

LA EVALUACIÓN DE IMPACTO EN POLÍTICAS PÚBLICAS: REFLEXIONES A PARTIR DE UN CASO DE EVALUACIÓN DE UN PROGRAMA DE REINSERCIÓN SOCIAL EN CHILE

RODRIGO SALAS PORTUGUEZ¹

rodrigosalas@iap.uchile.cl

DIEGO PIÑOL ARRIAGADA²

diego.pinol@iap.uchile.cl

Instituto de Asuntos Públicos
Universidad de Chile

Recientemente, en Latinoamérica han proliferado estrategias de intervención cuyo foco está centrado en la reinserción social de personas que producto de conflictos con la ley. El presente documento contiene una serie de reflexiones metodológicas surgidas con ocasión de la implementación de una evaluación de impacto de un programa de reinserción social en Chile. Para ello se enuncian un conjunto de condiciones requeridas de manera que una evaluación de impacto permita obtener conclusiones robustas respecto de su desempeño. Por otro lado, se establecen los desafíos específicos a superar para la evaluación de estos programas comenzando por una adecuada conceptualización y posterior operacionalización del concepto de reinserción, identificando finalmente la aplicabilidad de distintos diseños de evaluación de acuerdo a las características específicas de cada programa Evaluado.

1 Administrador Público, Master of Science in Social Research Methods y Magíster en Economía Financiera. Director del Centro de Estudios en Seguridad Ciudadana de la Universidad de Chile.

2 Sociólogo, Doctor (c) en Psicología. Coordinador de Investigación del Centro de Estudios en Seguridad Ciudadana de la Universidad de Chile.

Palabras clave: Reinserción, Metodologías de evaluación de impacto, Reincidencia, Seguridad Ciudadana.

THE IMPACT EVALUATION ON PUBLIC POLICIES: REFLECTIONS FROM THE EVALUATION OF THE SOCIAL REINSERTION PROGRAM IN CHILE

Recently in Latin America an important number of programs focused on social reinsertion have been developed and carried out for people with conflicts with the Law. The present article covers a sequence of methodological considerations emerging from the evaluation carried out on a reinsertion program designed and applied in Chile. For that purpose, a group of conditions necessary to meet to achieve robust evidence of program performance is individualized. On the other hand, a group of specific requirements for the evaluation of reinsertion programs is pointed out, beginning with some definitions of reinsertions and different approaches that can be used to operationalize that concept verifying finally the applicability of different evaluation designs according with the specific characteristics of each evaluated Program.

Keywords: Reintegration, Impact Evaluation Methodologies, Reoccurrence, Citizen Security.

INTRODUCCIÓN

Las evaluaciones de impacto de políticas públicas se configuran como una necesidad en un contexto donde los programas utilizan los escasos recursos de los que dispone un Estado para brindar una atención a los problemas de un país. Las evaluaciones de impacto permiten determinar si la inversión económica y técnica puesta en una intervención amerita su continuidad, o bien si es necesario buscar alternativas de solución del problema que justifica el programa. A través de un método riguroso, una evaluación de impacto permite informar a los diseñadores y entidades que financian la implementación de políticas públicas si un programa está generando los efectos que promete en su diseño, siendo la única modalidad robusta capaz de establecer, más allá de cualquier duda razonable, si la intervención logra cumplir los objetivos que se ha planteado.

En el siguiente artículo analizaremos y reflexionaremos sobre el proceso de evaluación de impacto de un programa de reinserción social en Chile, de modo de extraer lecciones para futuras evaluaciones en este campo. Siendo en muchos países de América Latina una demanda de la ciudadanía la búsqueda de soluciones concretas respecto del delito, los programas de este tipo han ido proliferando. Si a esto se suma el hecho que las cárceles en nuestra región viven graves problemas que disminuyen la probabilidad de generar procesos de cambio en los sujetos, resulta urgente determinar qué funciona y qué no funciona en la búsqueda por brindar una intervención a los sujetos privados de libertad que efectivamente contribuya a mejorar sus probabilidades de reinserción social.

El presente artículo ha sido concebido como un ensayo de reflexión metodológica, con la intención de generar procesos de análisis respecto de las condiciones que deben cumplirse para realizar este tipo de evaluaciones así como sus limitantes en la evaluación de este tipo de programas, en el contexto regional. Es una invitación a una reflexión metodológica que tiene por finalidad mejorar los procesos de evaluación de las políticas públicas.

1. LA TEORÍA DE LA EVALUACIÓN DE IMPACTO: CONSIDERACIONES PARA SU IMPLEMENTACIÓN

Teoría y supuestos a la base

El objetivo central de cualquier evaluación de impacto es la estimación del efecto de una intervención en los sujetos que la reciben. El primer paso para la consecución de dicho objetivo es lo que se denomina en la literatura como el problema fundamental de la estimación causal (Holland & Rubin, 1988), que se expresa en el reconocimiento del hecho de que es imposible observar en un mismo sujeto, en un mismo momento del tiempo, dos valores para una misma variable: el valor de un indicador de impacto de un programa dado que un sujeto ha recibido tratamiento versus el valor que dicha variable habría tenido para el mismo sujeto si es que no hubiese recibido el tratamiento. Este problema se expresa de la siguiente forma en evaluación de programas: no es posible observar en un mismo sujeto las condiciones de “tratamiento” y “control”. La inferencia causal requiere la generación de un escenario ficticio en el que se pueda estimar lo que hubiese ocurrido a un sujeto si es que no hubiese recibido tratamiento, esto es, se busca un *contrafactual*.

Existen múltiples estrategias para configurar escenarios contrafactuales, pero todas ellas coinciden en un aspecto central: la estimación de los impactos a nivel grupal en lugar de a nivel individual. Esta estrategia no sólo permite generar el escenario contrafactual a partir del cual se pueda efectuar inferencia causal, sino que además es consistente con el requerimiento de la evaluación de impacto de programas sociales en el sentido de que se espera que las conclusiones de la evaluación sean representativas de lo que sucedería si el programa se aplica en poblaciones mayores. No obstante, para poder realizar una adecuada estimación del efecto existe un conjunto de condiciones necesarias de cuyo cumplimiento depende significativamente la robustez y validez del ejercicio evaluativo. Dado que la indagación en los efectos de una intervención es un proceso demandante, que puede requerir esfuerzos significativos tanto de la institución a cargo como de quién evalúa, resulta fundamental verificar el cumplimiento de dichas condiciones. Éstas no dependen de la estrategia evaluativa, sino que más bien hace referencia a aspectos específicos del diseño e implementación de las intervenciones.

En primer lugar, resulta fundamental desde el diseño del programa que la teoría de cambio y la consecuente jerarquía de objetivos sea explícita. Se trata de que los diseñadores del programa identifiquen y declaren expresamente, con la mayor precisión posible, la manera en que a partir de las principales actividades contempladas, se procesan los insumos (recursos de distinta naturaleza) de manera tal de generar los productos (bienes u o servicios) que el programa define entregar a sus usuarios, cómo el que los usuarios al apropiarse y utilizar dichos productos permite que el programa alcance su propósito, y finalmente cómo dicho propósito se conecta con objetivos de política pública de mayor jerarquía (fin)³. Sin una descripción detallada de los mecanismos de causalidad que vinculan acciones a productos, y los productos al cumplimiento de los objetivos propuestos, no resultaría posible atribuir cualquier efecto constatado a la implementación del programa. De este modo, la evaluación de impacto permitirá corroborar la causalidad del programa en efectos observados.

