



# **DISEÑO Y ENCARGO DE EVALUACIONES DE IMPACTO CONTRAFCTUALES**

**GUÍA PRÁCTICA PARA LAS AUTORIDADES DE GESTIÓN DEL FSE**

**COMISIÓN EUROPEA**

Dirección General de Empleo, Asuntos Sociales e Inclusión  
Dirección G: Fondos, Programación y Ejecución  
Unidad G5: Mejora de la Legislación

*Contacto:* Linda Adamaite

*Correo electrónico:* [EMPL-G5-UNIT@ec.europa.eu](mailto:EMPL-G5-UNIT@ec.europa.eu) o [Linda.Adamaite@ec.europa.eu](mailto:Linda.Adamaite@ec.europa.eu)

*Comisión Europea*

*B-1049 Bruselas*

# **DISEÑO Y ENCARGO DE EVALUACIONES DE IMPACTO CONTRAFACTUALES**

**GUÍA PRÁCTICA PARA LAS AUTORIDADES DE GESTIÓN  
DEL FSE**

Texto original finalizado en septiembre de 2021

La edición de 2021 de *Diseño y encargo de evaluaciones de impacto contrafactuales. Guía práctica para las autoridades de gestión del FSE* ha sido elaborada por Jochen Kluve (Humboldt-Universität zu Berlin) y Andrea Naldini y Marco Pompili (Ismeri Europa). Se trata de una adaptación de la guía de 2013 (ISBN 978-92-79-28238-6; doi: 10.2767/94454), que fue elaborada por Stephen Morris (Unidad de Evaluación e Investigación de Políticas, Universidad Metropolitana de Manchester), Herta Tödting-Schönhofer (Metis GmbH, Viena) y Michael Wiseman (Instituto George Washington de Políticas Públicas).

Este documento se ha elaborado para la Comisión Europea. No obstante, únicamente refleja los puntos de vista de sus autores. La Comisión Europea no es responsable de ninguna consecuencia derivada de la reutilización de la presente publicación. Puede encontrarse más información sobre la Unión Europea en internet (<https://www.europa.eu>).

Luxemburgo: Oficina de Publicaciones de la Unión Europea, 2022

© Unión Europea, 2022



La política de reutilización de los documentos de la Comisión Europea se rige por la Decisión 2011/833/UE de la Comisión, de 12 de diciembre de 2011, relativa a la reutilización de los documentos de la Comisión (DO L 330 de 14.12.2011, p. 39).

Salvo que se indique otra cosa, la reutilización del presente documento queda autorizada en virtud de una licencia Creative Commons Attribution 4.0 International (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>). Esto significa que se permite la reutilización siempre que la fuente esté adecuadamente identificada y se indique cualquier cambio.

Para cualquier uso o reproducción de elementos que no sean propiedad de la Unión Europea, podrá ser necesario solicitar la autorización directamente de los respectivos titulares de derechos.

Portada: © Shutterstock, 2021

PDF ISBN 978-92-76-40729-4

doi: 10.2767/424520

KE-02-21-992-ES-N

## Índice

<b>INTRODUCCIÓN: ANTECEDENTES Y OBJETO DE LA GUÍA.....</b>	<b>5</b>
<b>CAPÍTULO 1. CONCEPTO Y ENFOQUES .....</b>	<b>11</b>
1.1. ESENCIA DEL ENFOQUE CONTRAFACtual .....	11
1.2. ¿POR QUÉ SON IMPORTANTES LAS EVALUACIONES CONTRAFACtuALES? .....	12
1.3. ¿POR QUÉ PLANTEAN DIFICULTADES TÉCNICAS LAS EVALUACIONES CONTRAFACtuALES?.....	13
1.4. VISIÓN GENERAL DE LOS DISEÑOS Y ENFOQUES DE LAS EIC .....	15
1.5. CÓMO INTEGRAR LAS EIC EN UN MARCO DE EVALUACIÓN MÁS AMPLIO .....	16
<b>CAPÍTULO 2. CONSIDERACIONES PRÁCTICAS PARA LA ELABORACIÓN DE EIC .....</b>	<b>24</b>
2.1. SELECCIONAR LAS INTERVENCIONES QUE SE SOMETERÁN A UNA EVALUACIÓN DEL IMPACTO .....	27
2.1.1. <i>Priorizar las intervenciones que se someterán a una evaluación del impacto.....</i>	<i>29</i>
2.1.2. <i>Seleccionar las intervenciones adecuadas para un enfoque contrafactual .....</i>	<i>30</i>
2.2. PREGUNTAS DE EVALUACIÓN Y VARIABLES DE RESULTADO .....	35
2.2.1. <i>¿Cuáles son los fines y objetivos de la intervención?.....</i>	<i>35</i>
2.2.2. <i>¿Cuál es la finalidad de la evaluación?.....</i>	<i>35</i>
2.3. DATOS PARA ESPECIFICAR EL GRUPO DE CONTROL Y MEDIR LAS VARIABLES DE RESULTADO.....	40
2.3.1. <i>¿Se dispone de datos adecuados o estos pueden facilitarse?.....</i>	<i>40</i>
2.3.2. <i>¿Cómo se puede especificar el grupo «tratado»? .....</i>	<i>53</i>
2.3.3. <i>Factores que deben tenerse en cuenta para especificar un grupo de control.....</i>	<i>54</i>
2.3.4. <i>¿Qué tipos de cuestiones relativas a los datos deben plantearse en el sistema de evaluación? .....</i>	<i>58</i>
2.3.5. <i>¿Cuáles son las principales limitaciones para analizar los datos y los resultados? .....</i>	<i>60</i>
2.3.6. <i>Una lista de comprobación para verificar la preparación y viabilidad de una EIC .....</i>	<i>63</i>
2.4. MÉTODO DE EIC QUE SE APLICARÁ .....	64
2.5. CALENDARIO Y PRESUPUESTO .....	65
2.5.1. <i>¿De qué recursos se dispone?.....</i>	<i>65</i>
2.5.2. <i>¿Cuándo se debe evaluar la intervención?.....</i>	<i>70</i>
2.6. REALIZAR LA EIC .....	72
2.6.1. <i>Seleccionar el evaluador .....</i>	<i>73</i>
2.6.2. <i>Supervisar la EIC.....</i>	<i>74</i>
2.6.3. <i>Presentar información.....</i>	<i>74</i>
2.6.4. <i>Utilizar los resultados.....</i>	<i>76</i>
<b>CAPÍTULO 3. CÓMO SELECCIONAR LA METODOLOGÍA ADECUADA PARA LLEVAR A CABO UNA EIC.....</b>	<b>79</b>
3.1. ALEATORIZACIÓN: EL ENFOQUE EXPERIMENTAL.....	79
3.2. DISEÑOS NO ALEATORIZADOS O CUASI EXPERIMENTALES.....	82
3.2.1. <i>Grupo destinatario y grupo de control sin aleatorización .....</i>	<i>83</i>
3.2.2. <i>Emparejamiento por puntuaciones de propensión.....</i>	<i>84</i>
3.2.3. <i>Diferencia en diferencias .....</i>	<i>87</i>
3.2.4. <i>Diseño de regresión discontinua .....</i>	<i>90</i>
3.2.5. <i>VARIABLES INSTRUMENTALES .....</i>	<i>94</i>
<b>CAPÍTULO 4. IMPULSAR EL PROGRAMA DE EIC .....</b>	<b>100</b>
4.1. MEJORAR EL NIVEL DE CONOCIMIENTO ENTRE LAS PARTES INTERESADAS.....	100

4.2. DESARROLLO DE CAPACIDADES .....	101
4.3. SUPERAR LAS BARRERAS JURÍDICAS.....	104
4.4. AVANZAR HACIA ENFOQUES MÁS PROSPECTIVOS .....	104
4.5. AMPLIAR EL ALCANCE DE LAS EIC .....	105
<b>GLOSARIOS .....</b>	<b>110</b>
ACRÓNIMOS.....	110
DEFINICIONES.....	111
<b>ANEXOS.....</b>	<b>120</b>
ANEXO 1. BIBLIOGRAFÍA ADICIONAL.....	120
ANEXO 2. PROPUESTA DE ESQUEMA PARA UN CURSO SOBRE EIC .....	124
ANEXO 3. EVALUACIONES DE IMPACTO CONTRAFACTUALES: EJEMPLOS MENCIONADOS EN LA GUÍA .....	125

## LISTA DE RECUADROS

RECUADRO 1 EJEMPLO DE UNA EIC PARA COMPARAR BENEFICIOS Y COSTES .....	22
RECUADRO 2 EIC INTEGRADA EN UN MARCO MÁS AMPLIO .....	26
RECUADRO 3 PREGUNTAS PARA SELECCIONAR LAS INTERVENCIONES QUE SERÁN OBJETO DE UNA EIC .....	28
RECUADRO 4 TIPOS MÁS COMUNES DE INTERVENCIONES Y GRUPOS DESTINATARIOS ELEGIDOS PARA EIC DEL FSE .....	31
RECUADRO 5 DEFINIR LOS GRUPOS DE CONTROL .....	34
RECUADRO 6 EJEMPLOS DE DATOS UTILIZADOS PARA EIC .....	42
RECUADRO 7 EJEMPLOS DE BASES DE DATOS INTEGRADAS PARA EIC .....	43
RECUADRO 8 MARCO REGULADOR DE LA UE SOBRE EL TRATAMIENTO DE DATOS PERSONALES .....	47
RECUADRO 9 PROTECCIÓN E INTERCAMBIO DE DATOS .....	52
RECUADRO 10 PREGUNTAS POLÍTICAS RELACIONADAS CON UN PROGRAMA DE FORMACIÓN .....	57
RECUADRO 11 INTERPRETAR LOS EFECTOS NETOS .....	57
RECUADRO 12 INCERTIDUMBRES EN LA INTERPRETACIÓN DE LOS RESULTADOS .....	62
RECUADRO 13 LA EXPERIENCIA POLACA CON LAS CONFERENCIAS DE EVALUACIÓN.....	78
RECUADRO 14 EJEMPLO DE ENSAYO ALEATORIO DE UN PROYECTO DEL FSE DESTINADO A LOS JÓVENES .....	81
RECUADRO 15 EJEMPLO DE EVALUACIÓN CON UN ENFOQUE DE EMPAREJAMIENTO.....	86
RECUADRO 16 EJEMPLO DE UNA EVALUACIÓN CON UN ENFOQUE DE DIFERENCIA EN DIFERENCIAS .....	89
RECUADRO 17 EJEMPLO DE UNA EVALUACIÓN CON UN ENFOQUE DE REGRESIÓN DISCONTINUA.....	93
RECUADRO 18 EJEMPLO DE UN ESTUDIO MEDIANTE UN ENFOQUE DE VARIABLES INSTRUMENTALES.....	96
RECUADRO 19 EJEMPLO DE UN PROYECTO DESTINADO A REFORZAR LA CULTURA DE LAS EIC Y LA CAPACIDAD PARA LLEVARLAS A CABO .....	103
RECUADRO 20 EJEMPLOS DE EVALUACIONES EN EL ÁMBITO EDUCATIVO .....	106
RECUADRO 21 EJEMPLO DE EVALUACIÓN DE LOS EFECTOS SOBRE LOS EFECTOS DIRECTOS DIFÍCILES DE CUANTIFICAR EN ALEMANIA.....	108

## LISTA DE GRÁFICOS

GRÁFICO 1 DIFERENTES TAREAS Y TIPOS DE EVALUACIÓN.....	17
GRÁFICO 2 ILUSTRACIÓN DEL ENFOQUE DEL MODELO LÓGICO O «TEORÍA DEL CAMBIO» .....	19
GRÁFICO 3 SECUENCIA PRINCIPAL DE LAS ACTIVIDADES DE UNA EVALUACIÓN DE IMPACTO CONTRAFACTUAL .....	25
GRÁFICO 4 TAMAÑOS DE LOS EFECTOS MÍNIMOS DETECTABLES CON MUESTRAS DE DISTINTOS TAMAÑOS .....	61
GRÁFICO 5 CALENDARIO SIMPLIFICADO DE RESULTADOS DE UN PROGRAMA DE FORMACIÓN.....	71
GRÁFICO 6 CONTROL ALEATORIZADO CON DOS GRUPOS .....	80
GRÁFICO 7 DISEÑO CUASI EXPERIMENTAL TIPO CON GRUPOS DE TRATAMIENTO Y DE CONTROL.....	83
GRÁFICO 8 ILUSTRACIÓN DEL ENFOQUE DE LAS PUNTUACIONES DE PROPENSIÓN .....	86
GRÁFICO 9 ILUSTRACIÓN DEL ENFOQUE DE LA DIFERENCIA EN DIFERENCIAS .....	89
GRÁFICO 10 ILUSTRACIÓN DEL ENFOQUE DE LA REGRESIÓN DISCONTINUA .....	91
GRÁFICO 11 ILUSTRACIÓN DE UN ENFOQUE DE VARIABLES INSTRUMENTALES.....	94

## LISTA DE CUADROS

CUADRO 1 CONTENIDO RECOMENDADO PARA UN SISTEMA DE EVALUACIÓN.....	26
CUADRO 2 TIPOS Y FUENTES DE DATOS .....	42
CUADRO 3 ESTRUCTURA DE LOS PRINCIPALES COSTES DE UNA EIC .....	69
CUADRO 4 INFORMACIÓN BÁSICA QUE DEBE CONTENER LA FICHA DE PRESENTACIÓN DE LA EIC.....	76
CUADRO 5 COMPARACIÓN DE LAS CARACTERÍSTICAS FUNDAMENTALES DE LOS PRINCIPALES ENFOQUES DE EIC .....	97
CUADRO 6 CARACTERÍSTICAS DE LAS EIC MENCIONADAS EN LA GUÍA COMO EJEMPLOS .....	125

# Introducción: antecedentes y objeto de la Guía

El período de programación de 2021-2027 comenzó con la dramática experiencia de la pandemia de COVID-19 y la consiguiente crisis económica. La Unión Europea (UE) redobló sus esfuerzos financieros y políticos para contribuir a la recuperación de las economías nacionales y aumentar el empleo. En este contexto, el Fondo Social Europeo Plus (FSE+) desempeña un papel clave a la hora de prestar una asistencia generalizada a los desempleados, con especial atención a los jóvenes y las mujeres, así como de apoyar las intervenciones contra la pobreza infantil, al promover la mejora de la educación e inclusión social de los grupos sociales más débiles en toda la UE. Para alcanzar los diversos objetivos que persigue el FSE+ y conseguir resultados rápidos en los ámbitos del empleo y la inclusión social, es preciso que los recursos se asignen de manera eficaz. Los enfoques empíricos para la elaboración de políticas son cada vez más importantes, y la evaluación es un instrumento fundamental para orientar las políticas públicas.

*Una guía para las autoridades de gestión que actualiza la experiencia de 2014-2020*

En el período 2014-2020, la Comisión Europea (CE) apoyó la capacidad de evaluación de los Estados miembros y de las autoridades de gestión (AG) y promovió el uso de la evaluación de impacto contrafactual (EIC). Los resultados positivos hallados en la EIC son pruebas tangibles de que el FSE tuvo unos efectos mayores de los que se habrían logrado de otro modo. Por primera vez se programó en los planes de evaluación un número considerable de EIC, las cuales se llevaron a cabo durante la ejecución de los programas del FSE. En muchos casos, surgieron dificultades durante la preparación y realización de dichas evaluaciones, o estas siguieron siendo experiencias aisladas y no se incluyeron en un marco de evaluación sistemático.

La presente Guía está destinada a las AG y otros organismos responsables de hacer efectivas las intervenciones y los programas financiados por el FSE+, y tiene por objeto facilitar la planificación, el diseño y el encargo de EIC. Tiene en cuenta las experiencias previas y proporciona asesoramiento práctico sobre algunos de los principales aspectos que deben tenerse en cuenta al preparar una EIC. La Guía actualiza la versión anterior para el período 2014-2020, y hace un mayor hincapié en los problemas encontrados en la aplicación práctica de las EIC. No obstante, en ella se analizan aspectos metodológicos y, cuando es posible, estos se simplifican e integran con ejemplos actualizados seleccionados de las evaluaciones del FSE llevadas a cabo durante el período 2014-2020.

Las EIC responden a las preguntas cruciales que permiten tomar decisiones políticas basadas en pruebas: ¿cuáles son los efectos causales de las intervenciones y «qué funciona»? Tratan de demostrar si los cambios en las circunstancias de los participantes pueden atribuirse realmente a las intervenciones financiadas por el FSE y los consiguientes logros de dichas intervenciones. Cuando se realizan correctamente, las EIC aportan pruebas

*Una EIC responde a la pregunta: «¿Qué funciona?»*

del «efecto neto»<sup>1</sup>, o impacto, de una intervención, lo que permite a los responsables políticos descartar explicaciones alternativas para los cambios en las circunstancias de los participantes o los logros que puedan haberse observado. Cuando las EIC proporcionan estimaciones de la presencia y la magnitud del efecto neto, estas contienen por fuerza una medida de incertidumbre, dependiendo de la exactitud metodológica y de la información disponible. El tipo de pruebas aportadas por estas evaluaciones permite a los responsables políticos valorar la eficacia de las intervenciones, comparar las distintas intervenciones y evaluar sus resultados relativos. Además, realiza importantes aportaciones a los análisis coste-beneficio o a los análisis coste/eficacia.

