como cualquier suceso destacable.

Cabe mencionar que se tratará de complementar la revisión de los objetivos del programa con información interesante que pueda surgir al momento de la evaluación, y pertinente al mismo.

Referencias

- Brandon, Welsh; Farrington, David (2004). *Surveillance for Crime Prevention In Public Space: Results and Policy Choices in Britain and America*. Inter-American Development Bank. Vol. 3(3), pp. 497-526 .
- Caliendo, Marco; Sabine Kopeinig, (2008). *Some Practical Guidance For The Implementation Of Propensity Score Matching*, Journal of Economic Surveys, Blackwell Publishing, vol. 22(1), pages 31-72
- Cho, Yong.; (1972). *A Multiple Regression Model for the Measurement of the Public Policy Impact on Big City Crime*. Policy Sciences, 3(4), pp. 435–455.
- Choumert, Johana; (2010). *An empirical investigation of public choices for green spaces*. Land and Use Policy, Vol. 27 , pp. 1123-1131.
- Cohen, Deborah; Ashwood, Scott; Scott, Molly; Overton, Adrian; Evenson, Kelly; Staten, Lisa; Porter, Dwayne; Mckenzie, Thomas; Catellier, Diane (2006). *Public Parks and Physical Activity Among Adolescent Girls*. Pediatrics, Vol. 16, No. 2, pp. 241-274.
- Dehejia, Rajeev H.; SadekWahba (1999). *Causal Effects in Nonexperimental Studies: Re-Evaluating the Evaluation of Training Programs*. Journal of the American Statistical Association. 94:1053-62.
- Escobal, Javier; Ponce, Carmen (2002). *The benefits of rural roads: enhancing income opportunities for the rural poor*. Lima: GRADE, 2002. Working paper, 40-I.
- Fan, Yingling.; Das, Kirti (2011). *Neighborhood green, social support, physical activity, and stress: Assessing the cumulative impact*. Health and Place. Vol. 17, pp. 1202-121
- Gibbs, Kenneth; (1974). *Evaluation of Outdoor Recreational Resources: A Note*. Land Economics .vol. 50 (3) pp. 309-311
- Imbens, Guido W.; Lemieux, Thomas (2008). *Regression discontinuity designs: A guide to practice*, Journal of Econometrics, Elsevier, vol. 142(2), pp. 615-635, Febrero.
- Independent Evaluation Department, Asian Development Bank (2009). *Impact of rural water supply and sanitation in Punjab, Pakistan*, Agosto 2009
- Jalan J.; M. Ravallion (1998). *Are There Dynamic Gains from a Poor-Area Development Program?* Journal of Public Economics, Vol. 67(1): 65-85.
- James-Burdumy, Susanne.; Dynarski, Mark (2007). *When Elementary Schools Stay Open Late: Results from the National Evaluation of the 21st Century Community Learning Centers Program* . Educational Evaluation and Policy Analysis. Vol. 29(4), pp. 296-318.
- Jim, CY; (1989). *Changing Patterns of Country-Park Recreation in Hong Kong*. The Geographical Journal, vol. 155 (2) pp. 167-178.

- Kandilov, Ivan; Renkow, Mitch (2010). *Infrastructure investment and rural economic development: An evaluation of USDA's broadband loan program*. Growth and Change Vol. 2, pp. 165-191. Junio 2010
- Kázmierczak, Aleksandra; (2012). *The contribution of local parks to neighbourhood social ties*. Landscape and Urban Planning.
- Lokshin, Michael; Yemtsov, Ruslan (2003). *Evaluating the impact of infrastructure rehabilitation projects on household welfare in rural Georgia*. World Bank Policy Research Working Paper 3155, Octubre 2003.
- Makinen, Kirsi; Tyrvaenen, Lissa;(2008). *Teenage experiences of public green spaces in suburban Helsinki*. Urban Forestry and Urban Greening, pp. 277-289
- OECD (2004). *Evaluating Local Economic and Employment Development. How to assess what works among programmes and policies*. OECD, Paris, France.
- OECD (2008). *A Review of Local Economic and Employment Development Policy Approaches in OECD Countries*. OECD Local Economic and Employment Development Programme. OECD, Center for Entrepreneurship, Paris, France.
- Pradhan, M.; R. Rawlings (2002). *The Impact and Targeting of Social Infrastructure investments: Lessons from the Nicaragua Social Fund*. The World Bank Economic Review, Vol. 16(2): 275-295
- Rogers, Cynthia; Tao, Jill (2004). *Quasi-Experimental Analysis of Targeted Economic Development Programs: Lessons from Florida*. Economic Development Quarterly 2004, 18.
- Rosenbaum, Paul R.; Donald B. Rubin (1983). *The Central Role of the Propensity Score in Observational Studies for Causal Effects*. Biometrika. 70: 41-55.
- Rubin, D.B.; N. Thomas (1986). *Matching Using Estimated Propensity Score: Relating Theory to Practice*. Biometrika, 52(1): 249-264.
- Stewart, Chris; Rapp-Paglicci, Lisa; Rowe, William; (2011). *The Impact of Neighborhood Factors on Mental Health and Academic Outcomes for Adolescents Participating in the Prodigy Programs*. Journal of Human Behavior in the Social Environment, vol. 21(4), pp. 383-397
- Spaaij, Ramón; (2012). *Buildingsocial and cultural capital among young people?in disadvantaged communities: lessons from a Brazilian sport-based intervention program*. Sport Education and Society. Vol. 17 (1), pp. 77-95.
- Smith, Jeffrey A., y Petra E. Todd (2005). "Does matching overcome LaLonde's critique of nonexperimental estimators?", *Journal of Econometrics*, 125, pp. 305-353

- Todd, Petra E. (2008). "Evaluating Social Programs with Endogenous Program placement and selection of the treated", *Handbook of Development Economics*, Volumen 4, Capítulo 60.