En la misma línea, ya desde el diseño se debería individualizar y operacionalizar con la mayor precisión posible los indicadores que dan cuenta del cumplimiento de los objetivos propuestos, así como los insumos de información que se utilizarán para su cálculo. En el caso específico de las evaluaciones de impacto resulta imprescindible la operacionalización de indicadores que permitan verificar el cumplimiento del objetivo general o propósito. La correcta identificación de dichos indicadores contribuye no sólo a la generar las condiciones para una estimación de los efectos, sino que también permite mejorar la consistencia del diseño. Resolver este aspecto nos permitirá determinar cuáles son los indicadores asociados a la inferencia de casualidad, o sea, qué mediremos con la evaluación de impacto.

Respecto a la implementación, la literatura especializada en materia de evaluación de programas sociales plantea un conjunto de requerimientos (P.J. Gertler, Martínez, Premand, Rawlings, & Christel M. J., 2009; Khandker, B. Koolwal, & Samad, 2010; Wolff, 2000 entre otros). Estas condiciones pueden agruparse en

3 Sin perjuicio del uso sistemático que organismos internacionales y entidades gubernamentales y/o no gubernamentales hacen de instrumentos como la Matriz de Marco Lógico para explicitar la teoría de cambio subyacente a las intervenciones sociales en las que participan, para efectos del presente artículo no resulta relevante el formato en que se presente dicha descripción, sino que más bien ésta se efectúe adecuadamente sin importar la modalidad específica mediante la cual se haga.

tres categorías distintas: condiciones respecto a la fidelidad de implementación, condiciones respecto a la equivalencia entre el grupo tratamiento y el grupo control, y condiciones respecto a la homogeneidad del entorno en que los programas son implementados.

Respecto a la fidelidad en la implementación, más allá de los requerimientos ya enunciados en relación con el diseño en primer lugar se espera que las acciones contempladas a realizar con los usuarios sean detalladas con un alto nivel de especificidad y que éste sea materializado expresamente en las acciones efectivas desarrolladas con los usuarios. En tal sentido, ya sea porque el diseño no contempla en detalle el contenido de las intervenciones o bien porque no se puede verificar efectivamente su realización, la calidad de las evaluaciones de impacto se deteriora significativamente de no cumplirse estas condiciones. La factibilidad de cumplir con esta condición se ve afectada en la medida en que conceptos como “gestión de casos”, “medición de riesgo”, “intervención multidisciplinaria” y otros similares no han sido adecuadamente operacionalizados en el diseño (Wolff, 2000). En la misma línea, en la medida en que las intervenciones son más complejas y, en consecuencia, aumenta la cantidad y diversidad de profesionales a cargo de éstas, se requiere que las competencias, habilidades e incluso los niveles de motivación por el trabajo de los equipos sean no sólo detallados con precisión en el diseño del programa, sino que estos elementos sean debidamente acreditados/verificados previo a la implementación del programa. Posteriormente se debe verificar permanentemente el cumplimiento de los aspectos detallados dando prioridad número uno a los sistemas de control de gestión dispuestos para monitorear la implementación.

Con respecto a la capacidad del programa de establecer un grupo de control que permita evaluar adecuadamente los logros de dicho instrumento, el hecho de que los programas intervengan poblaciones con problemas múltiples que ocurren simultáneamente y que para muchos de ellos no existan operacionalizaciones satisfactorias también es un elemento que complica la estrategia evaluativa. Es por lo anterior que resulta fundamental que la población objetivo del programa sea definida con precisión a partir de atributos adecuadamente operacionalizados y posteriormente verificados en los sujetos que conformarán ambos grupos. De no contarse con instrumentos validados que permitan operacionalizar las definiciones operativas robustas de conceptos como el riesgo de reincidencia entre otros y que

dichos instrumentos sean adecuadamente aplicados, se reduce significativamente la posibilidad de verificar la equivalencia entre los grupos tratamiento y control generados con el propósito de evaluar un programa. En la misma línea argumental, los indicadores que dan cuenta del logro de los objetivos planteados también deben estar adecuadamente operacionalizados y validados.

Finalmente, las características del entorno en que se desarrollaran los programas pueden influir severamente en su desempeño y, por lo tanto, es necesario que el plan declare expresamente las condiciones bajo las cuales se deberá implementar, así como también se verifique el cumplimiento efectivo de dichas restricciones durante el desarrollo de las intervenciones en todos aquellos ambientes en que el proyecto se desarrolló. En tal sentido, es fundamental tener en consideración que las condiciones en que se implementan los programas sociales distan significativamente del control que se puede ejercer de las variables intervinientes como se podría lograr en un laboratorio. Las condiciones a verificar no sólo dicen relación con los elementos físicos en donde las intervenciones se desarrollan sino que también con los niveles de capital social existentes antes y durante las intervenciones y la presencia de otras organizaciones que desarrollan intervenciones en paralelo a las intervenciones evaluadas. De no ser equivalentes (en promedio) los atributos del entorno en que se desenvuelven los sujetos que forman parte de los grupos tratamiento y control la calidad de las conclusiones del proceso evaluativo puede deteriorarse significativamente.

Otro tipo de riesgos en la conformación de grupo y la estimación causal refiere a los efectos heterogéneos del tratamiento o intervención del programa, ya que la evaluación de impacto supone que éste influye de manera similar en todos los miembros del grupo de tratados.

Adicionalmente, más efectos no deseados se enuncian a continuación:

- El efecto Hawthorne, cuando el mismo procedimiento de evaluación genera un comportamiento diferente en los sujetos (tratados y/o controles).
- El efecto John Henry, cuando el grupo de control se esfuerza más para compensar el hecho de no haber sido elegido como parte del programa.

- La anticipación, que es cuando los sujetos empiezan a modificar sus conductas antes de recibir el tratamiento.
- El sesgo por sustitución, que se configura cuando el grupo de control busca alternativas al programa para poder recibir atención.
- El efecto derrame, cuando los efectos de los tratados se traspasan al grupo de control por la interacción social entre ambos, o bien porque el programa modifica el contexto y genera un cambio en ambos grupos.

Otro aspecto a considerar en las evaluaciones es que el impacto del programa o el efecto en la población objetivo puede requerir cierto espacio de maduración para que se alcance plena efectividad, lo que supone determinar con precisión estos tiempos, además de generar una estrategia de seguimiento y de mediciones temporales en los sujetos pertenecientes a ambos grupos.

1.1 Inferencia Causal en la Evaluación de Impacto

El ejercicio práctico de la evaluación de impacto requiere de la conformación de un grupo de control y uno de tratamiento, los cuales deben ser estadísticamente idénticos de tal manera que las diferencias que existen entre ellos en las variables que el programa definió como resultado, puedan ser atribuidos a la intervención. Habitualmente en la literatura se observan estudios que se denominan de impacto pero los supuestos impactos son estimados sin contar con grupos de control o utilizando “falsos grupos” de control, tales como evaluaciones reflexivas (antes después) o bien utilizando a quienes no desean participar como controles de quienes sí participan (sesgo de autoselección; P. Gertler, Martínez, & Premand, 2017). Cuando se conforma un grupo control adecuadamente la inferencia causal se sustenta en la comparación del valor promedio de la variable de interés definida por el programa (indicadores de propósito o resultado) en el grupo que recibió tratamiento versus el valor de la misma variable en otro que no lo ha recibido. Esta comparación se denomina Efecto Promedio del Tratamiento o ATE por sus siglas en inglés⁴. Resulta clave la constitución de ambos grupos, siendo el principal requisito que ambos sean idénticos tanto en aquellas variables directamente observables como en aquellas que no lo son pero que podrían influir en el efecto del programa.

⁴ Average Treatment Effect.