La presente Guía se publica en un momento en que el FSE+ se enfrenta a unos retos sin precedentes. Habida cuenta del fuerte aumento de los recursos de la UE destinados a inversiones y el empleo, facilitados a través del Instrumento de Recuperación de la Unión Europea, resulta fundamental que los responsables políticos midan y comprendan los efectos de las intervenciones de las que son responsables. Los fondos públicos deben asignarse a intervenciones más productivas y eficaces que aceleren la recuperación y reduzcan los desequilibrios sociales. Por lo tanto, los encargados de desembolsar los recursos del FSE+ deben justificar sus decisiones demostrando que sus intervenciones son eficaces y aportan valor a la ciudadanía. La mejor manera de lograrlo es mediante la realización de más EIC de calidad.

El FSE+ es el principal instrumento europeo de apoyo al empleo y la inclusión social. En el período de programación 2014-2020, el FSE gastó casi 125 000 millones EUR en políticas activas del mercado de trabajo, educación e inclusión social que se pusieron en práctica en los veintiocho Estados miembros a través de programas operativos (PO). Como se indica en el Reglamento (UE) n.º 1303/2013 sobre Disposiciones Comunes, «se llevarán a cabo evaluaciones para mejorar la calidad de la concepción y la ejecución de los programas, así como para valorar su eficacia, eficiencia e impacto».

En el período de programación 2021-2027, se seguirán examinando el rendimiento y los resultados<sup>2</sup>, para lo cual será necesario reforzar los actuales sistemas y capacidades de seguimiento y evaluación, en particular los mecanismos de recopilación de datos. Los planes de evaluación continuarán siendo obligatorios y se hará mayor hincapié en la evaluación del impacto. Habida cuenta de los múltiples métodos existentes para reflejar los impactos de las operaciones financiadas por el FSE+, las autoridades de gestión deben decidir qué método, o combinación de métodos, es más adecuado para satisfacer los requisitos reglamentarios. La cuantificación rigurosa de los impactos de las intervenciones también implica una hipótesis contrafactual.

*Orientación a la obtención de resultados y evaluación de alta calidad*

---

<sup>1</sup> El efecto neto, o impacto, es el residuo entre el efecto total (o bruto) y lo que se habría conseguido sin la intervención. El efecto neto también puede ser negativo, cuando la intervención es menos efectiva que las dinámicas del mercado. Véase la p. 9 para obtener más detalles.

<sup>2</sup> Reglamento (UE) 2021/1060 del Parlamento Europeo y del Consejo, de 24 de junio de 2021, por el que se establecen las disposiciones comunes relativas al Fondo Europeo de Desarrollo Regional, al Fondo Social Europeo Plus, al Fondo de Cohesión, al Fondo de Transición Justa y al Fondo Europeo Marítimo, de Pesca y de Acuicultura, así como las normas financieras para dichos Fondos y para el Fondo de Asilo, Migración e Integración, el Fondo de Seguridad Interior y el Instrumento de Apoyo Financiero a la Gestión de Fronteras y la Política de Visados.

Una característica importante de la nueva normativa es que se centra en lograr un rendimiento sólido y está orientada a la obtención de resultados. Resulta fundamental contar con unas estrategias y técnicas de evaluación de alta calidad para adquirir conocimientos esenciales que muestren a todos los Estados miembros qué intervenciones «funcionan» y cuáles no. Es crucial reforzar la calidad de las evaluaciones y desarrollar pruebas fiables de su valor añadido.

En principio, el punto de partida para recabar pruebas sobre la eficacia de las intervenciones políticas es sencillo. Entre otras cosas, resulta necesario:

- determinar cuál es el problema al que se debe plantar cara,
- decidir los instrumentos que se utilizarán para solucionar el problema,

una fórmula que conecte los instrumentos y los resultados.

Para evaluar un régimen o instrumento de financiación, es fundamental disponer de indicadores claros y mensurables tanto de los insumos aplicados como de los productos y resultados asociados a los problemas. Es habitual establecer objetivos en lo que respecta a los productos y los resultados, y comparar los logros reales con los objetivos marcados. El seguimiento se utiliza para supervisar los insumos, los productos y los resultados a lo largo del tiempo, así como para hacer observaciones sobre la gestión. La teoría de la intervención subyacente a menudo apunta a resultados intermedios que también pueden convertirse en el foco de atención del seguimiento. Sin embargo, desde este punto, no es sencillo determinar si una intervención concreta funciona.

*La necesidad de resultados...*

Es necesario complementar la práctica de evaluación existente con enfoques que generen pruebas mucho más sólidas de los efectos netos o los impactos de las intervenciones. Medir los logros es una forma de rendir cuentas por los fondos utilizados. Las EIC dan respuesta a la pregunta fundamental de si una intervención es eficaz. Si bien estas evaluaciones tratan de establecer un nexo causal entre las intervenciones y los resultados, puede ser necesario recurrir a otros métodos de evaluación de los procesos y basados en la teoría para determinar los mecanismos causales subyacentes y favorecer que los impactos conseguidos en un sitio constituyan una base empírica para replicar las políticas en otros lugares.

*...y pruebas de los efectos netos*

En el período de programación 2014-2020, todos los Estados miembros y las autoridades de gestión adoptaron un plan de evaluación en el que se describían los objetivos, las actividades, los elementos organizativos y los productos de la evaluación. Los planes de evaluación contemplan distintos tipos de evaluaciones, como evaluaciones generales y constantes de los programas, y evaluaciones temáticas que buscan responder a preguntas de evaluación específicas o se centran en ejes o prioridades de inversión concretos de un PO. Un análisis de los planes de evaluación relativos a 177 programas del FSE realizado a finales de 2018<sup>3</sup> mostró que en este período de programación está previsto llevar a cabo 132 evaluaciones de impacto contrafactuales, una cifra que indica que se ha avanzado y que en el período de programación 2014-2020 se prestó más atención a los métodos contrafactuales, en comparación con el período de 2007-2013<sup>4</sup>.

*Experiencia de los Estados miembros con las EIC*

---

<sup>3</sup> Véase [Ismeri Europa, Ecorys e Instituto de Estudios de Empleo, 2019](#).

<sup>4</sup> Véase [Bratu, C. et al., 2014](#).

Según el repositorio de proyectos del Servicio de Asistencia para las Evaluaciones<sup>5</sup>, a junio de 2021 se habían completado 1 795 evaluaciones, de las cuales 1 001 correspondían a programas de la Iniciativa de Empleo Juvenil (IEJ) del FSE o a programas del FSE/IEJ y el Fondo Europeo de Desarrollo Regional (FEDER) (675 y 326, respectivamente). De las 1 001 evaluaciones referentes al FSE, 323 fueron evaluaciones del impacto y 106 aplicaron un enfoque contrafactual: 82 de las 234 evaluaciones del impacto se refirieron a programas del FSE/IEJ y 24 de las 89 evaluaciones del impacto correspondieron a programas del FSE/IEJ y del FEDER<sup>6</sup>.

*La utilización de las EIC va en aumento*

En los primeros años del actual período de programación, las evaluaciones contrafactuales de los programas del FSE/IEJ o FEDER se centraron en el período de programación anterior, mientras que desde 2018 casi todas las EIC evalúan los efectos de las intervenciones de 2014-2020<sup>7</sup>. Veinte de los veintisiete Estados miembros efectuaron evaluaciones contrafactuales de programas, entre ellos el FSE, lo que refleja también en este caso una adopción más generalizada de este enfoque en comparación con el pasado.

Aunque el planteamiento metodológico de las EIC varía entre los distintos Estados miembros, el enfoque contrafactual se aplicó fundamentalmente para evaluar el impacto de las políticas activas del mercado de trabajo (formación, incentivos, ayuda a la búsqueda de empleo, experiencias laborales) y, en mucha menor medida, en otros ámbitos, como las intervenciones en apoyo de las instituciones educativas y los estudiantes, y las políticas relativas a cuestiones sociales, como las intervenciones para luchar contra la pobreza o la exclusión social.

En términos metodológicos, el emparejamiento por puntuaciones de propensión es la técnica más utilizada, mientras que otros métodos se emplean en menor medida. Las EIC suelen centrarse en los efectos a corto o medio plazo (seis o doce meses), y los efectos a largo plazo se evalúan con menor frecuencia.

El análisis de las EIC llevado a cabo en el proyecto del Servicio de Asistencia para las Evaluaciones mostró algunas limitaciones, tales como que las personas de los grupos de control no eran lo suficientemente similares a las de los grupos de apoyo; que los datos eran poco fiables, por ejemplo, cuando procedían de encuestas *ex post* o entrevistas basadas en la capacidad de los encuestados para recordar los acontecimientos con precisión; o que la muestra era tan pequeña que sus resultados no eran estadísticamente significativos, entre otras. En algunos casos, los informes de evaluación no contienen suficiente información sobre los pasos, las opciones y las limitaciones metodológicas, que son elementos necesarios para informar a otros agentes, aparte de los que encargaron dichos informes, de los efectos de las medidas examinadas, a fin de que conozcan mejor los efectos.

*Pero las limitaciones siguen siendo frecuentes*

En resumen, a pesar de los progresos realizados en los últimos años, para llevar a cabo EIC es necesario contar con conocimientos técnicos y que

*Una Guía para profesionales*

<sup>5</sup> El Servicio de Asistencia para las Evaluaciones, financiado por la Dirección General de Política Regional y Urbana (DG REGIO) y la Dirección General de Empleo, Asuntos Sociales e Inclusión (DG EMPL), recaba información sobre las evaluaciones publicadas en los sitios web de AG desde el 1 de enero de 2015. Una red de expertos nacionales encuentra las evaluaciones, que son resumidas y evaluadas en términos de calidad y fiabilidad de sus conclusiones.

[https://ec.europa.eu/regional\\_policy/en/policy/evaluations/member-states/](https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/).

<sup>6</sup> El proyecto de Servicio de Asistencia examinó y resumió 90 evaluaciones de 106.

<sup>7</sup> De las 33 evaluaciones de impacto contrafactuales llevadas a cabo hasta 2018, 29 se refieren al período de programación anterior.

haya voluntad política. La presente Guía defiende las EIC y expone algunos de los problemas que las AG deben resolver para que sean provechosas. Además de los aspectos prácticos de las EIC, también se ponen de relieve las cuestiones más amplias a las que conviene prestar atención para mejorar la calidad de las evaluaciones del impacto. La Guía contiene una serie de recomendaciones básicas, que se espera que sean útiles para las AG. Sin embargo, uno de los principales mensajes es que para llevar a cabo EIC es preciso, entre otras cosas, planificar con atención la preparación de datos (especialmente de los datos administrativos) y definir metas y objetivos claros, a fin de evitar posibles problemas en la fase de ejecución o que las evaluaciones sean de poca calidad.

La Guía también contiene ejemplos de evaluaciones y prácticas para ayudar al lector. Aunque las evaluaciones recogidas por el Servicio de Asistencia para las Evaluaciones constituyen la principal fuente de información para especificar los ejemplos y prácticas presentados en la presente Guía, también se tuvieron en cuenta otras fuentes, y en especial las evaluaciones realizadas por el Centro Común de Investigación [Centro de Investigación sobre la Evaluación de Impacto (CRIE)] y, en menor medida, publicaciones académicas. En la medida de lo posible, los ejemplos y las experiencias recogidos en la Guía se refieren a intervenciones del Fondo Social Europeo.

La Guía está estructurada en cuatro capítulos.

En el **capítulo 1** se analiza la naturaleza de las EIC y por qué son importantes. Se ofrece un resumen introductorio de los enfoques para efectuar EIC y se hace especial hincapié en la distinción entre enfoques experimentales y cuasi experimentales. También se exponen las dificultades técnicas y prácticas que se presentan al llevar a cabo una EIC. Asimismo, se analizan, en términos generales, los tipos de cuestiones políticas en las que podrían centrarse estas evaluaciones y la relación entre los métodos de las EIC y otros enfoques de evaluación (por ejemplo: enfoques basados en la teoría, evaluación de procesos y análisis de la eficiencia).

*Una estructura que refleje las etapas de una evaluación*

En el **capítulo 2** se tratan una serie de cuestiones que las AG deben tener en cuenta a la hora de diseñar y realizar EIC. En la presente Guía se exponen algunos de los principales retos a los que se enfrentan habitualmente quienes preparan EIC y se formulan algunas recomendaciones para darles respuesta. Estas cuestiones sirven de orientación para quienes quieren encargar EIC de intervenciones financiadas por el FSE.

Las EIC aportan pruebas sólidas de la eficacia de los fondos, pero solo si están bien planificadas y se llevan a cabo adecuadamente. Para ello, las AG deben abordar ciertas cuestiones clave cuando realicen una evaluación. La manera exacta en que las AG aborden estas cuestiones y el orden en que lo hagan vendrán determinados por los aspectos prácticos y las disposiciones institucionales sobre el terreno en el Estado miembro de que se trate. La presente Guía pone de relieve algunas de estas cuestiones importantes y las señala a la atención de las AG. La lista de las principales cuestiones se articula atendiendo a las seis etapas del proceso de evaluación: 1) seleccionar las operaciones que serán evaluadas; 2) definir las preguntas de evaluación y las variables de resultado; 3) analizar los datos para especificar el grupo de control y medir las variables de resultado; 4) seleccionar el método de la EIC; 5) definir el calendario y el presupuesto;

6) realizar la evaluación. Se presta especial atención a la disponibilidad y la protección de los datos, un asunto que puede comprometer seriamente la posibilidad de llevar a cabo una EIC.

En el **capítulo 3** se profundiza en la cuarta etapa descrita anteriormente (es decir, en la selección del método de la EIC) y se centra en las características de los principales métodos utilizados en los análisis contrafactuales. En particular, se presentan cinco métodos: el método experimental o aleatorización, el emparejamiento por puntuaciones de propensión, el método de diferencia en diferencias, el diseño de regresión discontinua y las variables instrumentales. Este apartado de la Guía no busca formar a los lectores sobre el uso de los métodos de la EIC; su objetivo es dar a conocer los puntos fuertes y débiles de cada método y mostrar cuándo uno es más adecuado que otro.

En el **capítulo 4** se tratan cuestiones más amplias del desarrollo de perspectivas, tales como la necesidad de desarrollar la capacidad para llevar a cabo EIC eficaces, tanto en el seno de las AG (responsables políticos y funcionarios) como entre las instituciones de investigación, las comunidades académicas y las empresas de consultoría de los Estados miembros. En esta sección también se señala la necesidad de superar los obstáculos jurídicos en torno al acceso a los datos, de actualizar las EIC de los programas del FSE+ a diseños más avanzados, de ampliar la cobertura de la política de inclusión social y educación, de estimar los efectos para los efectos directos difíciles de cuantificar y de aplicar enfoques de meta evaluación.

En resumen, la presente Guía: 1) defiende las EIC, 2) determina las medidas fundamentales que deben tomarse para la realización satisfactoria de EIC y 3) considera a las EIC una parte esencial del panorama del FSE+. El objetivo último es aumentar la contribución del FSE+ al bienestar de los ciudadanos europeos.

Los autores desean reconocer el valioso apoyo y asistencia proporcionados por los miembros de la Unidad G5, Mejora de la Legislación, de la Dirección General de Empleo, Asuntos Sociales e Inclusión en la elaboración de este informe, y en particular por Linda Adamaite, Jeannette Monier y María José Cueto Faus. También desean agradecer a Andrea Pisano y Ernesto Belisario sus aportaciones a la sección relativa a la protección de datos y a Paweł Hess por su apoyo y contribución.

# Capítulo 1. Concepto y enfoques

En el presente capítulo se abordan cuestiones fundamentales sobre la naturaleza y la finalidad del enfoque contrafactual a la hora de evaluar los efectos causales de las intervenciones. Concretamente, se explica la esencia de la evaluación de impacto contrafactual, en particular respecto de los tipos de operaciones cofinanciadas a través del FSE+. También se examina la relación entre los enfoques contrafactuales y otras metodologías de evaluación, y se analiza por qué son importantes las EIC, especialmente en la actualidad. En este capítulo se examinan las cuestiones políticas sobre las que pueden versar las EIC y se presenta un breve resumen de algunos de los principales métodos contrafactuales pertinentes para evaluar las intervenciones cofinanciadas por el FSE+, que posteriormente se analizarán con mayor detalle en el capítulo 3.

## 1.1. Esencia del enfoque contrafactual

Las EIC tratan de determinar los efectos netos o impactos de las intervenciones. Su característica distintiva es que tienen por objeto respaldar las alegaciones de que una determinada intervención provoca un resultado específico, es decir, las afirmaciones de que un resultado concreto solo puede atribuirse a la intervención. Para ello, las EIC aíslan la intervención en cuestión y descartan explicaciones alternativas para el resultado observado.