La mejor manera de alcanzar esta condición es mediante la asignación aleatoria de los tratamientos, procedimiento utilizado en medicina ya desde 1747 (Dunn, 1997) y paulatinamente incorporado como herramienta en la investigación en ciencias sociales. Bajo este procedimiento denominado genéricamente diseños experimentales o Ensayo Controlado Aleatorizado (P. J. Gertler et al., 2009) o RCT por sus siglas en inglés,⁵ la asignación aleatoria de los tratamientos a sujetos representativos de la población objetivo de un programa permite generar dos grupos estadísticamente idénticos de manera tal que efectuar estimaciones ATE resulta trivial.⁶

Sin embargo, incluso en un escenario de evaluación mediante RCT pueden presentarse problemas que contaminan la estimación de los efectos. El primero de ellos surge de la posibilidad de que los sujetos podrían no respetar la condición de asignación resultante del RCT. Por ejemplo, algunos de los sujetos asignados a tratamiento podrían rechazarlo o eventualmente, sujetos asignados al grupo control podrían acceder a tratamiento.⁷ En este escenario la estimación de efectos podría efectuarse sin verificar si quienes fueron asignados al grupo tratamiento o al grupo control accedieron o no al tratamiento de acuerdo a la aleatorización. Esto es, se compararía el valor promedio de la(s) variable(s) objetivo del programa entre todos aquellos a los que se asignó tratamiento con el valor promedio de la(s) misma(s) variable(s) entre todos aquellos a los que no se asignó tratamiento, a pesar de que en ambos grupos podría haber sujetos que no cumplieron la asignación (sujetos del grupo tratamiento que no lo recibieron y sujetos del grupo control que accedieron). Esta modalidad de estimación se denomina Intención de Tratar o ITT por sus siglas en inglés.⁸ Considerando el hecho de que en no pocos programas sociales la participación es voluntaria (independiente de la posibilidad de una asignación aleatoria), las estimaciones ITT pueden aportar sentido a las estimaciones de efecto puesto que esta estimación constituye lo que en la literatura se denomina un ATE con cumplimiento imperfecto (Gerring, 2011).

5 Randomized Controlled Trial.

6 Un simple test t de diferencia de medias y/o proporciones según sea la o las variables objetivo es suficiente para estimar de manera robusta el efecto de un Programa cuando se ha implementado RCT.

7 Este problema se denomina Noncompliance.

8 Intention to Treatment.

Por cierto, desde otra perspectiva resultaría relevante el análisis de efectos sólo entre los sujetos que efectivamente respetaron su condición de asignación a tratamientos o controles, retirando de los cálculos a sujetos que fueron asignados a tratamiento y que por distintas razones no lo recibieron o sujetos que fueron asignados al grupo control pero que se las ingeniaron para recibirlo. Esta modalidad de estimación se denomina Efecto Promedio en los Tratados o ATT por sus siglas en inglés⁹. Las estimaciones ITT y ATT serían idénticas en situaciones de cumplimiento perfecto de la regla de asignación. Sin embargo, el hecho de que en programas de inscripción voluntaria los sujetos podrían abstenerse de tomar parte en las intervenciones deteriora la calidad de las estimaciones puesto que, eventualmente, los atributos de los sujetos que determinan su decisión de inscribirse también podrían condicionar el efecto del programa. Esta situación se denomina sesgo de autoselección.

La literatura plantea que esta situación puede mitigarse en la medida en que, mediante el uso de distintos procedimientos estadísticos, se pueda estructurar un grupo de control que comparta en promedio los mismos atributos que los sujetos del grupo de tratamiento. En cualquier caso, se trataría de una búsqueda de similitud sobre la base de atributos observables, por lo general contenidos en bases de datos que permiten caracterizar a los sujetos a partir de atributos observables ya medidos más no sobre la base de aquellos atributos no observables o que no están incorporados en la base de datos que permitiría estructurar el grupo control. Sin embargo, aun cuando la base de datos disponible permitiese lograr un grupo control “idéntico” a los tratados, la autoselección ocasiona que los efectos estimados de esta forma (ATT) a lo sumo puedan considerarse como efectos locales, esto es, efectos que sólo pueden extrapolarse a sujetos con voluntad de inscribirse y no necesariamente a toda la población objetivo de un programa. Esta modalidad se denomina Estimación Local de Efectos del Tratamiento o LATE por su sigla en inglés¹⁰ (P. Gertler et al., 2017). Es por ello que el sesgo de autoselección constituye uno de los principales problemas a mitigar en el contexto de la evaluación de programas.

9 Average Treatment on the Treated.

10 Local Average Treatment Effect.

Finalmente, en el caso en que dadas las características del programa o el escenario factual de su implementación, ya sea que el grupo tratamiento o el grupo control queden reducidos a una fracción de aquellos para los cuales se contempló preliminarmente tratamiento y aquellos que no (tal como ocurre en los casos en que se estiman efectos por la vía de la modalidad ATT con un grupo control ad hoc), las estimaciones de efecto son sólo extrapolables a sujetos similares a aquellos que se consideran para la estimación de efectos. Esta modalidad constituye otro ejemplo de estimaciones LATE. En otras palabras, el ATT es una LATE en que el incumplimiento se produce en el grupo tratamiento.

1.2 Diseño de evaluación de impacto

Tal como se ha relatado, las características del diseño y las condiciones bajo las cuales se implementa un programa determinan en gran medida no sólo la calidad de las estimaciones sino que además constituyen los principales elementos de juicio que señalan la estrategia metodológica más adecuada para cada situación en particular. Sin perjuicio de que los diseños experimentales mediante la modalidad RCT para la asignación de sujetos al grupo tratamiento y al grupo control constituyen la modalidad más robusta debido al carácter estadísticamente equivalente en atributos observados y no observados entre el grupo tratamiento y el grupo control, ya sea por las características específicas de los programas y/o de su implementación, en varias ocasiones esta estrategia no puede ser implementada. Tal es el caso por ejemplo de programas en que desde su diseño se propone como de “cobertura universal”, o de programas en los que al no haberse implementado aleatorización en la asignación de tratamientos desde el comienzo de las intervenciones, debido por ejemplo a razones éticas (Goldstein et al., 2018). Para estos casos en la evaluación de programas sociales se ha desarrollado un conjunto adicional de técnicas denominadas genéricamente como diseños cuasi-experimentales. En estos procedimientos el objetivo central es estructurar un grupo control ad hoc lo más equivalente posible al grupo tratado de manera de implementar estimaciones del Efecto Promedio Tratamiento ATE.

En lo siguiente se detallarán las tres estrategias más utilizadas para estimar el ATE y conformar los grupos de comparación.

PROPENSITY SCORE MATCHING (PSM)

Este método consiste en que, una vez determinados mediante los procedimientos establecidos por el programa para la selección de los sujetos que recibirán las prestaciones contempladas en el diseño, se busca sujetos similares a los tratados pero que en principio no recibirán tratamiento. Para ello es imprescindible contar con una base de datos con la mayor cantidad posible de atributos medidos tanto para el grupo que recibirán tratamiento como del que no, de manera tal de identificar de entre estos últimos aquellos más parecidos a los tratados. La similitud entre los tratados y los candidatos a ser parte del grupo control se determina sobre la base de las variables registradas en la base de datos en un momento previo a la intervención (línea base). Puesto que idealmente se busca identificar la similitud sobre la base de la mayor cantidad posible de atributos, y que por lo tanto se trataría de una comparación multidimensional, por razones prácticas la cercanía entre los sujetos de ambos grupos se cuantifica unidimensionalmente, estimando para cada uno de los sujetos incluidos en la base un puntaje único. Por regla general, la probabilidad de participar en el programa es estimada mediante modelos de regresión de variable dependiente dicotómica (logit o probit), en que las variables explicativas son los atributos de los sujetos contenidos en la base de datos, calculando a partir de ello un Propensity Score para cada sujeto tratado y cada control.

Una vez obtenidos los Propensity Score, para cada sujeto tratado se puede identificar a uno o más sujetos candidatos para ser incluidos del grupo control (procedimiento de matching propiamente tal). Si se trata sólo de un sujeto, la modalidad de selección se denomina Vecino Más Cercano (*nearest neighbor*), pudiendo también utilizarse varios controles por cada tratado mediante técnicas de matching tales como Caliper, Radius o procedimientos Kernel,¹¹ siendo la razón principal del uso de varios controles por cada tratado el aumentar los tamaños muestrales y con ello el poder estadístico de las estimaciones. La calidad del procedimiento se juzga esencialmente verificando que existe un grupo suficientemente grande de sujetos en ambos grupos con puntajes similares (verificación de soporte común) y que para distintos valores del score calculado (Eg: deciles de score) los sujetos

11 Mayor detalle sobre estos y otros aspectos de la técnica de matching puede ser encontrada en Caliendo & Kopeinig, 2008 y P.J. Gertler et al., 2009 entre otros.

en ambos grupos sean idénticos en las variables utilizadas en el procedimiento de pareo (verificación de balance). Una de las ventajas principales de este método es que la estimación de efectos puede efectuarse sólo por medio de mediciones *ex post* de los indicadores de efecto definidos por el programa, no requiere de controles de línea base de las variables de efecto del programa. Sin perjuicio de lo anterior, el método suele implementarse contrastando la diferencia entre el valor de línea base y *ex post* de los indicadores de resultado/impacto definidos para el grupo tratamiento con la misma diferencia calculada para los sujetos participantes del grupo control, esto es, se mezcla *matching* con diferencias en diferencias, procedimiento que mejora la robustez de las estimaciones.¹²

Sin embargo, más allá de la posibilidad de lograr un grupo control altamente similar al de tratados, dos son las principales debilidades del método. En primer lugar, las estimaciones de efecto sólo se efectúan considerando sujetos con score similares, singularidad que implica que de no haber suficiente soporte común las estimaciones serían sólo extrapolables a una población de dichas características, más no a sujetos con puntajes extremos (aquellos con muy poca o muy alta probabilidad de participar), razón por la cual la estimación ideal de efectos (ATE) sería a lo sumo local (LATE). Finalmente, la mayor debilidad del método consiste en que requiere de un supuesto fundamental denominado independencia condicional, el que implica que dadas las variables utilizadas para el pareo, éstas no son afectadas por el tratamiento mientras los *outcomes* potenciales son independientes de la asignación del tratamiento. En otras palabras, se asume que la asignación de tratamiento se efectúa sólo sobre la base de las variables observables utilizadas en el procedimiento de pareo y que todas las variables que influyen dicha asignación y los *outcomes* definidos por el programa son observables e incluidas en el procedimiento de pareo. De esta forma, si otra(s) variable(s) no observables y por lo tanto no utilizables en el procedimiento de pareo determinan la asignación de tratamiento y/o los *outcomes* del programa, las estimaciones podrían invalidarse por la eventual presencia de sesgos (Eg: sesgo de autoselección).

12 La estimación de efectos mediante diferencias en diferencias puede efectuarse por separado a *matching* así como también en conjunto con otros diseños cuasi experimentales. Mayor detalle de aspectos puede encontrarse en (P. Gertler et al., 2017; Khandker et al., 2010).

REGRESIONES DISCONTINUAS

Un segundo diseño cuasiexperimental para la conformación de grupos control se denomina regresión discontinua. Este procedimiento se utiliza principalmente en programas que seleccionan sus usuarios sobre la base de un puntaje; por ejemplo, si los programas determinan sus usuarios sobre la base de un indicador como ingresos, vulnerabilidad, riesgo u otros constructos creados para sintetizar en un solo índice un conjunto de atributos relevantes para la focalización del programa. De esta forma, el diseño del programa establece que todos aquellos sujetos con puntaje inmediatamente superior al punto de corte establecido, reciben el tratamiento. El supuesto fundamental que subyace a este diseño es el hecho de que sujetos que no reciben tratamiento pero que tienen valores cercanos al punto de corte, no deberían ser muy distintos a los sujetos participantes con puntajes inmediatamente por sobre el punto de corte, por lo que los primeros podrían ser un buen grupo control de los efectivamente tratados.

La gran ventaja de esta técnica es que la comparabilidad entre grupos estaría asegurada desde un principio (en los sujetos en las cercanías del punto de corte), por lo que a priori existe menor necesidad de información de línea base de ambos grupos que en matching. Sin embargo, el diseño cuasi experimental de regresión discontinua tiene su principal debilidad en que, dado que en la estimación de efectos se considera sólo a aquellos sujetos del grupo de los tratados y de los no tratados, los efectos estimados son locales (LATE), esto es, no extrapolables a sujetos en con puntajes significativamente inferiores o superiores al punto de corte establecido.

VARIABLES INSTRUMENTALES Y PROMOCIÓN ALEATORIA

Finalmente, un tercer diseño cuasi-experimental candidato a utilizarse en la evaluación de programas sociales se denomina método de Variables Instrumentales. Por lo general el método es utilizado en programas de cobertura universal o en situaciones en que la decisión de participar en el programa es voluntaria, razón por la cual el evaluador no está en condiciones de asegurar el total cumplimiento de la regla de asignación si es que se implementase un diseño de aleatorización en la distribución de los cupos del programa. Ahora bien, el incumplimiento en la regla de asignación también podría ocurrir por parte de sujetos que producto

de un diseño experimental sean asignados al grupo control pero que de alguna forma se las ingenian para recibir tratamiento. El ejemplo típico de esta situación está constituido por aquellas intervenciones consistentes en asistencias técnicas y/o capacitación en que los sujetos no tratados pueden arreglárselas para acceder a los contenidos y material bibliográfico entregado por el programa a sus participantes y en función de ello podrían, por ejemplo, incrementar sus conocimientos y eventualmente su empleabilidad, a pesar de que en el diseño de la evaluación no deberían haber recibido tratamiento.

Una solución para esta situación sería estimar el efecto sólo entre quienes respetaron la regla de asignación (por ejemplo ATT), lo cual claramente produciría estimaciones locales (LATE),¹³ pero dado que hay muchos programas que en la práctica no pueden obligar a ninguno de los sujetos de la población objetivo a recibir el tratamiento contemplado en el diseño, las estimaciones LATE son representativas del efecto del programa en un universo amplio de beneficiarios.

El método de Variables Instrumentales introduce una fuente externa de variación que permite influir en la decisión de participar de los sujetos que componen la población objetivo del programa. Uno de los mecanismos más utilizados como variable instrumental para programas de cobertura universal se denomina *Promoción Aleatoria*. Este diseño es muy similar a los modelos RCT siendo la principal diferencia entre ambos el hecho de que en lugar de aleatorizarse la asignación a tratamiento o a control se aleatoriza la entrega de un mecanismo de promoción (incentivo) diseñado con el único fin de incidir en la decisión de quienes lo reciben aumentando su probabilidad de inscripción. Lo esencial del dispositivo de promoción aleatoria es que cualquiera sea el mecanismo de incentivo seleccionado éste sea efectivamente exógeno al programa y que además no afecte de ninguna forma los indicadores de efecto establecidos por el programa. Por ejemplo, en el caso de un plan diseñado para aumentar los ingresos de las personas, la entrega de bonos en efectivo a sujetos que aleatoriamente fueron asignados a recibir la promoción sería una estrategia errónea ya que dichos bonos podrían tener incidencia sobre el

13 La modalidad de estimación de efectos Average Treatment on the Treated (ATT) es sólo un caso particular de Local Average Treatment Effect (LATE) en que el incumplimiento ocurre entre los asignados a tratamiento.

ingreso de los sujetos. Es por ello que en la práctica se suele utilizar un abanico amplio de promociones que van desde campañas publicitarias focalizadas, cartas de invitación personalizadas, vouchers,¹⁴ etc.