Su capacidad para descartar explicaciones alternativas se sustenta en la idea de la hipótesis contrafactual, es decir, en la respuesta a la pregunta: «¿Qué habría sucedido, en términos de efectos directos de interés, si no se hubiera llevado a cabo la intervención?». Para comprender bien el concepto de hipótesis contrafactual, y simplemente para aclarar la cuestión, resulta útil pensar en el ejemplo de una persona desempleada que participa en un programa de formación cuyo objetivo es fomentar el empleo. Para determinar el efecto de la formación en esta persona, el enfoque contrafactual concibe dos posibles resultados<sup>8</sup>. El primero es la situación laboral de dicha persona después de haber recibido la formación. Este es el resultado observado para la persona formada. El segundo posible resultado es la situación laboral de esta persona si no hubiera participado en el programa de formación, siendo todos los demás factores iguales. En estas circunstancias, este segundo resultado se denomina «resultado contrafactual». El impacto de la formación para la persona que la ha recibido se determina analizando la diferencia entre el resultado observado y el resultado contrafactual. Este es el efecto causal neto o impacto de la formación para esta persona. La única diferencia entre las circunstancias o condiciones que dieron lugar al resultado observado y el resultado contrafactual es su participación en la formación. Por lo tanto, cualquier diferencia entre ambos resultados debe ser el impacto de la formación en su situación laboral.

*El enfoque  
contrafactual*

En realidad, no observamos ni podemos observar los resultados contrafactuales de las personas expuestas a una intervención. Sin embargo, el objetivo principal de las EIC es proporcionar estimaciones convincentes de

*Definir los grupos  
de tratamiento y...*

<sup>8</sup> En [Holland, P., 1986](#), se puede consultar un análisis más detallado del modelo de nexos de causalidad de «posibles efectos directos».

los resultados contrafactuales para los grupos de personas o empresas a los que afectan las intervenciones cofinanciadas por el FSE. Así pues, los impactos se expresan, por ejemplo, en forma de diferencias en las medias o las proporciones entre los valores promedios observados y los valores contrafactuales «estimados». En la mayoría de los casos en que se llevan a cabo, las EIC tratan de comparar los resultados que una intervención (una medida o un instrumento) tiene para aquellas entidades (personas, pymes, etc.) que se beneficiaron de ella con los de un grupo que no estuvo sujeto a la intervención. En la terminología de estas evaluaciones, se distingue entre el grupo «tratado» o «de tratamiento» y el grupo «de control», que debe ser lo más similar posible al grupo tratado en todos los sentidos, excepto en lo que respecta a los tratamientos recibidos. Las estimaciones de los resultados contrafactuales se obtienen del grupo de control, prestando atención en concreto a las diferencias en las características de los dos grupos, tanto a las que se observan como a las que no. También se pueden comparar varios tratamientos diferentes exponiendo a las unidades elegibles a una serie de variantes del tratamiento (p. ej., otros tratamientos financiados por el FSE o intervenciones financiadas a través de otras fuentes), formando diversos grupos de tratamiento y comparando los resultados entre sí, o los resultados de un grupo de control no tratado.

Cuando el grupo de control no sea expuesto a ningún tratamiento, la evaluación responderá a la pregunta: «¿Qué impacto tiene recibir la intervención en comparación con no obtener ningún tipo de ayuda o apoyo?». En cambio, cuando los resultados de recibir el tratamiento de interés se comparen con los resultados de recibir otro tratamiento, la evaluación responderá a la pregunta: «¿Qué impacto tiene recibir la intervención en cuestión en comparación con someterse a una alternativa claramente definida?». En muchos casos, una EIC se puede diseñar para dar respuesta a cualquiera de estas preguntas fundamentales. La elección de qué pregunta se abordará dependerá de las prioridades de los responsables políticos y de las limitaciones prácticas para su diseño.

*...los grupos de control*

En aquellos casos en que se comparen dos tratamientos distintos, la diferencia entre ambos debería estar clara y ser significativa desde el punto de vista de la elaboración de políticas.

## 1.2. ¿Por qué son importantes las evaluaciones contrafactuales?

Las EIC aportan información importante sobre los efectos netos o impactos de las intervenciones. Proporcionan estimaciones de la magnitud de los impactos, su signo (positivo o negativo) y mediciones estadísticas de la incertidumbre. Ayudan a validar o rechazar la supuesta relación causal entre la intervención y los resultados, la cual subyace al diseño de la intervención. Estos efectos medidos pueden utilizarse para perseguir múltiples objetivos, tales como mostrar transparencia en el gasto de los recursos de los contribuyentes y rendir cuentas por dicho gasto; tomar decisiones políticas basadas en pruebas claras; o aprender de los resultados obtenidos en distintas regiones del mismo país o en diversos Estados miembros.

A los responsables de las intervenciones que quieran asegurarse de que sus programas sigan atrayendo financiación les interesará especialmente promover las EIC para demostrar que sus programas son rentables y aportan

*Transparencia y rendición de cuentas*

beneficios mensurables a los participantes, así como a la sociedad en su conjunto. Las pruebas aportadas por estas evaluaciones serán de especial interés para los responsables de la asignación de los recursos. Las AG estarán deseosas de demostrar que sus programas realmente «funcionan». Y, para hacerlo de manera convincente, tendrán que encargar EIC de gran calidad.

Estas características de las EIC proporcionan información importante a los responsables políticos, cuya tarea consiste en asignar recursos a diferentes intervenciones. Las decisiones relativas a la financiación de posibles intervenciones se toman en un contexto en el que los recursos son limitados. En este contexto, los responsables de la toma de decisiones necesitan pruebas sólidas de los impactos y la rentabilidad del programa, de modo que puedan utilizar los recursos disponibles de la mejor manera posible. En la evaluación de la eficiencia relativa de las intervenciones, los efectos netos estimados por las EIC pueden conocerse con mayor detalle mediante el estudio de la rentabilidad de una intervención o la realización de un análisis coste-beneficio completo.

*Apoyar las decisiones de asignación de recursos*

También se puede obtener otra información importante mediante la comparación de los efectos netos de operaciones similares ejecutadas en diferentes territorios. Si las EIC detectan diferencias significativas en sus efectos, es muy probable que la calidad de las operaciones difiera notablemente o que algunas operaciones se adapten mejor a sus contextos socioeconómicos que otras. Estas comparaciones, a las que se suele denominar «meta evaluaciones» o «meta análisis», permiten generalizar en mayor medida las conclusiones de las EIC y aportan importantes comprobaciones de las medidas políticas (p. ej., formación, servicios integrados, etc.)<sup>9</sup>. Para efectuar estas comparaciones con arreglo a criterios científicos, es necesario realizar EIC de manera sistemática y en una cantidad suficiente. Por consiguiente, la buena coordinación a escala nacional o de la UE es una condición fundamental para que estas comparaciones sean más frecuentes y puedan utilizarse para fundamentar las decisiones políticas.

*Comparar las operaciones de diferentes regiones o Estados miembros*

### 1.3. ¿Por qué plantean dificultades técnicas las evaluaciones contrafactuales?

Se podría decir que hay una serie de enfoques que constituyen intentos «poco fiables» de estimar los impactos de las intervenciones. Se hablará de ellos a continuación para mostrar las complejidades inherentes a las EIC, sin hacer referencia alguna a la práctica de evaluación real.

*Abandonar los enfoques simplistas*

Para empezar, es posible que un responsable político desee evaluar el impacto de un programa de formación para las personas desempleadas mediante la comparación de los ingresos de quienes participaron en una formación, una vez completada esta, con los ingresos de todas las personas desempleadas que no participaron en el programa. A continuación, el responsable político atribuye al programa de formación la diferencia de ingresos observada entre los participantes y los no participantes.

<sup>9</sup> Véase, por ejemplo: [Card, D.; Kluve, J. y Weber, A., 2017.](#)

Esta no es una estrategia válida para determinar el impacto de la formación en los ingresos, ya que las personas que no recibieron la formación podrían presentar características muy distintas a las que sí lo hicieron, y estas diferencias podrían influir en los resultados. Por lo tanto, este enfoque no excluye explicaciones alternativas para las diferencias de ingresos observadas. Por ejemplo, es posible que las personas que participaron en la formación tengan unas capacidades inherentes mayores que las que no lo hicieron. En otras palabras, puede que las personas desempleadas con mayores capacidades se ofreciesen voluntarias para participar en el programa de formación. Así pues, sus capacidades no solo afectan a la decisión de participar, sino también a los resultados, ya que los desempleados con mayores capacidades inherentes tienen más probabilidades de obtener mayores ingresos que aquellos con menores capacidades. En consecuencia, cualquier diferencia de ingresos observada entre el grupo de tratamiento y el grupo de control podría deberse tanto al programa de formación como a las diferencias de capacidad inherentes, de modo que se estaría sobreestimando el impacto de la intervención.

Si no es posible medir las capacidades y, por tanto, no se pueden tener en cuenta las diferencias en las capacidades inherentes entre los dos grupos a la hora de estimar los impactos mediante la comparación de los efectos directos para ambos, se dice que el impacto estimado del programa de formación adolece de **sesgos de selección**. Para subsanar este problema, los evaluadores intentan recopilar la máxima información posible sobre los factores importantes que influyen en la decisión de participar y los consiguientes efectos directos. Estos datos se utilizan para seleccionar un grupo de control válido entre los no participantes, es decir, un grupo lo más similar posible al grupo de tratamiento en lo que respecta a estos factores, y realizar análisis estadísticos teniendo en cuenta estos factores. De esta forma, los evaluadores invocan a menudo la hipótesis de que la admisión en el programa viene determinada por estos factores observables. Por lo general, esta «hipótesis de identificación» no es fácil de probar, por lo que el evaluador debe argumentar de manera convincente que la hipótesis es plausible en el contexto dado, habida cuenta de los factores institucionales y de la teoría del comportamiento que se conocen.

*Contrarrestar el sesgo de selección*

Un segundo enfoque «poco fiable» podría ser que el responsable político se fijase en los ingresos de las personas que participen en el programa antes y después de la formación y atribuyese el cambio a la intervención formativa. En esencia, este enfoque parte de la base de que, si no se interviene, los ingresos medios no varían.

*Antes y después del cambio*

Una vez más, en la mayoría de los casos no se trata de una estrategia válida para medir el impacto de la formación en los ingresos, a menos que pueda invocarse de forma plausible la hipótesis de estabilidad temporal. Esto se debe a que los ingresos de las personas que reciban la formación cambiarán inevitablemente con el paso del tiempo, sin que ello guarde relación alguna con la formación. Por ejemplo, es habitual observar que los ingresos de las personas que participan en una formación disminuyen antes de que esta tenga lugar, en parte debido a factores transitorios. En muchos casos, sus ingresos volverían a repuntar con independencia de la intervención formativa<sup>10</sup>. El enfoque poco fiable de cuantificar el impacto de la formación

<sup>10</sup> Este patrón se conoce con el nombre de «caída de Ashenfelter», en honor al economista que habló de este fenómeno por primera vez. Véase [Ashenfelter, O., 1978](#).

atendiendo a la diferencia entre los ingresos percibidos inmediatamente anteriores a la admisión en el programa y los ingresos percibidos una vez finalizado este ignora el hecho de que, en muchas ocasiones, los ingresos habrían aumentado de todas formas.

Para ajustar estos diseños, es preciso medir la hipótesis contrafactual, es decir, cuánto habrían cambiado los ingresos de las personas que recibieron la formación de no haberse llevado a cabo la intervención formativa. Por ejemplo, este resultado contrafactual puede obtenerse de un grupo de control de características similares cuidadosamente emparejado, que no esté expuesto a la intervención formativa y cuyos ingresos se observen en los mismos momentos que los de los de quienes sí participen en la formación. A continuación, se suele invocar la hipótesis de las tendencias comunes, que establece que la tendencia de los ingresos de las personas que han recibido la formación y del grupo de control habría sido la misma en ausencia de la intervención.

Los límites de estos enfoques «poco fiables» motivan la búsqueda de métodos de evaluación más convincentes. No obstante, como se ha señalado anteriormente, la aplicación de métodos más convincentes plantea más dificultades técnicas. En la siguiente sección de este capítulo se ofrece un breve resumen de algunos de los enfoques específicos de las EIC que probablemente sean más pertinentes en el contexto del FSE.

## 1.4. Visión general de los diseños y enfoques de las EIC

Cuando se centran en los efectos de una intervención en los participantes, los resultados contrafactuales se suelen estimar utilizando datos recopilados de grupos de no participantes que son similares a las personas que sí participaron en la intervención evaluada. En el cuadro 5, se resumen brevemente los enfoques, algunas de sus ventajas y limitaciones, y los tipos esenciales de datos que requieren.

En las EIC, la principal distinción se hace entre los diseños que son experimentales y los que son cuasi experimentales. El enfoque experimental suele denominarse «ensayo aleatorio controlado» (EAC) y, a veces, «experimentación social».

El enfoque experimental se considera el modelo de referencia entre los métodos de EIC destinados a evaluar los efectos de las intervenciones que pueden comprobarse y manipularse en períodos relativamente cortos, y en la mayoría de las circunstancias representa el ideal. Un buen diseño de la evaluación del impacto debe tratar de obtener estimaciones imparciales de los resultados contrafactuales. En muchos usos, se puede considerar que los enfoques experimentales dan lugar a esas estimaciones imparciales. Al examinar los distintos enfoques de la EIC, a menudo conviene empezar describiendo el enfoque experimental, ya que los métodos cuasi experimentales básicamente intentan imitar el ideal experimental.

*Diseño  
aleatorizado: el  
modelo de  
referencia*

En el análisis de los diseños de las EIC, las características clave de cada enfoque se exponen de la manera más sencilla posible con vistas a aclarar los principios subyacentes. En realidad, estos métodos pueden tener aplicaciones considerablemente más complejas, y cuestiones como la falta de adhesión —es decir, cuando las personas que deben participar en la

intervención no lo hacen— puede agrandar de forma significativa las dificultades existentes.

Existen numerosos enfoques que en esencia tratan de imitar la aleatorización. Se los denomina «**cuasi experimentales**». No es posible exponerlos todos en la presente Guía ni proporcionar una descripción técnica completa y pormenorizada de cada uno de ellos. No obstante, en términos generales, estas son las metodologías cuasi experimentales que es más probable que se apliquen en el contexto del FSE: 1. emparejamiento por puntuaciones de propensión; 2. diferencia en diferencias; 3. regresión discontinua; 4. variables instrumentales. El Cuadro 5 del capítulo 3 contiene un resumen de los principales enfoques y sus ventajas relativas. Su presentación se ha simplificado considerablemente para poner de relieve los principios clave de cada enfoque. En el anexo 1 se propone bibliografía complementaria sobre las metodologías cuasi experimentales.

*Diseños no aleatorizados o cuasi experimentales*

## 1.5. Cómo integrar las EIC en un marco de evaluación más amplio

Las evaluaciones contrafactuales responden a determinados tipos de preguntas sobre los efectos causales de las intervenciones. Estos enfoques están limitados en la medida en que podrían abordar otras cuestiones relativas a una intervención. Resulta útil distinguir entre las preguntas de evaluación relativas a la **explicación causal** y las referentes a la **descripción causal**. Las EIC tienen por objeto **describir** las consecuencias de una intervención. Estos métodos son menos adecuados para **explicar** los mecanismos y los contextos a través de los cuales surgen las relaciones causales. Esta distinción es importante, ya que contribuye a aclarar la función característica de las EIC<sup>11</sup>.

*Explicación y descripción causales*

Una EIC bien diseñada indicará al responsable político si una intervención diseñada para influir en unos resultados concretos ha dado lugar a un cambio en ellos. Proporcionará pruebas de la magnitud de cualquier impacto o efecto, indicará al responsable político si el impacto fue positivo o negativo, y le proporcionará una medida de incertidumbre. Sin embargo, a las evaluaciones de impacto contrafactuales no se les da tan bien explicar por qué y cómo se produjeron los impactos que se están midiendo con ellas. Por el contrario, a menudo resulta difícil determinar, a partir de una EIC, por qué una intervención no tuvo ningún impacto, si así se demuestra.

*Qué pueden y qué no pueden revelar las EIC a los responsables políticos*

En la mayoría de los órganos responsables de la formulación de políticas, las partes interesadas que preguntan por la descripción y la explicación causales suelen tener intereses y puntos de vista distintos. Los responsables de programas y los profesionales tienden a centrarse en cuestiones relativas a la explicación causal. Por su parte, los encargados de la asignación de recursos y las autoridades decisorias responsables de la presupuestación suelen centrarse en preguntas relacionadas con la descripción causal. En la práctica, la diferencia entre explicación causal y descripción causal puede ser difusa. En algunas circunstancias, las EIC pueden explicar por qué se produjeron determinados impactos, por ejemplo, mediante el análisis de los impactos de las intervenciones en subgrupos importantes. Sin embargo, resulta esencial analizar atentamente los tipos de preguntas que formulan las

<sup>11</sup> Véanse Shadish, W. R.; Cook, T. D. y Campbell, D. T., 2002, y [Stern, E. et al., 2012](#).