2. EVALUACIÓN DE IMPACTO EN PROGRAMAS DE REINSERCIÓN: EL CASO DEL PRS

El Programa de Reinserción Social de Gendarmería de Chile es una intervención a la población penal chilena, que supone un conjunto de intervenciones con la población del programa con el fin de disminuir la reincidencia de los sujetos intervenidos. Este programa se basa en los principios y enfoque metodológico del programa de Riesgo Necesidad y Capacidad de Respuesta (*Risk-Need-Responsivity* en inglés), modelo generado por Andrews y Bonta entre los años 1990 y 2000 (Andrews, Bonta & Wormith, 2011), y que busca afectar una serie de variables que según los autores son las de mayor efecto en la reincidencia.

Este programa tiene una serie de requisitos de acceso o de selección de participantes, siendo el más relevante el hecho que el sujeto (posterior a un diagnóstico) tenga un riesgo de reincidencia medio o alto, para lo cual se les aplica una escala de riesgo de reincidencia y se clasifica a los sujetos según el puntaje obtenido en riesgo bajo, medio o alto. La población con bajo riesgo de reincidencia no participa del programa porque de acuerdo con la teoría base, la intervención podría producir un efecto negativo en ellos (aparte del hecho que no la necesitan).

En lo siguiente analizaremos los aspectos teóricos antes enunciados respecto de la evaluación de impacto, de modo de generar una contribución al debate sobre la evaluabilidad o factibilidad de realizar evaluaciones de impacto a este tipo de intervenciones.

Un primer aspecto problemático de dicha evaluación es la **variable definida como resultado para el programa**, ya que si bien éste está destinado a disminuir la

14 Dependiendo de los objetivos del programa, el uso de vouchers monetarios puede ser recomendable como estrategia de promoción. Tal puede ser el caso de programas con objetivos en materia de salud.

reincidencia de los sujetos tratados, el nombre y una serie de documentos del programa utilizan el concepto de reinserción social.

Esto hace referencia a un problema más general, pues cuando nos enfrentamos con la necesidad de estimar el efecto que una intervención en el ámbito penal se tienden a utilizar dos conceptos, reincidencia y reinserción, que a menudo se utilizan como sinónimos o como términos similares, pero que refieren a dos temas absolutamente distintos entre sí, no obstante poseen ciertos puntos de encuentro o intersección.

La reincidencia consiste en la ocurrencia de un delito o de una conducta transgresora (luego analizaremos la diferencia entre ambos), de una persona que ha sido juzgada y condenada con anterioridad (CESC, 2012). Dado que las ofensas o conductas criminológicas pueden o no ser conocidas por el sistema penal, se establece una diferenciación entre aquellos hechos que han sido juzgados y condenados de aquellos que no han resultado en la aprehensión, juzgamiento y condena de las personas que los han cometido. Aquellos no conocidos por el sistema penal (que no han resultado en una detención, juzgamiento o condena) se denominan reiteración, mientras aquellos que sí han sido juzgados y condenados, son reincidencia.

Para efectos del análisis criminológico, el que estos hechos sean o no conocidos por el sistema penal (incluida la policía) no tiene ningún interés particular, ya que su materia de interés es la conducta transgresora.

Para efectos de un análisis de las consecuencias de un programa en esta materia, la reiteración presenta dos problemas: la baja confiabilidad del autoreporte de los sujetos (al tener que reconocer una persona que ha cometido una infracción, o sea, declararse culpable de un hecho no conocido por el sistema penal), y la inexistencia de registro formales de este tipo de conductas (ya que no han sido detectadas por algún órgano del Estado). La reiteración sólo podría ser estimada a partir de la generación de un vínculo entre quien recolecta y quien entrega la información, de modo que se genere un ambiente de confianza mutua para que el sujeto reconozca dichos actos, algo que en casos de programas de seguimiento post-condena podría ser factible, no obstante el carácter institucional de dichas estrategias será un impedimento.

La reincidencia, por su parte, también presenta complejidades: primero, descansa en la capacidad del sistema penal de detectar conductas que han infringido la ley, por lo que depende de la gestión policial; segundo, depende que dicha conducta haya sido procesada, por lo que depende de la eficiencia del sistema persecutorio, que tiene a su vez una serie de factores (entre los que está la calidad de la evidencia como los criterios de los actores del sistema penal, entre otros aspectos); tercero, depende que dicha acción haya sido condenada, o sea, que se haya probado la culpabilidad del sujeto y además que la consecuencia penal sea una condena (y no una salida alternativa), por ende, depende tanto del marco legal como de la forma de actuar de los agentes involucrados en el sistema de justicia.

La principal ventaja que tiene la reincidencia como indicador es que es una acción que los tribunales han probado que el sujeto ha cometido (no depende del autoreporte), y supone que en dicho juicio se han cuidado todos los derechos concurrentes (por lo que hay garantías de confiabilidad de la resolución o condena dictada), por lo que la infracción ha sido probada de forma confiable. Además, la literatura es clara en señalar que los hechos que sólo han sido detectados por las policías se denominan re-arrestos o re-detenciones, y que los hechos detectados y sometidos a proceso son nuevos procesamientos o juicios. A estas categorías se suma el re-encarcelamiento, que corresponde a aquel subgrupo de sujetos que han sido condenados a penas de cárcel, dejando fuera las condenas no privativas en libertad (que por ejemplo en Chile constituyen la mayor parte de la población penal).

La reincidencia, entonces, es una reiteración de la conducta delictiva que ha sido conocida por el sistema penal y en que se ha probado la culpabilidad del infractor. Usando la terminología ya enunciada, se trata de una reiteración comprobada. No obstante, refiere a la naturaleza jurídica del comportamiento transgresor o infractor, por lo que para la criminología éste es un conjunto de hechos muy limitado para comprender la conducta criminal. La reincidencia son conductas transgresoras conocidas, juzgadas y probadas por el sistema de justicia, no obstante, una serie de conductas no detectadas por dicho sistema también constituyen una conducta criminal.

Una corriente criminológica de gran difusión a nivel internacional es el riesgo de reincidencia, que ya no se ocupa de estimar las conductas cometidas por los sujetos, sino la propensión a cometer dichos hechos. Derivado de la comprensión causal y de la posibilidad de modificar la conducta criminal, se trabaja sobre la base de los factores de riesgo de reincidencia, que han sido identificados por investigadores como Andrews y Bonta entre otros (CESC-FPC, 2015), sobre la base del análisis de cohortes de sujetos que han cometido infracciones o transgresiones a lo largo de su vida, logrando establecer las variables que inciden en mayor medida en las probabilidades de cometer dicha conducta. En base al estudio de infractores de anglosajones (ingleses, canadienses, estadounidenses) ha definido un conjunto de variables que permiten predecir, con un alto nivel de confiabilidad, si los sujetos realizarán conductas transgresoras o no. Un mayor riesgo de reincidencia, bajo esta lógica, supone que las probabilidades de cometer una nueva infracción son más altas; y además, que la modificación de éstas (a través de un tratamiento o intervención) de un factor de riesgo disminuye las probabilidades que ello ocurra.

Las críticas a esta noción de la conducta infractora han provenido de diversas fuentes (en los mismos países que vieron nacer la teoría), siendo una de las corrientes teóricas de mayor difusión la teoría del desistimiento. La teoría del desistimiento no discute la presencia o no de factores de riesgo, sino que se ocupa de realizar un giro respecto de qué se debe hacer o intervenir para modificar la conducta. Para ello, analizando los mismos datos de trayectorias delictivas, se ha fijado en aquellos sujetos que modificaron dicho comportamiento y en qué elementos (factores, variables) explican este cambio. Esto supone pasar de trabajar sobre la base de los riesgos, hacia el trabajo en los precursores del cambio, por lo que la intervención se modifica completamente.

La reinserción, que no tiene una adecuada o concisa definición (en inglés el término más similar es reentry o reintegration), se alinea con esta teoría, en cuanto busca conocer el proceso que enfrentan los sujetos en su etapa de integración a la comunidad posterior al cumplimiento de una pena (Arellano, 2010). Si bien la reinserción no se circunscribe a la identificación de los factores que hacen desistir a los sujetos, se alinea con el giro epistemológico que se preocupa por cómo los

sujetos buscan generar un cambio en sus vidas, no obstante dicho cambio puede ser en ámbitos distintos de la comisión de infracciones o transgresiones. La reinserción por ejemplo, se ocupa de identificar los aspectos esenciales para reintegrarse en su familia, en su trabajo o estudios, en su comunidad, con sus amigos, entre otras áreas de su vida. La reinserción no circunscribe su análisis a la conducta infractora.

Ahora bien, ¿cómo se relaciona la conducta infractora o transgresora con la reinserción? Los procesos de reintegración a la vida en libertad son un desafío complejo, en especial cuando los sistemas de apoyo social (fundamentales para el éxito de esta etapa) no funcionan de manera adecuada, y dado que el contexto donde se desenvuelven los sujetos pueden aumentar la probabilidad que vuelvan a cometer una infracción a la ley, y que en caso de ser detectada por el sistema de justicia podría resultar en una reincidencia. Ante la existencia de una reincidencia, el proceso de reinserción social del sujeto se verá truncado o al menos retrocederá en algunas áreas que podrían haber presentado un avance, por lo que una reincidencia resultará en un efecto nocivo en su proceso de reinserción social. Ahora bien, esto no significa que la no-reincidencia sea reinserción y viceversa, ya que como se ha mencionado, podrían existir conductas transgresoras no detectadas por el sistema de justicia (reiteraciones) que igualmente podrían estar afectando su reintegración; o bien, que no existan conductas infractoras pero que el proceso de reinserción no haya tenido los resultados esperados o deseables. Por ende, la reincidencia afecta la reinserción social pero no la explica, como tampoco la reiteración es un indicador confiable respecto de si el proceso de reintegración ha tenido avances o no.

Dicho de otro modo, la reinserción social es un proceso complejo que busca establecer tanto los aspectos críticos como los facilitadores para una adecuada reintegración, en donde la transgresión o conducta infractora es una variable más en juego, una importante, pero sólo una. Además de la diferenciación conceptual, la reinserción social, tratándose de un proceso de cambio basado en una serie de aspectos potenciadores y facilitadores, se centrará en determinar qué variables posibilitan la reintegración, y no aquellos que disminuyen el riesgo de volver a cometer una infracción o no. Por lo que la reinserción social se perfila bajo el giro epistemológico del desistimiento, ocupándose de un entramado mucho más complejo de cambios que sólo la conducta infractora.

En el caso de la evaluación del PRS hay una opción teórico-metodológica por definir a la reincidencia como indicador de resultado, no obstante dado que se implementa en un contexto de política pública donde la reinserción social es la finalidad principal, sería deseable determinar el vínculo entre la reincidencia y el proceso de reinserción de los sujetos intervenidos, a fin de estimar un impacto más certero o preciso. En un marco más amplio este problema se asocia con la dificultad de operacionalizar de forma adecuada en los programas, ya que los conceptos determinan de forma directa qué se evaluará en una medición como la de impacto.

Relativo a este mismo concepto, la evaluación del PRS también se ve dificultada por el hecho que las escalas utilizadas para la asignación de los sujetos al programa no han tenido un proceso de validación riguroso, por lo que es muy factible que hayan sujetos que participan del programa pero no debieran estar en él.

Un segundo tema dice relación con la **calidad de la implementación del programa**, para lo que el ejemplo del programa PRS nos sirve para graficar aspectos claves de los procesos de evaluación de impacto. En efecto, la materialización efectiva del programa permite ejemplificar algunos de los problemas ya señalados, pero destacaremos tres de ellos como los más relevantes. El primer aspecto refiere al nivel de estructuración del programa, un aspecto muy sensible para una evaluación de impacto, dado que cada sujeto asignado al grupo tratamiento podría recibir prestaciones muy diferentes a las de otro asignado al mismo grupo. En el caso del PRS se ha hecho un diseño del programa suficientemente estructurado que permite minimizar o al menos controlar dicho aspecto.

Un segundo aspecto refiere a la capacitación del personal técnico a cargo, ya que en caso de no tener procesos de inducción y entrenamiento de los profesionales, dicho diseño puede no ser traspasado a las intervenciones con los usuarios. Este es un problema no sólo de fidelidad, sino de calidad, en cuanto cada prestación entregada en el marco del programa tenga un alto estándar de calidad. Este aspecto en el caso del PRS se aborda de forma pertinente, realizando una intensiva inducción a los profesionales tratantes, como también procesos de capacitación permanente respecto del modelo de trabajo, las metodologías de implementación, los procesos de evaluación, entre otros aspectos, lo que asegura que la fidelidad del programa sea la pertinente para este tipo de intervenciones.

Un tercer aspecto refiere a la supervisión de las intervenciones, ya que si bien el diseño y el entrenamiento de los profesionales se considera idóneo, no es sino a través de procesos de monitoreo donde se observa si las prácticas al interior del programa se alinean con lo estructurado en el diseño. Nuevamente esto es tanto un aspecto de fidelidad como de calidad, ya que la implementación por diversas razones puede tener dificultades y obstáculos en la práctica que deben ser subsanados. En el caso del PRS se constatan procedimientos de supervisión que pueden ser considerados pertinentes, no obstante resulta necesario que se maximicen los esfuerzos tanto para asegurar la calidad del programa, como que los procedimientos de registro puedan efectivamente contar con datos respecto de las intervenciones brindadas a cada sujeto. En consideración que el acceso a una u otra prestación depende tanto de factores del sujeto como del contexto, para efectos de una evaluación de impacto es necesario registrar minuciosamente el número y característica de las acciones en cada sujeto controlado.

Estos aspectos de implementación tienen incidencia directa en uno de los supuestos base de las evaluaciones de impacto, en tanto el tratamiento debe ser lineal y no heterogéneo. Si bien el caso del PRS parece tener una serie de aspectos abordados, aquellos no controlados (como el número de prestaciones recibidas por cada sujeto y el nivel de logro del Plan de Intervención Individual) se transforman en un riesgo para lograr realizar estimaciones de causalidad confiables.

En un tercer tema, **la conformación de los grupos de tratamiento y control** tuvo una serie de complejidades en el caso estudiado. Por una parte, dado que hay un proceso voluntario de adscripción, significa que un set de atributos de los sujetos determinan la decisión del individuo de inscribirse en un programa de reinserción (Ej: voluntad de reinsertarse). Estos atributos que podría determinar el que los resultados observados en el grupo de tratados no sean el producto de la intervención, sino que más bien el reflejo de los atributos de los individuos que les inclinaron a participar en el programa, generándose eventualmente un sesgo por autoselección. Por otro lado, dado que las personas que no son seleccionadas y forman parte teóricamente del grupo de control disminuyen sus probabilidades de obtener beneficios de salida anticipada de la cárcel, el hecho de no participar del grupo de tratamiento puede significar en un efecto grave a nivel psicológico.

A nivel de los efectos descritos en la literatura se pueden configurar una serie de aspectos problemáticos en la conformación de los grupos. Primero, dado los beneficios que supone formar parte del programa, es muy factible que previo a la intervención comience un proceso de cambio en los sujetos que no es aducible al programa, lo que se denomina efecto de anticipación. Segundo, dado que las personas que no son seleccionadas pueden tener acceso a otro tipo de oferta programática en sus penales, es muy factible pensar que el carácter instrumental que muchas veces se da en contexto carcelario a la participación en programas se logra por una vía distinta, como por ejemplo acceder a otros programas o buscar ayuda directa con los profesionales de los recintos penitenciarios, lo que se denomina efecto por sustitución. En tercer lugar, y vinculado a lo anterior, dado que los sujetos del grupo de control al tener una motivación inicial por reinsertarse (o salir en libertad, visto desde un carácter estrictamente instrumental) pueden realizar cambios en su comportamiento de forma autónoma sin la necesidad de acudir a otras prestaciones, lo que se denomina efecto John Henry. En cuarto lugar, aunque en un nivel menor de probabilidad, es el hecho que un número no determinado de personas que forman parte del programa pueden tener vínculos de parentesco y amistad con otras personas en un recinto penal, podría configurarse un efecto derrame hacia su grupo de referencia.