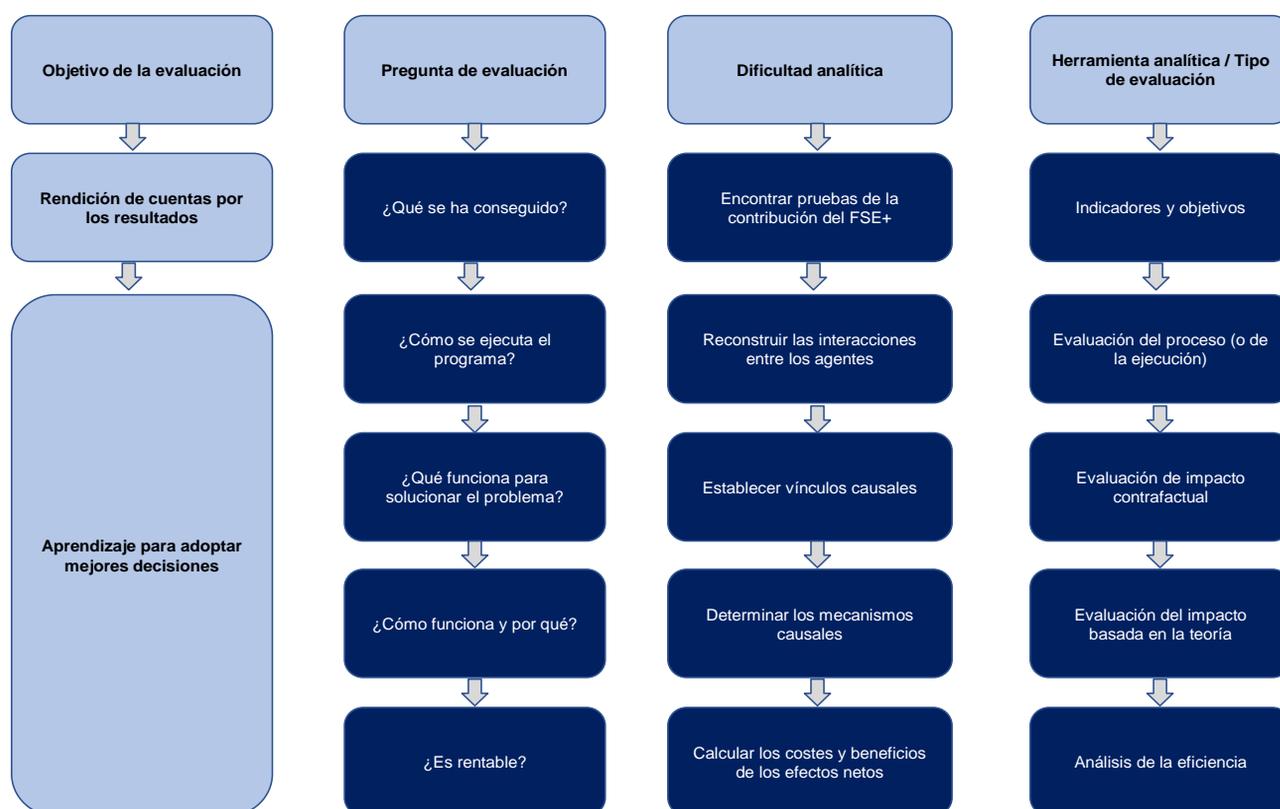
partes interesadas en relación con una intervención y seleccionar el enfoque adecuado para responder a cada una de ellas. Cuando la pregunta principal es si una intervención funciona, una evaluación de impacto contrafactual es, en muchas circunstancias, adecuada. Por contra, en aquellos casos en los que la pregunta principal es cómo funciona una intervención, son más apropiados los métodos de evaluación de procesos y basados en la teoría.

Estos distintos tipos de preguntas y finalidades se resumen en el gráfico que figura a continuación.

Este debate conduce a la conclusión de que las EIC deben desarrollarse en el marco del plan de evaluación, que debe comprender diferentes formas de evaluación dirigidas a responder a distintas preguntas para distintas partes interesadas en la política. En la práctica, un plan de evaluación rara vez incorporará una EIC sin una evaluación del proceso, si es que lo hace alguna vez.

En nombre de la evaluación se pone en práctica una gran variedad de enfoques que persiguen diversas finalidades. La pregunta clave es cómo se pueden combinar estos enfoques de maneras útiles para impulsar el aprendizaje político. El verdadero reto en este ámbito reside en combinar distintos tipos de evaluación de la manera adecuada y con diferentes fines dentro del ciclo de programación. Como se ha comentado y como se muestra en el gráfico que figura a continuación, la EIC, la evaluación del proceso y los enfoques basados en la teoría se complementan entre sí.

**Gráfico 1 Diferentes tareas y tipos de evaluación**



Fuente: adaptado de [Martini, A., 2009](#).

Una estrategia de evaluación sólida debe abarcar los siguientes elementos:

- una evaluación del impacto basada en la teoría,
- una evaluación del proceso (o de la ejecución),
- una evaluación de impacto contrafactual (EIC),
- un análisis de la eficiencia.

*Combinar distintos tipos de evaluaciones*

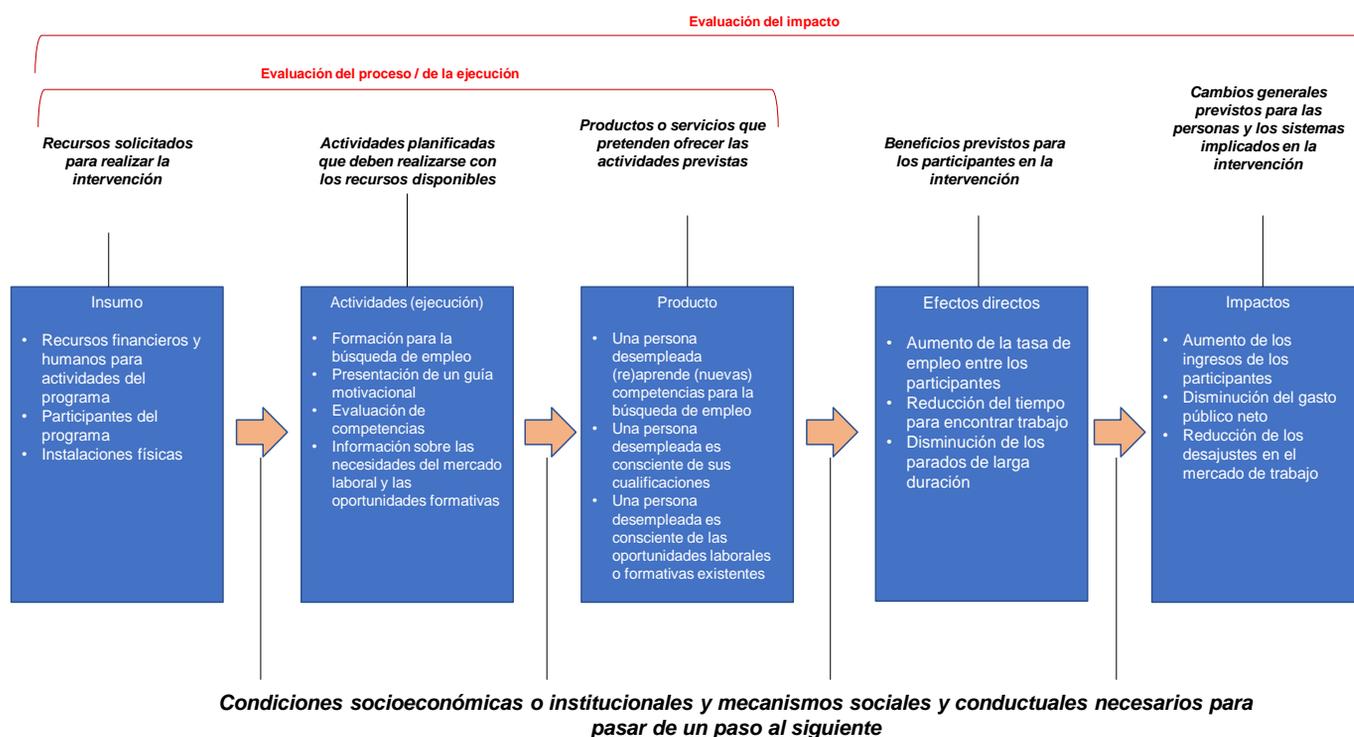
En la presente Guía, solo se analizarán en profundidad los enfoques contrafactuales de la evaluación del impacto. En el contexto de las EIC, los enfoques basados en la teoría sirven para comprender la intención del diseño que subyace a una intervención.

Las **evaluaciones basadas en la teoría** se utilizan en algunas circunstancias para describir el funcionamiento previsto de la intervención y comprobar si se puede observar el cambio en los resultados previsto por la teoría de la intervención. En este sentido, los enfoques basados en la teoría pueden utilizarse para evaluar el impacto respondiendo a la pregunta de «cómo» se ha producido, y podrían usarse para examinar el impacto de una intervención cuando no sea posible llevar a cabo una EIC. En el gráfico siguiente se muestra un ejemplo tipo de la teoría del cambio para una intervención destinada a mejorar los servicios de empleo. Una evaluación basada en la teoría examina si las pruebas confirman la secuencia de pasos causales, desde el insumo hasta el impacto, como señala la teoría del cambio, y en qué condiciones y a través de qué mecanismos sociales fue posible. Una descripción pormenorizada del uso de los enfoques basados en la teoría para determinar el impacto escapa al alcance del presente documento.

*La evaluación basada en la teoría se refiere a una «teoría del cambio»*

En el contexto de la EIC, la evaluación basada en la teoría examina cómo se planifica y diseña una intervención y cómo se pretende que funcione. Básicamente, el enfoque consiste en colaborar con las partes interesadas en una intervención para desarrollar una descripción común de la «teoría del cambio» subyacente a una intervención, tal como se simplifica en el gráfico que figura a continuación. Todas las intervenciones llevan incorporada una lógica del programa que vincula los insumos y las actividades con los productos, los resultados intermedios y, por último, los resultados a largo plazo. Por consiguiente, la articulación de una teoría del cambio también es una manera importante de sustentar las EIC, pues facilita la reconstrucción del proceso de ejecución y su posible influencia en los productos y efectos directos, así como la determinación de las variables de resultado más representativas que deben comprobarse en el análisis contrafactual.

## Gráfico 2 Ilustración del enfoque del modelo lógico o «teoría del cambio»



Fuente: adaptado de la [Fundación W.K. Kellog, 2004](#), y de [Bredgaard, T., 2015](#).

La evaluación basada en la teoría puede vincularse a las evaluaciones de impacto contrafactual de varias maneras útiles. Una teoría del cambio (o lógica de intervención) claramente definida puede orientar el diseño de una EIC. Entre otros aspectos, una teoría del cambio bien definida puede indicar al responsable de diseñar una evaluación del impacto lo siguiente:

*...que contribuye a las evaluaciones de impacto contrafactual*

- ¿Qué resultados son importantes y deben medirse?
- ¿Cuál podría ser el signo y la magnitud probables de los impactos de la intervención?
- ¿Quién es el grupo destinatario previsto y cómo se puede seleccionar un grupo de control?
- ¿Cuánto tiempo podrían tardar en materializarse los efectos y los resultados del programa?
- ¿Qué datos podrían ser necesarios para medir la participación en la intervención?
- ¿Cómo de factible es utilizar el grupo de control para medir la hipótesis contrafactual?

El desarrollo de una teoría del cambio también puede ayudar a detectar posibles efectos imprevistos que puedan tenerse en cuenta a la hora de diseñar una EIC. En cierta medida, una teoría del cambio claramente articulada también podría ayudar al evaluador a interpretar los resultados del estudio de una EIC. Sin embargo, en términos de interpretación, una **evaluación del proceso** también puede aportar mucha información.

*Evaluación del proceso*

En el contexto de las EIC, la **evaluación del proceso** tiene dos objetivos: evaluar la «fidelidad» y evaluar la diferencia entre las experiencias de los miembros del grupo de tratamiento y el grupo de control.

La evaluación de la «fidelidad» examina hasta qué punto una intervención, en la forma en que se ha llevado a cabo, es fiel a su diseño. Una evaluación del proceso analiza qué servicios se pusieron realmente a disposición de los participantes de la intervención. ¿Se corresponden con lo previsto en la teoría de la intervención? ¿A qué responde la variación en la forma de llevar a cabo la intervención entre unos lugares y otros, si la hubiera? La mayoría de las intervenciones tienen una lógica de gestión y de efecto:

- La lógica de gestión se refiere a cómo se espera que los órganos de ejecución respondan a las normas e incentivos del programa.
- La lógica de efecto se refiere a la forma en que se espera que respondan las personas a las que se dirige la intervención, teniendo en cuenta la intervención que realmente se lleva a cabo.

La parte del análisis del proceso relativa a la fidelidad proporciona información sobre lo que realmente se logró con una intervención y, por tanto, sobre qué contribuye verdaderamente a los efectos observados. También proporciona información importante para la gestión de los proyectos.

La evaluación de la diferencia resulta especialmente importante en el contexto de las evaluaciones contrafactuales. Es habitual centrarse, como ya se ha hecho en buena parte en la presente Guía, en los impactos de las intervenciones. Pero antes que el impacto en los resultados está el impacto en los insumos, es decir, la diferencia de oportunidades entre el grupo de tratamiento y el grupo de control que realmente se logra con una intervención. En principio, todas las EIC pueden «darse la vuelta» para convertir el grupo de tratamiento en el grupo de control, a fin de evaluar el resultado para las personas que integran el que, antes de la inversión, era el grupo de control. La idea es que es necesario conocer la experiencia del grupo de control en la misma medida que se conoce la del grupo de tratamiento, ya que la EIC atribuye la causalidad de la diferencia en los resultados a la diferencia entre los insumos del grupo de tratamiento y el grupo de control.

Volviendo de nuevo al programa de formación, se pueden concebir dos circunstancias de partida sumamente distintas. En una, el programa de formación se imparte en un contexto general en el que no existe nada igual. A los integrantes del grupo de control, sencillamente, no se les ofrece ninguna alternativa. No obstante, otra posibilidad es que haya otras opciones. Por ejemplo, podría haber empresas especializadas en la preparación profesional que impartan formación. De ser así, el análisis del proceso debe contener, en la medida de lo posible, una evaluación de la diferencia en la participación en la formación por parte del grupo de tratamiento y del grupo control, en lugar de limitarse a suponer que todas las dimensiones del grupo de tratamiento están fuera del alcance del grupo de control.

Si bien las evaluaciones del proceso pueden encargarse de forma totalmente independiente a otros tipos de evaluación, su importancia tanto para la gestión como para la EIC hace que resulte esencial planificar conjuntamente las evaluaciones del proceso y del impacto.

Un buen análisis del proceso puede contribuir a conseguir la fidelidad, y las evaluaciones del proceso proporcionan una explicación causal de una

*Evaluación de la fidelidad*

*Diferencia entre el grupo tratado y el grupo de control*

*Toda EIC requiere una evaluación del proceso*

intervención. Sin una evaluación del proceso bien diseñada, a menudo cuesta interpretar plenamente los resultados de una EIC o determinar los costes que deben incluirse en la evaluación coste-beneficio, una vez que se dispone de estimaciones del impacto.

Como se ha señalado anteriormente, la evaluación del proceso también puede contribuir a interpretar las conclusiones de las evaluaciones del impacto con una descripción del contexto en el que tuvo lugar una intervención. Comprender el contexto es fundamental, dado que proporciona las condiciones para que una intervención tenga éxito y permite valorar en qué medida dicha intervención podría tener efectos similares si se llevase a cabo en otros lugares, en diferentes zonas geográficas o en un momento distinto. Esto es especialmente importante para analizar la posibilidad de transferir los enfoques políticos y para poner de relieve las buenas prácticas existentes en el aprendizaje y el intercambio transnacionales. El análisis del proceso contribuye a mejorar la confianza en la denominada «**validez externa**» de los resultados de la evaluación.

En la mayoría de los casos, el análisis de la eficiencia implica, bien un análisis coste/eficacia, bien un análisis coste-beneficio completo.

*Determinar los índices de coste/eficacia*

El **análisis coste/eficacia** consiste en comparar los costes de la intervención con sus efectos o impactos, determinados por una EIC. En pocas palabras, el índice de coste/eficacia se obtiene dividiendo el impacto de una intervención —expresado en las unidades de medida o en unidades normalizadas— por el coste neto de realizar la intervención por unidad tratada.

El índice de coste/eficacia de un programa de formación cuyo objetivo es ayudar a las personas desempleadas a encontrar trabajo podría indicar los fondos necesarios por participante para que cada participante deje de estar en paro y encuentre trabajo.

*El ACB para comparar el beneficio con el coste neto*

El índice de coste/eficacia constituye un indicador importante para los responsables de asignar los recursos entre distintos programas. Los índices obtenidos a partir de distintas intervenciones permiten a los responsables de la asignación de los recursos hacer valoraciones relativas respecto de qué intervenciones ofrecen una mayor rentabilidad.

En lugar de expresar los efectos del programa en su unidad de medida o en unidades normalizadas, un **análisis coste-beneficio** (ACB) intenta monetizar las estimaciones del impacto derivadas de una EIC y compararlas con los costes netos de una intervención. El análisis coste-beneficio tiene por objeto determinar si los beneficios monetarios de un programa superan sus costes netos. Un análisis coste-beneficio de un programa de formación típico del FSE compararía los beneficios que la intervención tiene para sus participantes, el Gobierno y la sociedad en su conjunto con los costes netos de la intervención. En el caso de los participantes, una EIC muestra los beneficios del programa, que suelen consistir en la mejora de la empleabilidad y el aumento de sus ingresos. Para obtener los beneficios netos, habría que restar los impuestos abonados por los participantes y otros costes del empleo. Desde el punto de vista del Gobierno, los beneficios de la intervención consistirán en un aumento de los ingresos fiscales y una disminución de las prestaciones, si bien tendría que correr con la mayor parte de los costes de la intervención. Los beneficios para la sociedad en su conjunto se derivan de sumar los beneficios para los participantes y el

Gobierno y restar de ellos la suma de los costes para los participantes y el Gobierno.

Los análisis de coste/eficacia y los ACB todavía no son prácticas extendidas en las evaluaciones del FSE. Sin embargo, estos análisis son muy útiles para decidir si se debería volver a financiar una intervención en el futuro o para determinar cuál es la intervención más eficaz entre una serie de intervenciones similares (véase el ejemplo del recuadro siguiente).

Las estimaciones del impacto de una EIC son un componente clave de los análisis de coste/eficacia y los análisis coste-beneficio. Para los primeros, proporcionan una medición de la eficacia, mientras que, en el caso de los segundos, son una fuente primordial para estimar los beneficios en valor monetario. Lo que también está claro es que tanto los análisis de coste/eficacia como los análisis coste-beneficio requieren la recopilación de datos precisos sobre los costes a partir de los cuales puedan calcularse los costes netos. Estas actividades se suelen conocer con el nombre de «estudio de costes». En algunas evaluaciones complejas que combinan varios métodos, los estudios de costes suelen integrarse en la evaluación del proceso, en la que los instrumentos de investigación pueden adaptarse para recopilar datos importantes sobre los costes.

### Recuadro 1 Ejemplo de una EIC para comparar beneficios y costes

En [Bazzoli, M. et al., 2018](#), puede consultarse un ejemplo de uso de un enfoque contrafactual para estimar los beneficios y los costes. El estudio se centra en los programas de formación profesional llevados a cabo en la Provincia Autónoma de Trento (Italia) en 2010-2011, en el marco de los cuales se impartieron más de trescientas horas de actividades de formación. Se evaluaron dos grupos de cursos: los financiados con recursos provinciales y los financiados por el Fondo Social Europeo (FSE), en los cuales participaron 954 y 205 personas, respectivamente.

Estos fueron los principales pasos que se siguieron para efectuar un ACB: 1) los autores evaluaron el impacto de los cursos de formación en la probabilidad de que los participantes encontraran un empleo durante los tres años posteriores al inicio de su curso, mediante un emparejamiento por puntuaciones de propensión<sup>12</sup>; 2) se calculó el impacto de los cursos en sus ingresos brutos hasta finales de 2013; 3) los autores estimaron el importe de las declaraciones fiscales adicionales (derivadas de los efectos sobre los ingresos) y el ahorro en gasto público generado por la reducción del número de beneficiarios de prestaciones por desempleo; 4) se compararon los beneficios y los costes de los cursos.

Se utilizaron varios conjuntos de datos administrativos: a) datos de seguimiento relativos a los participantes y sus características; b) datos de las personas desempleadas de los registros de los centros de los Servicios Públicos de Empleo (SPE) (Centri per l'impiego) para definir el grupo de control; c) datos de la base de datos COB, el archivo de notificación obligatoria por parte de las empresas de los contratos laborales enviados a los Servicios Públicos de Empleo y utilizados para determinar la situación laboral del grupo tratado y el grupo de control, tanto antes como después de participar en el curso de formación; d) datos de los archivos de ingresos fiscales, para calcular los ingresos de las personas; y e) datos del Instituto Italiano de Seguridad Social (INPS) para obtener información sobre las prestaciones por desempleo percibidas por las personas.

<sup>12</sup> Más concretamente, los autores aplicaron el estimador de bloqueo con ajuste de la regresión.

Después de treinta y seis meses, la probabilidad de que los integrantes del grupo tratado de los cursos financiados con recursos provinciales tuvieran trabajo era aproximadamente un 5 % superior a las de los miembros del grupo de control, mientras que el impacto de los cursos del FSE era mucho mayor, de entorno al 28 %. En los tres años posteriores a la intervención, las personas que participaron en los cursos financiados con cargo a los recursos provinciales ganaron, en promedio, 2 250 EUR anuales más que el grupo de control, mientras que quienes participaron en los cursos del FSE ganaron 4 106 EUR anuales más que el grupo de control. Los datos también permitieron a los autores estimar los beneficios para las administraciones públicas en términos de aumento de los ingresos fiscales y reducción del gasto en prestaciones. Las personas del grupo de cursos financiados mediante recursos provinciales pagaron 126 EUR más en concepto del impuesto sobre la renta que el grupo de control por cada año objeto de examen, mientras que la estimación para las personas del grupo de cursos del FSE ascendió a 318 EUR anuales. En cuanto a la disminución de las prestaciones por desempleo abonadas por la administración pública, el impacto de los cursos fue insignificante, muy probablemente porque el valor monetario de estas prestaciones dependía de cuánto tiempo se hubiera trabajado antes de quedarse en paro, y muchos participantes eran jóvenes sin apenas experiencia laboral.

Los costes de las intervenciones ascendieron a 4 800 EUR por participante en los cursos financiados con recursos provinciales y a 14 500 EUR en el caso de los cursos del FSE. Los autores también compararon los costes y los beneficios individuales de ambos tipos de cursos. En promedio, constataron que los costes fueron superiores a los beneficios en el período 2010-2013<sup>13</sup>.

---

<sup>13</sup> Puede consultarse otro ejercicio «similar» en [Lammers, M. y Kok, L., 2021](#).

## Capítulo 2. Consideraciones prácticas para la elaboración de EIC

En el presente capítulo se examinan las cuestiones prácticas que deben tenerse en cuenta a la hora de preparar una evaluación. Debe utilizarse cuando se planifiquen actividades de evaluación, para decidir qué intervenciones deben someterse a una EIC y para decidir las principales preguntas a las que se debe dar respuesta en el diseño de una EIC.

*Preparar una EIC*

Se asume que, en un comienzo, el responsable de programas de una AG [o un responsable de un organismo intermediario (OI) encargado de llevar a cabo una intervención del FSE] se plantea qué intervenciones evaluar y las estrategias adecuadas para incorporar una EIC. También se asume que los funcionarios de una AG no serán quienes realicen las evaluaciones, sino que contratarán o encargarán los servicios de evaluación a expertos externos. Aunque será un contratista quien efectúe la EIC, la AG (o el OI) tendrá que planificar y preparar una evaluación del impacto antes de encargarla.

La estrategia de evaluación, así como los diversos tipos de evaluaciones descritos en el capítulo anterior, debe establecerse en el plan de evaluación.

Los planes de evaluación son obligatorios para todos los programas y deben ser aprobados por el Comité de Seguimiento, a más tardar, un año después de la aprobación del programa<sup>14</sup>. Estos planes deben elaborarse al inicio del período de programación y contener disposiciones relativas al proceso de evaluación (la gobernanza de las evaluaciones y el vínculo entre evaluación y seguimiento), actividades de evaluación reales (p. ej., una lista indicativa de las evaluaciones que deben llevarse a cabo, el alcance de cada evaluación, las principales preguntas, los datos necesarios, los posibles usos, un calendario indicativo, la estructura de gestión), el calendario de las evaluaciones, el presupuesto global y el desarrollo de la capacidad de evaluación.

Si bien los planes de evaluación suelen ser de carácter general, la planificación de una EIC requiere una preparación más detallada. Lo ideal sería que se preparase en el momento de redactar el plan de evaluación, aunque más adelante pudieran añadirse algunos detalles. Sin embargo, la AG o el OI debe tener en cuenta que el establecimiento de contactos con las partes interesadas y otras medidas necesarias para recabar datos relacionados con las intervenciones no suelen ser fáciles y deben planificarse con bastante antelación.

La presente Guía se centra en las maneras de desarrollar un sistema de evaluación para intervenciones específicas que sean candidatas a una EIC. Dicho sistema debe formar parte del plan de evaluación o, en su defecto, podría establecerse como paso operacional derivado de un plan de evaluación.

*Desarrollar un sistema de evaluación para intervenciones específicas*

No todas las intervenciones financiadas por el FSE pueden someterse a una evaluación contrafactual. Los responsables políticos deben decidir en qué

<sup>14</sup> Los planes de evaluación son obligatorios de conformidad con el artículo 44, apartados 5 y 6, del Reglamento (UE) 2021/1060 del Parlamento Europeo y del Consejo. Aunque un plan puede incluir las evaluaciones planificadas para varios programas, todos los programas deben estar cubiertos por un plan de evaluación.

aspectos centrar su atención. Habrá que llevar a cabo un proceso de selección de las intervenciones que deberán ser objeto de una evaluación del impacto. En esta Guía se señalan algunos aspectos que las AG deberán tener en cuenta cuando seleccionen las intervenciones oportunas. Además, su objetivo central es ayudar a los responsables de encargar EIC a hallar soluciones para algunos de los retos que plantea la creación de una evaluación del impacto satisfactoria y, de este modo, desarrollar sistemas de evaluación para las distintas EIC que se estén planteando.

La presente Guía asume que, tras seleccionar las intervenciones que deben someterse a una EIC, las AG tendrán que preparar un sistema de evaluación para cada intervención elegida. En este contexto, el término «sistema» se utiliza para distinguir esta actividad de los «planes» de evaluación formales que exige el Reglamento (UE) 2021/1060 (Reglamento sobre Disposiciones Comunes o RDC) para el período de programación 2021-2027. En particular, el término «sistema» se refiere al conjunto de actividades normalizadas necesarias para definir y efectuar una EIC, que deben prepararse antes de su puesta en práctica, como se muestra en el gráfico siguiente.

**Gráfico 3 Secuencia principal de las actividades de una evaluación de impacto contrafactual**



Estos sistemas constituirán la base para encargar EIC y sentarán las bases para que los contratistas puedan realizar un estudio riguroso y valioso. En el resto del capítulo se examinan las cuestiones que deben tenerse en cuenta al planificar una evaluación. Los sistemas de evaluación deben adaptarse a las circunstancias específicas en las que se produce la intervención. Puesto que es imposible especular sobre cuáles serán estas circunstancias específicas, la presente Guía se centra en las cuestiones que a) deberían incluirse en los programas o b) deberían incitar a la reflexión sobre las dificultades a las que se enfrentarán dichos sistemas.

*Cuestiones que deben abordarse en la planificación de la evaluación*

Tras examinar algunas de las cuestiones que se deben abordar cuando se valore qué intervenciones podrían someterse a una EIC, e incluso si es posible emprender una EIC con los tipos de datos disponibles, la atención pasa a centrarse en algunas de las cuestiones clave que deben tenerse en cuenta cuando se desarrolle un sistema de evaluación. Esto debe ponerse por escrito antes de encargar una EIC o un estudio de evaluación más amplio, a fin de poder preparar los términos de referencia y designar a un contratista. En el cuadro que figura a continuación se recoge el contenido principal de este sistema de evaluación.

**Cuadro 1 Contenido recomendado para un sistema de evaluación**

Principales pasos para preparar y llevar a cabo una EIC	Contenido
<b>1. Seleccionar las operaciones que se van a evaluar</b>	<ul style="list-style-type: none"> <li>- Las operaciones del FSE, o los tipos de operaciones (si es posible, partir de una tipología predeterminada), que se evaluarán con la EIC.</li> <li>- Una breve descripción de la lógica de trabajo de las operaciones seleccionadas (objetivos, principales criterios de elegibilidad y población destinataria, tipos de asistencia, modelo de ejecución, fechas aproximadas de activación y finalización, presupuesto indicativo).</li> </ul>
<b>2. Definir las preguntas de evaluación y las variables de resultado</b>	<ul style="list-style-type: none"> <li>- Preguntas de la evaluación.</li> <li>- Funcionamiento previsto de las operaciones («teoría del cambio») y variables de resultado (p. ej., situación laboral, cambios en los ingresos, situación de pobreza, puntuación media en los exámenes, etc.).</li> </ul>
<b>3. Analizar los datos para especificar el grupo de control y medir las variables de resultado</b>	<ul style="list-style-type: none"> <li>- Los datos administrativos (p. ej., registros de desempleo, registros fiscales, datos administrativos de seguros, registro de escuelas o base de datos de estudiantes, etc.) o de otro tipo (p. ej., encuestas, macrodatos) que deben utilizarse.</li> <li>- Las variables de la base de datos o encuesta que se utilizará en la EIC.</li> <li>- Las series temporales de las variables necesarias.</li> <li>- Las principales normas y cuestiones relativas al acceso a los datos (accesibilidad directa por parte de la AG, necesidad de un acuerdo con otras administraciones, normas de privacidad y restricciones).</li> </ul>
<b>4. Seleccionar el método de EIC</b>	<ul style="list-style-type: none"> <li>- El posible método de EIC que se adoptará (puede detallarse más adelante, pero el uso de ensayos aleatorios controlados u otros métodos debería indicarse de antemano para favorecer que el proceso de evaluación sea coherente).</li> </ul>
<b>5. Definir el calendario y el presupuesto</b>	<ul style="list-style-type: none"> <li>- El calendario de las EIC y los principales hitos: a) decisión sobre un enfoque experimental o cuasi experimental, b) preguntas detalladas de la evaluación, c) preparación de los términos de referencia, d) identificación del evaluador, e) preparación de los datos, f) análisis de los datos, g) informes intermedios y final, h) validación de los resultados, i) difusión de los resultados y las lecciones aprendidas.</li> <li>- El presupuesto disponible para la EIC.</li> </ul>
<b>6. Realizar la evaluación</b>	<ul style="list-style-type: none"> <li>- Selección del evaluador.</li> <li>- Supervisión de la realización de la EIC.</li> <li>- Presentación de informes sobre la EIC.</li> <li>- Distribución de los resultados (principales partes interesadas a las que se debe involucrar, principales herramientas).</li> </ul>

**Recuadro 2 EIC integrada en un marco más amplio**

Muchas de las EIC de intervenciones financiadas por el FSE que se llevan a cabo en los Estados miembros están integradas en marcos de evaluación más amplios.

- En Alemania, en el contexto del PO Bund del FSE 2014-2020, la evaluación contrafactual del programa para la integración de los parados de larga duración en el mercado laboral, financiado en el marco de la prioridad de inversión (PI) 9.i, forma parte de una estrategia de evaluación más amplia que prevé la presentación de informes provisionales anuales que también examinan temas relacionados con la ejecución de las intervenciones. De 2017 a 2021 se elaboraron cuatro informes<sup>15</sup>. Lo mismo cabe decir de la evaluación de las medidas del FSE en apoyo de la integración de los parados de larga duración en Baden-Württemberg, donde el análisis contrafactual se realizó tras un análisis más cualitativo centrado en las evaluaciones de las intervenciones por parte de los participantes<sup>16</sup>.

<sup>15</sup> Véanse [Boockmann, B. et al., 2017](#); [Boockmann, B. et al., 2018](#); [Boockmann, B. et al., 2019](#); y [Boockmann, B. et al., 2021](#). En 2019 y 2021, se llevó a cabo una EIC.

<sup>16</sup> Véanse [Hunger, K. y Sattler, K., 2017](#), y [Scheller, F. y Seidel, K., 2020](#). En 2020, se llevó a cabo una EIC.