En atención a estas complejidades, la necesidad de conformar un grupo de control en ejercicio evaluativo del PRS, el contrafactual se conformó mediante la aplicación de la técnica de Propensity Score Matching, procurando controlar un conjunto amplio de variables buscando entre los no participantes a aquellos que compartieran en mayor medida los atributos de los tratados. No obstante el esfuerzo efectuado, dado que una gran cantidad de información sobre la población penal no tiene registro en bases de datos, la capacidad efectiva de control de estos efectos por la vía del Matching es una incógnita cuya respuesta no se puede verificar con la información disponible.

Es así como las estimaciones de causalidad en la evaluación de impacto del PRS, al no resolver una serie de complejidades desde la operacionalización de variables, pasando por la calidad de la implementación como por la conformación de grupos de control, adolecen de limitaciones significativas, lo que supone que el resultado

generado debe ser considerado con cautela, y sobre todo como un ejercicio preliminar de evaluación del efecto del programa sobre la reincidencia de los sujetos que en él participan.

3. EVALUABILIDAD DE LOS PROGRAMAS DE REINSERCIÓN: ABRIENDO LA DISCUSIÓN

Los programas diseñados para influir sobre la probabilidad de que personas que se encuentran en situación de cárcel reincidan en conductas ilícitas a su egreso se caracterizan por una serie de particularidades que determinan significativamente su evaluabilidad.

En primer lugar, si bien generalmente para la evaluación de este tipo de programas se consideran principalmente indicadores de reincidencia, esta aproximación resulta incompleta dada la multidimensionalidad del fenómeno en estudio por lo que se debe abordar el logro a nivel de reinserción social. Si bien en diversos países los programas de reinserción social han querido disminuir la complejidad de las distintas variables intervinientes focalizando su intervención en el componente criminológico y realizando intervenciones centradas en la modificación de la conducta infractora bajo la lógica de la disminución del riesgo de reincidencia, es una decisión que posee ventajas como desventajas. Las principales ventajas se vinculan con que la medición del riesgo de reincidencia, como de las intervenciones bajo este enfoque, vienen protocolizadas y estructuradas, por lo que no requiere (en perspectiva de los implementadores de estos programas) de un trabajo arduo de diseño ni de implementación, sino que ajustar la gestión de la intervención penal a esta estructura. La segunda ventaja es que, siendo apoyadas estas estrategias por un número importante de estudios a nivel internacional, la probabilidad de éxito es mayor; la tercera dice relación con un tema de costos, ya que si bien supone adquirir licencias y capacitar al personal para su puesta en marcha, es una inversión mucho menor a la de diseñar un programa propio; y la cuarta ventaja es que entrega legitimidad a la intervención penal, ya sea porque es una decisión basada en evidencia y porque la probabilidad de éxito es teóricamente alta, permite al sistema de ejecución tener a disposición argumentos en contra de ciertos cuestionamientos a su eficacia como efectividad.

Las desventajas refieren, primero, a que la intervención no logra ser integral, sino que focaliza su acción en aquellas variables conductuales, dejando de lado aquellos factores que no tendrían efecto sobre esta modificación conductual, con lo que se pierden posibilidades de contribuir al proceso (mucho más complejo) de reinserción social. En este sentido, los programas deben apelar a una intervención que aborde toda la complejidad del proceso de reinserción social, ya que el sesgo de focalizarse sólo en el aspecto conductual disminuye las probabilidades de dar sustentabilidad a los logros.

Una segunda desventaja es que no se realizan adecuados procesos de adaptación, tratándose de programas diseñados para una población distinta y para un contexto diferente, se necesita una metodología que entregue sensibilidad cultural a la intervención, buscando un adecuado balance entre ésta y la fidelidad al programa origen. Lo mismo tiende a suceder con los instrumentos de diagnóstico y para evaluar resultados de estos programas, que al no tener una adecuada validación, terminan entregando estimaciones erróneas y que tienen graves consecuencias en los usuarios y en los resultados buscados.

Una tercera desventaja se vincula a que los procesos de transportabilidad de los programas no tiende a ser realizado rigurosamente, lo que deriva en problemas de calidad de la implementación, calidad que además de depender de un programa estructurado de forma adecuada, depende tanto de aspectos organizacionales, de los implementadores como de los usuarios. La calidad de la implementación depende de observar, supervisar y mejorar los procesos de puesta en práctica de los programas, lo que no depende exclusivamente de tener un buen diseño, sino de contextos y actores participantes que pueden ser favorables u obstaculizadores.

En este sentido, una evaluación de impacto requiere de ciertas condiciones para su desarrollo, no sólo para que el ejercicio evaluativo sea factible, sino además para que la estimación casual genere insumos a los diseñadores de política pública. Al no haber analizado y controlado aspectos como los antes enunciados, el resultado obtenido de una estimación de impacto puede terminar generando más dudas que certezas, ya que se dificulta el análisis de qué aspectos del programa, del contexto y/o de los actores involucrados, pueden estar afectando los resultados (sean positivos o negativos).

En segundo lugar, una vez definida la población objetivo de un programa a ejecutarse al interior de un recinto penitenciario, por ejemplo, sujetos con un nivel de riesgo de reincidencia determinado, y/o que cumplen con ciertos requisitos específicos tales como cantidad de tiempo remanente de las condenas, buen comportamiento, entre otros; estos programas podrían contemplar cobertura universal, modalidad que en principio sería incompatible con la creación de un grupo control a priori que permitiese estimar efectos por la vía de diseños como RCT, matching o promoción aleatoria. Sin embargo, al menos dos hechos pueden hacer factible la creación de grupos control. El primero de ellos es que si la cantidad de individuos que constituyen la población objetivo y que cumplen con los criterios de elegibilidad establecidos por el programa supera la cantidad de cupos disponibles ello no sólo abre espacio a la obtención de un grupo control sino que más aún, siendo dicho grupo de sujetos homogéneos por tener los mismos niveles de riesgo de reincidencia y/o requisitos de elegibilidad, la aleatorización en la asignación de los tratamientos podría ser un criterio más justo de selección de los participantes sobre la base de otros criterios menos robustos en términos estrictamente metodológicos o derechamente subjetivos. De manera análoga, si los recursos disponibles para la ejecución del programa no permiten que se ejecute a gran escala a nivel nacional, en lugar de aleatorizar las intervenciones entre los sujetos al interior de un recinto la unidad de aleatorización podría ser precisamente los recintos, modalidad que por regla general de existir un número suficiente de establecimientos tratados y no tratados permite lograr estimaciones de efecto robustas.¹⁵

En tercer lugar, por lo general la decisión de participación de los reclusos en programas de reinserción suele estar asociada a la expectativa de obtener beneficios de distinta naturaleza durante el cumplimiento de las penas. Esta situación puede ocasionar dos problemas de distinta naturaleza respecto del procedimiento de evaluación. El primero de ellos es la eventualidad de sesgo de autoselección. Este

15 En principio, cada recinto debe considerarse como un cluster de sujetos. El hecho de que por lo general la asignación de reclusos a distintos recintos está determinada por atributos como la naturaleza de los delitos cometidos y/o las condenas, genera simultáneamente una homogeneidad intra recintos y una heterogeneidad entre recintos. Ambos fenómenos dificultan la comparabilidad necesaria para implementar la evaluación. Desde otro punto de vista, en la medida que la variabilidad de las variables consideradas en la evaluación es mayormente explicada por la heterogeneidad entre recintos que por heterogeneidad intra recintos (alta correlación intra clase), se requerirán mayor cantidad de recintos (clusters) para que las estimaciones sigan siendo robustas.