- En Piamonte (Italia), la evaluación contrafactual de los efectos sobre el empleo de los bonos de empleo dirigidos a personas vulnerables, que se financiaron en el marco de las PI 8.i y 9.i, fue uno de los pasos de una evaluación continua más exhaustiva. Dos informes iniciales examinaron los problemas en la ejecución y las percepciones subjetivas de los participantes (abril de 2018 y febrero de 2019), mientras que otros dos informes realizados en julio de 2019 y a finales de 2020 se centraron en los impactos en el empleo utilizando un enfoque contrafactual<sup>17</sup>.
- En Las Marcas (Italia), la EIC que examinó los impactos de las intervenciones dirigidas a los parados de larga duración en 2020 es un análisis del impacto con un enfoque temático que se puso en marcha tras llevar a cabo, en 2019, un análisis más general del impacto de las operaciones del FSE dirigidas a los desempleados (informe sobre la colocación)<sup>18</sup>.
- En Polonia, el PO Knowledge Education Growth (en español, «Conocimiento, Educación, Crecimiento») para 2014-2020 encargó una serie de evaluaciones (ocho informes<sup>19</sup>) entre finales de 2015 y mayo de 2020 para analizar el apoyo del FSE y de la IEJ a los jóvenes desde diferentes perspectivas y mediante la aplicación de diversos métodos analíticos (p. ej., análisis cualitativos a través de encuestas y entrevistas, macromodelos, actividades sobre el terreno y un enfoque contrafactual).

## 2.1. Seleccionar las intervenciones que se someterán a una evaluación del impacto



La selección de las intervenciones que se someterán a una evaluación del impacto comprende tres pasos clave:

*Crterios para seleccionar las intervenciones*

1. Deben determinarse las cuestiones estratégicas.
2. Una vez que las prioridades estratégicas estén claras, es preciso evaluar cada intervención para dilucidar si se ajusta a los requisitos básicos de un enfoque contrafactual y en qué medida es innovadora o contribuye de manera significativa a la base de conocimientos.
3. Debe aclararse la disponibilidad real o potencial de los tipos de datos necesarios para llevar a cabo una EIC. Hasta la fecha, este tercer paso

<sup>17</sup> Véanse [Pomatto, G., 2017](#); [Pomatto, G., 2019](#); [Poy, S., 2019](#), y [Poy, S., 2020](#). En 2019 y 2020, se llevó a cabo una EIC.

<sup>18</sup> Véanse [Pompili, M. y Giorgetti, I., 2020](#), y [Pompili, M. y Giorgetti, I., 2020a](#)

<sup>19</sup> Véanse [Instytut Badań Strukturalnych, Imapp e IQS, 2015](#); [Baran, J. et al., 2016](#); [Baran, J. et al., 2017](#); [Baran, J. et al., 2018](#); [Baran, J. et al., 2018a](#); [Palczyńska, M. et al., 2019](#); [Kalinowski, H., 2020](#), y [Kalinowski, H. et al., 2020](#). En 2017 y 2020, se llevó a cabo una EIC.

ha constituido un gran obstáculo para llevar a cabo evaluaciones contrafactuales de las intervenciones del FSE y merece especial atención.

### Recuadro 3 Preguntas para seleccionar las intervenciones que serán objeto de una EIC

Las EIC no son adecuadas para todas las intervenciones, y su realización indiscriminada no suele resultar rentable. Las autoridades de gestión deben decidir cómo asignar los recursos para maximizar los beneficios. Se espera que el plan de evaluación refleje estas decisiones y que en la planificación de las EIC se tengan en cuenta tres elementos principales: a) las prioridades estratégicas, b) la viabilidad de una EIC, y c) la disponibilidad de los datos necesarios.

La magnitud, la formulación de políticas, las áreas de incertidumbre y la necesidad de conocimientos influyen en **la estrategia de evaluación**. Las AG deben hacerse las siguientes preguntas:

- ¿La gran cantidad de fondos asignados a esta intervención hace que sea especialmente importante justificar el gasto? Estas intervenciones son relativamente fáciles de identificar porque reciben el grueso de los fondos asignados a cada objetivo específico, tal como se define en el artículo 2 del Reglamento (UE) 2021/1057 relativo al FSE+.
- ¿Es la medida el centro de atención de un proceso de reforma y es probable que los resultados de la evaluación contribuyan a una revisión crítica de la iniciativa? Estas intervenciones están vinculadas a reformas recientes de la política laboral, educativa o social; aunque es posible que no reciban grandes sumas, son cruciales para el éxito de la reforma.
- ¿Es la intervención innovadora y se está poniendo a prueba mediante una intervención piloto o un ensayo antes de su expansión? Es posible que estas intervenciones no reciban una enorme cantidad de recursos, pero requieran una evaluación temprana para decidir si conviene seguir adelante y expandirlas o es mejor poner fin al experimento.
- ¿Se centra la intervención en ámbitos que requieren pruebas adicionales de su eficacia? Este grupo incluye aquellas intervenciones que no han sido evaluadas en el pasado o cuyas últimas evaluaciones son tan antiguas que deben ser actualizadas. En los programas del FSE+ hay muchas intervenciones de este tipo, ya que las EIC no se han generalizado hasta hace poco y se desconocen los «efectos netos».

La **viabilidad** se refiere tanto a las características de las intervenciones como a las circunstancias en las que se introducen. Los responsables de la planificación deben poder responder afirmativamente a las siguientes preguntas:

- ¿Es el tratamiento de la intervención específico, distintivo y suficientemente homogéneo?
- ¿Es suficientemente significativa la comparación entre el grupo de tratamiento y el grupo de control para medir el impacto?
- ¿La población destinataria es amplia y está claramente definida?
- ¿La teoría que vincula la intervención con los efectos directos previstos es coherente desde el punto de vista lógico?
- ¿Puede diferenciarse claramente el grupo de tratamiento dentro de la población destinataria?
- ¿Es suficientemente grande el grupo de tratamiento?
- ¿Es posible definir un grupo de control creíble?
- ¿Puede mantenerse la diferencia entre las experiencias del grupo de tratamiento y el grupo de control durante un período de tiempo suficiente para medir el impacto?

Los **datos** son cruciales. La esencia de la EIC es la medición, para la cual se requiere información cuantitativa, tanto sobre el grupo de tratamiento y el grupo de control como sobre el contexto en el que se lleva a cabo la evaluación. Por lo general, la teoría de la intervención y la estrategia utilizada para establecer la hipótesis contrafactual determinan qué datos se necesitan. Al seleccionar las intervenciones que serán sometidas a una EIC, la AG que la esté planificando debe preguntarse lo siguiente:

- ¿Qué es esencial saber sobre los miembros del grupo destinatario y el grupo de control?

- ¿Qué es esencial saber sobre la naturaleza de la intervención tal y como se lleva a cabo en el grupo de tratamiento?
- ¿El grupo de control recibe algún tratamiento, o ninguno en absoluto? ¿Se dispone de datos al respecto?
- ¿De qué datos de fuentes administrativas y de otra índole se dispone?
- ¿Se dispone de datos que describan las trayectorias individuales?
- ¿Pueden vincularse datos individualizados procedentes de diversas fuentes?

Más adelante en este mismo capítulo se ofrece información más detallada sobre estas cuestiones.

### **2.1.1. Priorizar las intervenciones que se someterán a una evaluación del impacto**

Antes de priorizar las intervenciones concretas que serán sometidas a una EIC, deben tenerse en cuenta cuestiones estratégicas más amplias. La atención debe centrarse en seleccionar aquellas intervenciones para las que las evaluaciones del impacto sean la mejor apuesta para descubrir qué funciona. Unas evaluaciones bien diseñadas y rigurosas no solo son beneficiosas para la AG y el Estado miembro que las encargan, sino también para otros Estados miembros y sus AG, otras partes interesadas y la Comisión.

#### **Contribución a la justificación de los gastos**

Dado que las EIC se centran en responder a preguntas cruciales para los responsables políticos, y en particular para los responsables de las decisiones relativas a la asignación de los recursos, es lógico centrar los esfuerzos de evaluación del impacto en aquellos programas e intervenciones que requieren una gran cantidad de recursos. Cuanto más tiempo y otros recursos absorbe un programa o intervención particular, más importante resulta comprender si funciona y, por tanto, si los beneficios generados superan los costes incurridos. Es posible que deban replantearse las intervenciones costosas que no generan valor social o económico, mientras que tal vez proceda incrementar la financiación y la atención que se brinda a otras de probado valor añadido.

*Atención a las intervenciones que requieren una gran cantidad de recursos*

Los resultados de las recientes evaluaciones de las intervenciones del FSE financiadas en el período de programación 2014-2020 han puesto de manifiesto que es necesario realizar ajustes estratégicos y concentrarse más en objetivos políticos clave. Se ha demostrado que las intervenciones en materia de empleo y movilidad laboral son menos eficaces para las personas de edad avanzada y para las personas más alejadas del mercado laboral, unos grupos que necesitan instrumentos nuevos y más efectivos<sup>20</sup>. Aunque las intervenciones en materia de política social muestran una gran variación en lo que respecta al coste por participante y tipo de operación, por lo general no se efectúa un análisis coste-beneficio sistemático<sup>21</sup>. Las intervenciones en el ámbito de la educación y la formación arrojaron resultados positivos, pero el limitado número de estudios de impacto impide evaluar por completo sus efectos a largo plazo<sup>22</sup>. Las EIC ofrecen la posibilidad de cribar las intervenciones a fin de dilucidar cuáles son los enfoques más eficaces y

<sup>20</sup> [Fondazione G. Brodolini, Metis GmbH, Applica, Ockham IPS, 2020.](#)

<sup>21</sup> [ICF, Cambridge Econometrics y Eurocentre, 2020.](#)

<sup>22</sup> [Ecorys e Ismeri Europa, 2020.](#)

eficientes para determinados grupos destinatarios, maximizando así la «relación calidad-precio» de los nuevos programas del FSE+.

### **Contribución de una intervención a un proceso de reforma**

A menudo, las intervenciones que son un componente clave de un programa de reforma más amplio obtendrán una financiación considerable. Como es natural, el hecho de que una intervención del FSE sea fundamental para una estrategia de inclusión social, o un componente crucial de un programa activo del mercado de trabajo, hará que se le preste más atención.

*Intervenciones que contribuyen a la innovación política*

### **Innovadoras y experimentales**

Las intervenciones piloto nuevas e innovadoras son claras candidatas a ser sometidas a una EIC. Para comprobar los efectos de intervenciones mediante una prueba piloto o ensayo es preciso efectuar una evaluación rigurosa. La evaluación a través de una EIC bien diseñada es aún más importante cuando existe la determinación clara de expandir o desplegar la intervención de forma más amplia si se acaba determinando que es eficaz.

### **Contribución al aprendizaje**

La necesidad de centrar la atención y los recursos en áreas específicas de los programas, así como en intervenciones específicas en estas áreas, es mayor cuando existen pocos datos o ninguno en absoluto sobre lo que funciona en el ámbito de actuación en cuestión. Por ejemplo, cuando hay una verdadera incertidumbre respecto de la política en el futuro y existe el riesgo de depender en exceso de pruebas que quizás no sean directamente pertinentes (p. ej., pruebas de otros países).

*Generar pruebas*

Las evaluaciones de gran calidad pueden considerarse un bien público. Los beneficios que generan en términos de aprendizaje se extienden a partes interesadas más allá de las de una AG específica. En consecuencia, es importante valorar qué partes interesadas podrían beneficiarse de la evaluación del impacto propuesta. Podría tratarse de organismos intermediarios (OI) o de órganos encargados de intervenciones en el marco del mismo programa, de otras AG o de OI en el Estado miembro en cuestión, o de órganos e instituciones responsables de los fondos nacionales o regionales. Otro interesado externo obvio al que se debe tener en cuenta es la Comisión Europea, y también hay partes interesadas de otros Estados miembros que podrían aprender de una evaluación. Los responsables políticos y los responsables de programas pueden realizar una importante contribución al aprendizaje mutuo con solo tener en cuenta las necesidades de interesados que no sean las partes interesadas inmediatas.

Una última consideración estratégica que debe hacerse a la hora seleccionar los ámbitos a los que debe prestarse atención cuando se prepare una EIC es tomar como referencia las intervenciones que puedan mostrar los beneficios de las EIC y utilizarlas como modelo.

*Defender los métodos de la EIC*

#### **2.1.2. Seleccionar las intervenciones adecuadas para un enfoque contrafactual**

Tras examinar cuestiones estratégicas más amplias que podrían motivar la selección de intervenciones concretas para someterlas a una EIC, en esta sección se analizan las características específicas de las intervenciones que podrían hacerlas adecuadas para un enfoque contrafactual. Estas características son numerosas y variadas. Mientras que algunas particularidades de una intervención podrían hacerla idónea para una EIC en

determinadas circunstancias, en otras ocasiones frustran los intentos de llevar una a cabo. Por este motivo, no es posible proporcionar una lista exhaustiva de los aspectos que deben tenerse en cuenta. No obstante, sí merece la pena mencionar las características de las intervenciones que hacen más probable que una EIC tenga éxito.

#### **Recuadro 4 Tipos más comunes de intervenciones y grupos destinatarios elegidos para EIC del FSE**

La mayoría de las EIC de intervenciones del FSE se centran en políticas activas del mercado de trabajo dirigidas a las personas desempleadas y a los subgrupos de estas afectados por una desventaja específica. La gran cantidad de análisis del impacto de intervenciones dirigidas a los jóvenes es el resultado de los reglamentos de la IEJ, que exigen que dichas evaluaciones se lleven a cabo a intervalos específicos.

Aproximadamente la mitad de las EIC identificadas en el proyecto del Servicio de Asistencia para las Evaluaciones<sup>23</sup> desde 2015 están relacionadas con el objetivo temático 8: «Fomento del empleo sostenible y de calidad y apoyo a la movilidad laboral». Del mismo modo, en la mayoría de las EIC que examinan intervenciones financiadas en el marco del objetivo temático 9, la atención se centra en lo eficaces que son las intervenciones para integrar a las personas desempleadas vulnerables en el mercado laboral (por ejemplo, la intervención alemana destinada a los parados de larga duración financiada en el marco del PO Bund del FSE 2014-2020).

La formación, las prácticas u otras formas de adquirir experiencia laboral o acceder a empleo subvencionado son los tipos de apoyo a las personas desempleadas que se analizan con más frecuencia. Así se desprende claramente de los ejemplos analizados en la presente Guía (véanse los ejemplos de la evaluación italiana en Las Marcas, la evaluación de los cursos de formación para migrantes en Alemania y la formación profesional para ninis en Letonia). Es menos habitual que se analicen los servicios de asesoramiento y adecuación entre la oferta y la demanda. Es el caso del ejemplo sueco que figura en la presente Guía, en el que se evalúa una actuación piloto para reforzar el apoyo a los desempleados que prestan los SPE mediante un enfoque aleatorizado. Las intervenciones del FSE en apoyo del trabajo por cuenta propia o la creación de empresas no suelen evaluarse, según los datos del Servicio de Asistencia<sup>24</sup>.

En algunos casos, las EIC analizan conjuntamente distintos tipos de intervenciones, con lo que se corre el riesgo de mezclar diferentes lógicas de intervención y se reduce la fiabilidad de la comparación entre el grupo de tratamiento y el grupo de control.

Cabe destacar el intento de evaluar los efectos de las intervenciones para las personas vulnerables, no solo en términos de resultados en materia de empleo, sino también en lo que respecta a los efectos directos difíciles de cuantificar. Una EIC llevada a cabo en Alemania se centró en este aspecto, pues evaluó el impacto de los planes de creación de empleo para parados de larga duración en distintas dimensiones subjetivas de la salud, la satisfacción con la vida, el sentimiento de pertenencia y los indicadores de estatus social de estas personas.

Es menos habitual que las intervenciones del FSE en el ámbito educativo se evalúen mediante un enfoque contrafactual. Se realizan menos EIC en relación con el objetivo temático 10 que con el objetivo temático 8 o 9, y a menudo estas se centran en las intervenciones financiadas en el marco de la PI 10.IV, relativas a la formación profesional y sus efectos en lo tocante a la integración en el mercado laboral. Dos razones que lo explican son la accesibilidad más limitada a los conjuntos de datos y unas normas de privacidad más estrictas<sup>25</sup>. No obstante, se han realizado intentos en este sentido: en Asturias (España), se evaluó una intervención en centros de enseñanza secundaria orientada a desincentivar el abandono escolar prematuro; en Podalskie (Polonia), se llevaron a cabo EIC para examinar los efectos de un proyecto destinado a promover la formación profesional entre los jóvenes estudiantes; y en Portugal se analizaron unas becas dirigidas a alumnos de enseñanza superior.

Las EIC pueden llevarse a cabo a distintos niveles políticos (es decir, aplicarse a uno o varios ejes prioritarios, objetivos específicos u

<sup>23</sup> Véase [https://ec.europa.eu/regional\\_policy/en/policy/evaluations/member-states/](https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/).

<sup>24</sup> Algunos ejemplos son: [Borik, V. et al., 2015](#); [Ires Piemonte, 2019](#), y [Openfield, 2019](#).

<sup>25</sup> Véase, por ejemplo, [Ismeri Europa, Ecorys e Instituto de Estudios de Empleo, 2019](#).

operaciones<sup>26</sup> de un programa), y pueden abarcar programas nacionales o regionales del FSE+, centrarse en grupos destinatarios homogéneos (hombres o mujeres, jóvenes o poblaciones vulnerables, parados de larga duración, etc.) o en tipos concretos de intervención (p. ej., formación, servicios de inclusión social o desincentivación del abandono escolar prematuro [véase el recuadro anterior]).

Los ejemplos de los Estados miembros indican que varios de los instrumentos utilizados en el FSE+ son adecuados para efectuar una EIC, como la formación, los incentivos al empleo y los servicios relacionados con el mercado de trabajo (p. ej., asesoramiento laboral, orientación). Por otra parte, las intervenciones de rotación de puestos y de trabajo compartido, los incentivos para empresas emergentes o el apoyo a sistemas y estructuras, así como las intervenciones en los ámbitos de la educación y la inclusión social, son más difíciles de someter a una EIC.

Resulta ilustrativo valorar qué intervenciones son más prometedoras desde el punto de vista de la EIC, atendiendo a las siguientes preguntas:

**¿Es la intervención específica, distintiva y suficientemente homogénea?**

El tratamiento o los tratamientos proporcionados a través de una intervención deben distinguirse de los de otras intervenciones. Además, debe existir un contraste notable entre lo que reciben los participantes en una intervención y los beneficios que obtienen otros grupos similares de personas. Si los tratamientos son difusos hasta el punto de que no es posible determinar un grupo específico de beneficiarios, entonces no es posible o deseable adoptar enfoques contrafactuales.

*Tratamiento  
claramente  
distinguible*

Los métodos de EIC se vuelven muy complejos y difíciles si el estado de tratamiento de una unidad concreta (una empresa o una persona) afecta al posible resultado de otras unidades (a través de los denominados «efectos de equilibrio general» más amplios). En los programas de formación, esto puede ocurrir cuando quienes han completado el programa dificultan a corto plazo que otras personas que no han recibido la formación encuentren empleo. Cuando se considere que esto constituye un problema importante (por ejemplo, en el caso de las intervenciones a gran escala), tal vez sea necesario efectuar un análisis macroeconómico para evaluar el alcance de los efectos de sustitución y desplazamiento. Las AG deben solicitar asesoramiento especializado cuando sea probable que se produzcan tales efectos.

La propia intervención debe ser relativamente homogénea. Esto significa, que, en términos generales, todos los participantes en una intervención deben recibir el mismo paquete de medidas o estar expuestos a él. Si las medidas aplicadas a los participantes en una única intervención son demasiado variadas, esto tiene diversas implicaciones para la EIC. En primer lugar, en realidad quizás no tenga sentido hablar de una intervención coherente, sino más bien de intervenciones con procesos causales independientes en curso. En segundo lugar, será difícil interpretar los

*Intervenciones  
homogéneas*

<sup>26</sup> De conformidad con el artículo 2, apartado 4, del Reglamento (UE) 2021/1060 del Parlamento Europeo y del Consejo, «operación» se refiere a: a) un proyecto, contrato, acción o grupo de proyectos seleccionados en el marco de los programas de que se trate; b) en el contexto de los instrumentos financieros, una contribución de un programa a un instrumento financiero y el posterior apoyo financiero prestado a los perceptores finales por dicho instrumento financiero.

impactos que se notifican como efectos netos medios de un grupo de intervenciones dispares. En tercer lugar, aunque podría estar justificado realizar un análisis de los subgrupos, si hay demasiados dentro de un grupo de tratamiento, las limitaciones del tamaño de la muestra podrían restringir la capacidad de comunicar las conclusiones que puedan utilizarse.

### **¿Se compara el tratamiento con la ausencia de tratamiento o existen otras formas de tratamiento pertinentes?**

El FSE cofinancia políticas nacionales y regionales de inclusión social y relacionadas con el mercado laboral. Así pues, cualquier sistema de EIC debe sopesar con cuidado si la intervención es claramente identificable y si las personas tienen la oportunidad de recibir servicios de otros programas (nacionales o regionales) y fuentes de financiación. Lo importante es que los tratamientos evaluados modifiquen realmente las oportunidades o los recursos de que disponen los participantes en comparación con los que están a disposición del grupo de control, y que las diferencias puedan medirse y vigilarse.

Estos problemas de «tratamiento complejo» suelen ser específicos del contexto, y complican el diseño y la realización de las EIC. Su presencia pone de relieve la importancia de planificar minuciosamente la evaluación, mediante el desarrollo del sistema de evaluación, antes de llevarla a cabo.

*Tratamientos complejos*

### **¿Existe un grupo destinatario amplio y bien definido?**

Las EIC requieren una muestra de gran tamaño, en comparación con otras formas de evaluación. Es esencial que los grupos destinatarios estén integrados por un número suficiente de personas, y debe ser posible localizar grupos de control de un tamaño suficiente. Esta cuestión se examina en detalle más abajo.

*Tamaño de la muestra grande*

Es importante que la intervención que se plantee someter a una EIC se dirija a un grupo bien definido. Si no se tiene una idea clara de quiénes son los grupos destinatarios, es difícil definir un grupo de control significativo. Algunas intervenciones tratan deliberadamente de captar a personas para recibir tratamiento a través de mecanismos oficiosos, alentando procesos que no están predefinidos o no son demasiado prescriptivos (p. ej., se puede involucrar a personas procedentes de medios desfavorecidos, como jóvenes ninis o grupos desfavorecidos, mediante procedimientos ocasionales y caso por caso), lo que puede dificultar la identificación precisa de las personas tratadas y del grupo de control correspondiente.

*Determinar la identidad del grupo destinatario*

### **¿Existe un mecanismo causal claro?**

Como se ha mencionado anteriormente, cuando se presentaron los principales métodos de evaluación y la «teoría del cambio», a menudo resulta útil llevar a cabo una evaluación basada en la teoría antes de la EIC o en combinación con esta. Desarrollar una teoría del cambio, o una lógica detallada de intervención, puede ayudar de distintas maneras a los encargados del diseño de una EIC y, lo que es más importante, puede servir para determinar si una intervención se sustenta en un mecanismo causal coherente. Es poco probable que las intervenciones sin un mecanismo causal claro y convincente produzcan unos impactos de una magnitud suficiente para ser identificados estadísticamente a través de una EIC.

*Mecanismo político diferenciado*

### ¿Es posible definir cuantitativamente los efectos directos?

Es necesario obtener medidas cuantificables de los efectos directos. Dichos datos e indicadores pueden proceder de fuentes administrativas o de encuestas específicas.

*Necesidad de medir los resultados*

En algunas circunstancias, los resultados previstos de una intervención podrían requerir disposiciones específicas que se midan cuantitativamente. Por ejemplo, una intervención podría perseguir un cambio de actitudes, creencias u opiniones. En tales casos, deben administrarse encuestas para medir estos cambios. Algunas intervenciones tienen resultados bastante vagos o poco definidos. Una vez más, el desarrollo de una lógica de intervención puede ayudar a comprender mejor qué se quiere conseguir con una intervención y cómo se pretende generar cambios en los resultados de interés.

### ¿La intervención se ha presentado de un modo que permite encontrar un grupo de control significativo?

Para establecer un grupo de control significativo, es importante valorar cómo se seleccionan las unidades tratadas (personas o empresas) para una intervención o por qué deciden participar; si puede proporcionarse la misma fuente de datos (p. ej., la misma encuesta) a la muestra de control y el grupo de tratamiento; y, por último, si es necesario seleccionar muestras de control que estén sujetas a las mismas condiciones del mercado laboral que el grupo de tratamiento. En el recuadro que figura a continuación se ponen de relieve algunos ejemplos.

*Mecanismo de selección para el tratamiento*

Si una intervención es obligatoria y se lleva a cabo para toda la población destinataria de forma más o menos simultánea, podría resultar difícil localizar una parte no tratada de la población destinataria para que actúe como grupo de control.

#### Recuadro 5 Definir los grupos de control

En los ejemplos de EIC examinados en esta Guía, la selección de los grupos de control se basó en las características de las intervenciones (por ejemplo, los criterios de elegibilidad) y en la disponibilidad de datos adecuados.

En comparación con la experiencia con las EIC en el período de programación anterior, la determinación y selección del grupo de control se basa más a menudo en datos administrativos, especialmente en los registros de desempleo. La estrategia más común consiste en encontrar a personas con características similares registradas como desempleadas en los SPE durante un período determinado, como exigen los criterios de elegibilidad, que podrían formar parte del grupo de control. Esto se aplica a algunas evaluaciones en Italia (Las Marcas, provincia de Trento y Piamonte), Polonia (Lubelskie y Podlaksie), así como a las evaluaciones de cursos de formación para migrantes, planes de creación de empleo y medidas de integración para la población desempleada llevadas a cabo en Baden-Württemberg (Alemania).

En otras evaluaciones, aunque no están cubiertas por los ejemplos expuestos en detalle en la presente Guía, las personas que se inscribieron y no resultaron elegidas para recibir tratamiento fueron seleccionadas posteriormente para formar parte del grupo de control. Así ocurrió en el caso de la evaluación italiana de la Garantía Juvenil, la evaluación alemana del proyecto piloto de Trabajo Ciudadano financiado en 2007-2013, y la evaluación del programa PIPOL ejecutado en Friuli-Venecia Julia<sup>27</sup>. Sin embargo, no se suele recurrir a esta estrategia, ya que en la mayoría de los países y regiones los sistemas de información de seguimiento no contienen información sobre las personas que solicitaron participar en

<sup>27</sup> Véanse, respectivamente, [Isfol, 2016](#); [IAW Institut für Angewandte Wirtschaftsforschung, ISG Institut für Sozialforschung und Gesellschaftspolitik, GmbH, 2015](#), e [Ismeri Europa, 2018](#).

un programa, pero finalmente no lo hicieron.

Alemania enfocó la evaluación de las intervenciones dirigidas a parados de larga duración<sup>28</sup> de manera diferente, ya que evaluó la «intención de tratar», en lugar de los «efectos promedio del tratamiento en las personas tratadas». En este caso, el grupo tratado estaba compuesto por personas potencialmente elegibles durante todo el período de ejecución del programa, con independencia de si realmente fueron tratadas o no (a partir de 2015), y el grupo de control estaba integrado por personas que reunían los mismos criterios de elegibilidad, pero que eran paradas de larga duración antes de la ejecución del programa (2010-2012).

Solo en el ejemplo sueco, que aplicaba un enfoque aleatorizado, el grupo de control se compuso aleatoriamente: por diseño, el tratamiento (refuerzo del apoyo prestado por los SPE) se proporcionó aleatoriamente a un grupo de jóvenes, mientras que el grupo de control recibió el apoyo que los SPE prestaban habitualmente.

En las pocas evaluaciones centradas en las medidas del FSE dirigidas a empresas, se estableció la línea divisoria entre el grupo de tratamiento y el grupo de control entre los solicitantes que reciben financiación y los que no, como en algunas evaluaciones danesas, que comparan los resultados de las empresas financiadas a través del FSE con los de una muestra de empresas con características similares, pero que no recibieron dicho apoyo<sup>29</sup>.

## 2.2. Preguntas de evaluación y variables de resultado



### 2.2.1. ¿Cuáles son los fines y objetivos de la intervención?

Cuando se establezca un sistema de evaluación, es aconsejable, en primer lugar, describir los fines, objetivos y características clave de la intervención.

En muchos casos, ya existirán documentos que indiquen sus fines y objetivos. Sin embargo, en el caso de las EIC, es importante especificar los resultados y los cambios que se quieren lograr con la intervención y, por tanto, los impactos esperados.

A menudo resulta beneficioso articular la teoría del cambio de una intervención, que establece cómo se prevé que sus diferentes insumos y actividades estén vinculados a los productos, los efectos directos (o resultados) y, por consiguiente, los impactos.

### 2.2.2. ¿Cuál es la finalidad de la evaluación?

Cuando se desarrolle un sistema de evaluación para una EIC, es importante reflexionar detenidamente acerca de la finalidad de la evaluación. Si no se tiene una idea clara de por qué es necesaria la evaluación, es poco probable

*Combinar el diseño de la EIC con las perspectivas de la lógica de intervención*

<sup>28</sup> Véase [Bratu, C. et al., 2019](#).

<sup>29</sup> Por ejemplo: [Danmarks Statistik et al., 2017](#), y [Danmarks Statistik et al., 2018](#).

que aporte las pruebas necesarias. En el contexto de las evaluaciones de intervenciones financiadas por el FSE, deben tenerse en cuenta una serie de preguntas:

- ¿Cuál es la finalidad y la naturaleza de la evaluación en el contexto de los requisitos reglamentarios y las directrices de la CE?
- ¿Quiénes son las principales partes interesadas de la evaluación?
- ¿Qué uso se hará de los resultados de la evaluación?
- ¿A qué preguntas específicas deberá responder la evaluación?

### ¿Cuáles son los fines y la naturaleza de la evaluación?

En primer lugar, hay que definir **la motivación** para llevar a cabo la evaluación. De conformidad con el Reglamento (UE) 2021/1060, «el Estado miembro o la autoridad de gestión llevará a cabo evaluaciones de los programas en relación con uno o varios de los siguientes criterios: eficacia, eficiencia, pertinencia, coherencia y valor añadido de la Unión, con el fin de mejorar la calidad del diseño y la ejecución de los programas. Las evaluaciones podrán también llevarse a cabo en relación con otros criterios pertinentes, como la inclusividad, la no discriminación y la visibilidad, y referirse a más de un programa»<sup>30</sup>. Como se ha mostrado anteriormente, las conclusiones de una EIC suelen referirse a la eficacia (la medida en que se han logrado los efectos directos previstos) y la eficiencia (la rentabilidad o la relación coste-beneficio de la intervención).

*El fin y la naturaleza de la evaluación*

En términos más generales, la CE anima a los Estados miembros a que sigan enfoques orientados a los resultados cuando formulen sus políticas y a que lleven a cabo evaluaciones que satisfagan las exigencias internas de los Estados miembros en lo que respecta a su alcance, diseño y calendario. En este sentido, en los programas del FSE+, también pueden introducirse EIC para responder a preguntas de evaluación específicas o de conformidad con la política nacional de evaluación.

En segundo lugar, es necesario determinar **la naturaleza de la evaluación**<sup>31</sup>:

- Las **evaluaciones del impacto** examinan los efectos de un programa, o un conjunto de programas, en relación con las prioridades nacionales y de la UE (pueden estimar el impacto macroeconómico del FSE, o centrarse en políticas y temas específicos o en prioridades horizontales como la infancia y la igualdad de oportunidades).
- Las **evaluaciones del proceso (o la ejecución)** apoyan la ejecución de un programa, para lo cual analizan los avances y los métodos de ejecución y formulan recomendaciones para mejorarlo.

En principio, el enfoque contrafactual puede aplicarse a las evaluaciones del impacto, mientras que para efectuar una evaluación del proceso hay que recurrir a otros métodos (véase también el Gráfico 1).

El RDC no exige un número específico de evaluaciones del impacto en el período de programación 2021-2027, como sí hace para 2014-2020. En su lugar, solicita que se adopte una estrategia de evaluación capaz de evaluar cómo el apoyo de los fondos europeos ha contribuido a alcanzar los objetivos

<sup>30</sup> Artículo 44, apartado 1, del Reglamento (UE) 2021/1060 o Reglamento sobre Disposiciones Comunes (RDC).

<sup>31</sup> Véase Comisión Europea, 2007.

del programa en relación con todos los principales perfiles estratégicos<sup>32</sup>. Además, el RDC no determina en qué prioridades o intervenciones debe centrarse la evaluación, sino que deja esta decisión en manos de la AG y del plan de evaluación. Esto significa que la estrategia de evaluación de cada programa debe definir la combinación de evaluaciones del impacto y del proceso y en qué prioridades e intervenciones deben concentrar sus principales esfuerzos. Antes de junio de 2029<sup>33</sup>, debe realizarse una evaluación del impacto final y general, pero no hay otras restricciones que delimiten los plazos para efectuar las demás evaluaciones, que deben establecerse en el plan de evaluación.

### ¿Quiénes son los principales destinatarios?

Es preciso definir a quién está dirigida la evaluación. En función de la naturaleza de la evaluación, podría incluir a responsables políticos, AG y responsables de programas, otras AG u organismos de ejecución del Estado miembro, o autoridades nacionales o regionales que gestionen programas similares. Cuando instituciones ajenas a la gestión del programa sean las que proporcionen los datos, también debe considerarse que estos organismos (los «propietarios de los datos») son partes interesadas. Es importante incluir a todas las principales partes interesadas en un grupo director de una evaluación, a fin de que el proceso de diseño y realización de la evaluación sea una responsabilidad conjunta, así como a algunos expertos en la evaluación procedentes de instituciones académicas o públicas para que proporcionen asesoramiento técnico a la AG.

*Especificar las partes interesadas*

### ¿Qué uso se hará de los resultados de la evaluación?

Una vez que se determine quiénes son los destinatarios de la evaluación, puede definirse qué uso se dará a los resultados. En la práctica, esto puede lograrse implicando al grupo director en la formulación de las preguntas de evaluación y en los debates sobre los términos de referencia.