se produciría en caso de utilizar a sujetos que deciden voluntariamente participar como grupo tratamiento como control y a aquellos que deciden no hacerlo como controles. En efecto, a pesar de la expectativa cierta de obtener beneficios al participar en programas de reinserción, el hecho de que la voluntad de participar esté determinada por atributos personales de los sujetos podría sesgar significativamente las estimaciones de efecto, sobre todo cuando dichos atributos no son directamente observables. En efecto, si bien es cierto que sin importar el diseño de evaluación seleccionado, no sólo es posible sino que deseable que previo a la implementación del programa se verifique la comparabilidad entre los sujetos que recibirán tratamiento y aquellos que no, el hecho de que se confirme dicha comparabilidad sobre la base de atributos observables no excluye la posibilidad de que los sujetos no sean equivalentes en atributos no observables, problema cuyas consecuencias serían más relevantes en la medida en que dichos atributos estén asociados a la probabilidad al logro de los objetivos que los programas se proponen. En tal sentido si la única estrategia factible de implementar es matching, el equipo de evaluación deberá tomar todas las precauciones que sea posible de manera de generar las condiciones que permitan verificar que el supuesto de independencia condicional se cumpla.

El segundo problema que podría producirse debido al hecho de que la participación de los sujetos en los programas de reinserción puede estar fundada en la expectativa de acceder a beneficios de salida anticipada, es el hecho de que de implementarse diseños como RCT entre sujetos elegibles, aquellos que no fuesen asignados a tratamiento podrían en la práctica empeorar su comportamiento por el hecho de haber sido excluidos de la posibilidad de acceder a beneficios. Si tal fuese el caso, eventualmente la calidad del grupo control se vería afectada, lo que podría redundar en una sobreestimación de los efectos del programa, ya que la probabilidad de reincidencia al egresar podría ser mayor que la de los tratados, no por el hecho de haber recibido tratamiento sino que en respuesta a no haberlo recibido.

Por otro lado, en el caso de programas que seleccionan a sus usuarios sobre la base de un puntaje continuo, por ejemplo riesgo de reincidencia y un punto de corte definido para sobre dicho puntaje, la regla de selección determina la factibilidad de implementar el diseño de regresión discontinua, esto es, establecer grupos control y tratamiento incluyendo en ambos casos a sujetos con puntajes inmediatamente

por encima y por debajo de dicho punto de corte, partiendo del supuesto de que ambos grupos son comparables. Esta modalidad de estimación no requeriría la negación de tratamiento a ninguna persona que cumpla con la regla de asignación establecida. Más aún, sin perjuicio de que siempre resulta recomendable verificar la comparabilidad de los grupos tratamiento y control sobre la base de todas aquellas variables para las cuales se cuente con información ex ante para ambos grupos, y siempre y cuando los sujetos que dados los criterios de admisión no podrían participar en el programa no puedan manipular su puntaje de manera de acceder a tratamiento, las estimaciones de efecto por la vía de regresión discontinua no requieren del uso de información ex ante de los sujetos, pudiendo todas las estimaciones efectuarse con información ex post, sin necesidad del uso de supuestos como la independencia condicional (como en el caso de matching). Sin perjuicio de las ventajas recién citadas, y aun cuando se cumplan las condiciones de soporte común y balance que es necesario verificar cuando se utiliza el diseño de regresión discontinua, dado el carácter local de las estimaciones obtenidas al utilizar esta técnica, el evaluador debe recordar que a lo sumo sus resultados son extrapolables a sujetos con puntajes (niveles de riesgo) cercanos al punto de corte establecido como regla de selección del programa.

Por último, en el caso de programas de reinserción de acceso universal pero de participación voluntaria se debe considerar la posibilidad de implementar promoción aleatoria, procedimiento que en el caso que la promoción diseñada efectivamente funciona como un incentivo a participar en el programa, se caracterizan por un menor nivel de localidad (LATE) que las estimaciones RCT en situaciones de cumplimiento imperfecto, tomando todas las precauciones que sea necesario para verificar que el incentivo diseñado es exógeno a los atributos de los sujetos y a los indicadores que de acuerdo al diseño del programa dan cuenta del cumplimiento de sus objetivos más relevantes.

En virtud de la discusión antes presentada, se evidencia que en el caso de los programas que buscan aportar o contribuir a la reinserción social, más allá de los importantes requisitos necesarios para efectuar estimaciones robustas y las dificultades específicas identificadas para cumplirlas en la estimación de los efectos de este tipo de programas, estas condiciones son factibles de lograrse implementando diversas estrategias con las que se puede mitigar el efecto de los problemas descritos.

REFERENCIAS BIBLIOGRÁFICAS

- Andrews, D.; Bonta, J. and Wormith, J. (2011). The Risk-Need-Responsivity (RNR) Model. Does Adding the Good Lives Model Contribute to Effective Crime Prevention? *Criminal Justice and Behavior*, Vol. 38, N° 7, 735–755.
- Arellano, B. (2010). Reinserción Comunitaria la experiencia de un recluso. *Revista de Estudios Criminológicos y Penitenciarios*. Gendarmería de Chile. N° 16.
- Caliendo, M. & Kopeinig, S. (2008). Some Practical Guidance For The Implementation Of Propensity Score Matching. *Journal of Economic Surveys*, 22(1), 31–72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Centro de Estudios en Seguridad Ciudadana, CESC (2012). Modelo de Medición de la Reincidencia de Adolescentes y Jóvenes Infractores de Ley. Servicio Nacional de Menores (SENAME). Gobierno de Chile.
- Centro de Estudios en Seguridad Ciudadana y Fundación Paz Ciudadana, CESC-FPC (2015). Estudio Diseño Organizacional y Modelo de Gestión del Servicio Nacional de Reinserción Social. CESC - Fundación Paz Ciudadana (documento de trabajo).
- Dunn, P.M. (1997). James Lind (1716–94) of Edinburgh and the treatment of scurvy. *Archives of Disease in Childhood. Fetal and Neonatal Edition*. <https://doi.org/10.1136/fn.76.1.F64>
- Gerring, J. (2011). *Social Science Methodology*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139022224>
- Gertler, P.J., Martínez, S., Premand, P., Rawlings, L. B. & Christel M.J., V. (2009). *La evaluación de impacto en la práctica*. <https://doi.org/978-0-8213-8681-1>
- Gertler, P., Martínez, S. & Premand, P. (2017). *La evaluación de impacto en la práctica*. <https://doi.org/978-0-8213-8681-1>
- Goldstein, C. E., Weijer, C., Brehaut, J. C., Fergusson, D. A., Grimshaw, J. M., Horn, A. R. & Taljaard, M. (2018). Ethical issues in pragmatic randomized controlled trials: A review of the recent literature identifies gaps in ethical argumentation. *BMC Medical Ethics*, 19(1), 1–10. <https://doi.org/10.1186/s12910-018-0253-x>
- Holland, P. W. & Rubin, D. B. (1988). Causal Inference in Retrospective Studies. *Evaluation Review*. <https://doi.org/10.1177/0193841X8801200301>
- Khandker, S., B Koolwal, G., & Samad, H. (2010). *Handbook on Impact*. World Bank Publications (Vol. 41).
- Wolff, N. (2000). Using randomized controlled trials to evaluate socially complex services: problems, challenges and recommendations. *The Journal of Mental Health Policy and Economics*. [https://doi.org/10.1002/1099-176X\(200006\)3:2<97::AID-MHP77>3.0.CO;2-S](https://doi.org/10.1002/1099-176X(200006)3:2<97::AID-MHP77>3.0.CO;2-S)