Los resultados de una EIC contribuyen con frecuencia a dos decisiones clave:

- si una intervención en curso debe continuar, y
- si debe llevarse a cabo un nuevo tipo de intervención de forma más amplia (p. ej., mediante su expansión).

En el primer caso, una EIC podría tratar de evaluar la eficacia de un programa existente o en curso en tiempos de tensión presupuestaria, y si los recursos utilizados podrían usarse de otro modo. En esta situación, es poco probable que la intervención se haya evaluado previamente mediante hipótesis contrafactuales.

En el segundo caso, podrían existir limitaciones en la ejecución de las intervenciones. Por ejemplo, es posible que una intervención se lleve a cabo en una región o zona concreta de un Estado miembro, o solo durante un período de tiempo limitado. En estos contextos, los resultados de una EIC podrían utilizarse para determinar si la intervención en cuestión es eficaz y, por lo tanto, puede llevarse a cabo con éxito en otro lugar. En esas situaciones, se dice que las intervenciones se han puesto a prueba o comprobado antes de un despliegue más amplio.

<sup>32</sup> Artículo 44, apartado 1, del RDC.

<sup>33</sup> Artículo 44, apartado 2, del RDC.

## ¿A qué preguntas se debe responder?

Una vez que se han establecido los objetivos de la intervención y la finalidad y los usos fundamentales de la evaluación, y se ha definido claramente quiénes son los destinatarios, debería ser posible especificar con cierto grado de detalle las preguntas a las que tendrá que responder la EIC. En muchas circunstancias, distintos públicos y partes interesadas tendrán preguntas de carácter causal que querrán que se exploren en la EIC. La AG o el evaluador deben recopilar estas preguntas a través de una amplia consulta, teniendo en cuenta los distintos puntos de vista y las sugerencias de las personas implicadas en las diversas fases de la intervención. Posteriormente, la AG o el evaluador deben decidir qué prioridad otorgan a cada pregunta y centrar la EIC en las más importantes y oportunas.

Estas son algunas de las cuestiones que se podrían tener en cuenta para ultimar la lista de las principales preguntas de investigación para una EIC:

*Preguntas clave de investigación*

- ¿La intervención dio lugar o contribuyó a los efectos directos previstos a corto, medio y largo plazo? ¿Los efectos a corto plazo diferían notablemente de los efectos a largo plazo? Debe darse prioridad a las preguntas que aborden estas cuestiones.
- ¿Es posible disponer de una medición cuantitativa de los efectos directos? La EIC debe fundamentarse en un conjunto adecuado de datos (datos administrativos o encuestas directas enviadas a los participantes), independientemente de la fuente preferida.
- ¿En qué medida pueden atribuirse a las intervenciones los cambios en las circunstancias de los participantes o en el contexto socioeconómico? Este tipo de pregunta requiere que se midan los efectos netos de la intervención para compararlos con los participantes de control y los indicadores de contexto.
- ¿Los efectos de la intervención fueron los mismos para todos los miembros del grupo destinatario? Por ejemplo, ¿una intervención dirigida a los parados de larga duración tuvo el mismo impacto en hombres y mujeres? ¿Y en los menores de veinticinco años y los mayores de cincuenta? La capacidad de la EIC para examinar los efectos en diferentes subgrupos resulta valiosa si los grupos de tratamiento y de control están integrados por un número lo suficientemente elevado de personas.
- ¿Ha sido rentable la intervención (en comparación con las alternativas)? ¿Cuál es su relación coste-beneficio? Este análisis de la eficiencia requiere que la EIC mida con exactitud los efectos, así como información precisa sobre los costes y beneficios directos e indirectos.
- ¿Se dispone de abundante información y conocimientos sobre los impactos de intervenciones similares? ¿En qué medida este conocimiento es aplicable a la intervención objeto de examen? Cuando una política se conozca sobradamente y sus efectos se hayan investigado en profundidad, podría ser útil centrar las preguntas de evaluación en aspectos específicos de la política, con lo que se evitaría repetir otros análisis. Un análisis detallado de la bibliografía existente puede servir de base para las decisiones a este respecto y, en cualquier caso, constituye un apoyo útil a la hora de diseñar la EIC.

Es importante tener una idea clara de las diversas preguntas de investigación que tendrá que responder una EIC antes de poder llevarse a cabo. La celebración de diálogos acerca de las preguntas que tratará la evaluación es un elemento clave de cualquier sistema de evaluación.

Es importante determinar qué preguntas son prioritarias y no sucumbir a la tendencia popular de sobrecargar una evaluación con demasiadas. Conviene encontrar un equilibrio entre garantizar que la evaluación sea pertinente para múltiples partes interesadas con intereses diferentes y hacerla manejable. Si una evaluación responde a una cantidad excesiva de preguntas de investigación, puede perder el norte y acabar abordando una amplia gama de cuestiones de una manera que diste de ser óptima. A menudo, «menos es más»: la definición de las prioridades es una fase crítica del proceso de planificación de la evaluación.

*Asignar una prioridad a las preguntas*

Para determinar qué prioridad tienen las distintas preguntas de evaluación, es necesario excluir las duplicaciones y asignar una puntuación a cada pregunta de acuerdo con los principios pertinentes. Algunos de ellos pueden ser la importancia y el compromiso real de la parte interesada que formuló la pregunta, la adecuación y congruencia de la pregunta con la teoría del cambio del programa, la pertinencia de la pregunta para el objetivo general de la evaluación, la viabilidad de la pregunta en relación con los datos, el tiempo y los recursos disponibles<sup>34</sup>. La clasificación resultante ordenará las preguntas por importancia y permitirá seleccionar las más pertinentes.

En algunos casos, es posible agrupar múltiples preguntas de evaluación en el marco de una pregunta más general; por ejemplo, bajo el paraguas de una pregunta como: «¿Cuál ha sido el efecto neto de la intervención?». Otras preguntas, como las relativas a los efectos sobre los distintos grupos de participantes, quizás incluso en diferentes períodos de tiempo, podrían agruparse. Sin embargo, este proceso de agrupación siempre debe dar lugar a un conjunto de preguntas manejable y viable.

### ***¿Qué criterios de evaluación pueden asociarse a la evaluación?***

Aunque ya se ha mencionado la relación entre los criterios y las preguntas de evaluación, conviene profundizar más en ella. Los criterios de evaluación (eficiencia, eficacia, valor añadido de la UE, coherencia, etc.) son necesarios para atribuir un valor a cualquier prueba recopilada y evaluar una política (eficiencia, etc.), mientras que las preguntas de evaluación son necesarias para que la solicitud de la persona que desea encargar una evaluación sea explícita y se centre en las principales cuestiones políticas en juego. Sin embargo, existe un nexo entre los criterios y las preguntas de evaluación. Por lo general, cada pregunta se refiere a un criterio específico, y esta relación es importante tanto para las AG como para los evaluadores, ya que vincula el diseño de la evaluación, que por fuerza se basa en las preguntas, con los requisitos contemplados en el RDC para 2021-2027, que están asociados a los criterios de evaluación de una manera diferente.

*Criterios y preguntas de evaluación*

A continuación se presentan algunos ejemplos de las preguntas típicas de evaluación, agrupadas con sus criterios correspondientes. Han sido seleccionadas y adaptadas de la caja de herramientas para la mejora de la legislación<sup>35</sup> (herramienta 47, relativa a los criterios y preguntas de

<sup>34</sup> Véase, por ejemplo, [Centros para el Control y la Prevención de Enfermedades \(CDC\), 2013](#).

<sup>35</sup> Véase [Comisión Europea, 2017](#).

evaluación) de la Comisión Europea, con vistas a mostrar ejemplos que estén en consonancia con las preguntas que es preciso someter a un análisis del impacto y, cuando sea posible, a una EIC:

- *Ejemplos típicos de preguntas sobre la eficacia*
  - ¿Cuáles han sido los efectos cuantitativos de la intervención?
  - ¿En qué medida pueden atribuirse estos cambios o efectos a la intervención?
  - ¿En qué medida pueden vincularse a la intervención de la UE los factores que influyen en los logros observados?
- *Ejemplos típicos de preguntas sobre la eficiencia*
  - ¿En qué medida la intervención ha sido rentable?
  - ¿En qué medida están justificados los costes de la intervención, habida cuenta de los cambios o efectos que ha logrado?
  - Si existen diferencias significativas en términos de costes (o beneficios) entre unos territorios y otros, ¿cuál es la causa? ¿Qué relación guardan estas diferencias con la intervención?
- *Preguntas típicas sobre el valor añadido de la UE*
  - ¿Cuál es el valor añadido resultante de las intervenciones del FSE+, en comparación con el generado por intervenciones nacionales o regionales similares?
  - ¿Cuáles serían las consecuencias más probables de detener o retirar la intervención actual del FSE+?

## 2.3. Datos para especificar el grupo de control y medir las variables de resultado



### 2.3.1. ¿Se dispone de datos adecuados o estos pueden facilitarse?

Las conversaciones mantenidas con las AG y los expertos en evaluación de toda la UE sugieren que el acceso a los datos adecuados es uno de los principales retos para efectuar EIC. Una consideración práctica clave es si se dispone de los tipos de datos necesarios. En esta sección se presenta una clasificación simplificada de los tipos de datos requeridos, junto con un análisis de las fuentes de las que podrían obtenerse o de los tipos de ejercicios primarios de recopilación de datos que podría ser necesario llevar a cabo. También se trata la cuestión crucial de la protección de datos.

Debe hacerse una aclaración importante respecto de la planificación adecuada. En cierto modo, los intentos de llevar a cabo EIC en el pasado se han visto frustrados por la falta de datos, debido a que no se establecieron

*Planificar la recopilación de datos*

planes adecuados con la suficiente antelación. Para las intervenciones existentes, es importante especificar a los miembros de los grupos tratados y no tratados y establecer mecanismos para recopilar datos de ellos, puesto que serán el centro de atención de la evaluación. En el caso de las nuevas intervenciones, deben tomarse medidas tempranas para velar por que se recojan los tipos de datos adecuados en los momentos oportunos.

### ¿Qué tipos de datos se necesitan?

*Principales fuentes de datos*

En términos generales, para realizar una EIC se necesitan tres tipos de datos. En algunos casos, una única fuente de datos puede contener uno o más de estos tipos. En concreto, estos son los registros del grupo de tratamiento y el grupo de control, los registros de los efectos directos y los registros de datos contextuales.

- **Registros del grupo de tratamiento y del grupo de control:** se requieren fuentes de datos que permitan a los evaluadores identificar a unidades concretas del grupo de tratamiento y el grupo de control (empresas, personas o, potencialmente, zonas geográficas).
- **Registros de los efectos directos:** como se muestra en los gráficos 6 y 7 del capítulo 3 de la presente Guía, para efectuar una EIC es necesario que se midan los efectos directos tanto para el grupo de tratamiento como para el grupo de control. Lo ideal sería recabar datos sobre los efectos directos para ambos grupos utilizando los mismos métodos de recopilación y midiendo los resultados en los mismos momentos.
- **Registros de datos contextuales:** se necesitan datos que permitan seleccionar un grupo de control y un grupo de tratamiento con una buena correspondencia entre sí, además de comprobar en el análisis cualquier diferencia restante entre ambos. Es importante recopilar tantos datos como sea posible sobre las características de las unidades y los factores que podrían estar relacionados tanto con la decisión de participar en una intervención como con los posibles resultados, y en particular los indicadores de resultados medidos antes del tratamiento. Entre los datos contextuales también pueden figurar aquellos que describen los mercados de trabajo locales (por ejemplo, las tasas de desempleo locales o mediciones de la rigidez del mercado de trabajo) y aquellos que permitirán el análisis por subgrupos.

En el cuadro 2, que figura a continuación, se presentan estos tres tipos de datos y se indican las fuentes de las que podrían obtenerse. Después de él figuran dos recuadros. En el primero se presentan ejemplos de los datos utilizados en EIC del FSE, mientras que el segundo muestra ejemplos de los conjuntos de datos integrados utilizados para analizar los problemas del mercado laboral y evaluar las políticas de dicho mercado.

**Cuadro 2 Tipos y fuentes de datos**

Tipos de datos	Fuentes
<b>Registros del grupo de tratamiento</b>	<ul style="list-style-type: none"> <li>- Registros de la participación en intervenciones (en general, en poder de los beneficiarios)</li> <li>- Datos de seguimiento del FSE+ (características de la intervención, inicio y finalización, registros de derivación, registros de solicitudes)</li> </ul>
<b>Registros del grupo de control</b>	<ul style="list-style-type: none"> <li>- Datos administrativos tales como registros de la seguridad social, educación y prestaciones por desempleo</li> <li>- Registros de participación (personas que eran elegibles, pero que no participaron por motivos distintos de la elegibilidad<sup>36</sup>)</li> <li>- Encuestas nacionales existentes, como la EPA</li> </ul>
<b>Registros de los efectos directos (necesarios tanto para el grupo de tratamiento como para el grupo de control)</b>	<ul style="list-style-type: none"> <li>- Datos administrativos: p ej., los registros de la seguridad social y el desempleo también pueden utilizarse para generar mediciones de los efectos directos (percepción de prestaciones/ingresos procedentes de la seguridad social), registros fiscales y de seguros sociales (ingresos y efectos directos en materia de empleo)</li> <li>- Registros administrativos del sector educativo (pruebas normalizadas sobre las competencias adquiridas, tasas de titulación, matriculación y asistencia)</li> <li>- Censo oficial de la empresa o registros fiscales (productividad o volumen de negocios antes y después de la formación interna o de las nuevas contrataciones)</li> <li>- Registros de empleo o de producción procedentes de las estadísticas oficiales (en análisis comparativos territoriales para medir los niveles de empleo y PIB)</li> <li>- Estudios a medida de los grupos de tratamiento y de control</li> </ul>
<b>Datos contextuales/variables de control (necesarios tanto para el grupo de tratamiento como para el grupo de control)</b>	<ul style="list-style-type: none"> <li>- Sistemas administrativos; p. ej., registros de prestaciones que proporcionen los historiales de solicitudes previas al tratamiento; registros fiscales y de seguros sociales, ingresos históricos y registros de empleo</li> <li>- Estadísticas oficiales sobre el mercado laboral o la educación (p. ej., encuesta de población activa, datos básicos a escala regional o nacional que también proporcionen microdatos individuales para su preparación específica)</li> <li>- Encuestas del grupo de control y el grupo de tratamiento. Cuando las normas de tratamiento sean claras, se puede especificar <i>ex ante</i> a los grupos de control y se pueden recopilar datos de referencia</li> <li>- Sistemas de seguimiento: en algunas circunstancias, pueden utilizarse sistemas de seguimiento para recabar las mediciones tanto del grupo de tratamiento como del grupo de control, tales como sistemas de solicitud en los que los solicitantes no seleccionados puedan servir de control</li> </ul>

**Recuadro 6 Ejemplos de datos utilizados para EIC**

Entre los ejemplos examinados en detalle en la presente Guía, es probable que las evaluaciones alemanas sean las que se basen en los conjuntos de datos administrativos más nutridos. Las biografías laborales integradas (BLI) aglutinan diversas fuentes administrativas de información gestionadas por la Agencia Federal de Empleo y contienen información sobre la vida laboral de una persona (a excepción del trabajo por cuenta propia) y los episodios de desempleo, sobre las características sociodemográficas, sobre las transferencias (prestaciones por desempleo) y sobre la participación en políticas activas del mercado de trabajo. Las BLI tienen una larga historia y ha sido necesario realizar esfuerzos e inversiones constantes

<sup>36</sup> Son los participantes «elegibles, pero no admitidos». Por lo general, su inadmisión se debe al fin de la financiación o a otras causas externas (p. ej., enfermedad de los profesores, avería del transporte, etc.). Estas personas pueden ser el grupo control de preferencia, dado que cumplen los mismos criterios de elegibilidad y tienen el mismo deseo de participar que los propios participantes. Sin embargo, a menudo no son suficientes para crear un grupo de control, o sus detalles no se registran.

para desarrollar y gestionar el conjunto de datos. En el recuadro 7 se ofrecen más detalles sobre las BLI<sup>37</sup>. En la evaluación alemana de los sistemas de creación de empleo, estos datos se complementaron con una encuesta estructurada en tres rondas para recoger información sobre los efectos directos difíciles de cuantificar (tal y como son percibidos por las personas) en diferentes fases de las intervenciones.

En las evaluaciones italianas examinadas en detalle en la Guía, se utilizaron datos similares. Los archivos de las personas desempleadas inscritas en los SPE se usaron para especificar posibles grupos de control, y se utilizaron datos administrativos sobre los cambios en la situación laboral («Comunicazioni obbligatorie» o «COB»), obtenidos mediante el registro de todos los contratos laborales de empresas y empleadores públicos (excluido el trabajo por cuenta propia), para medir las variables de resultado. Dado que los COB se gestionan a escala regional, la calidad de la información varía. Sin embargo, en los últimos años se han introducido normas a escala nacional y la calidad ha mejorado. En una evaluación (para la provincia de Trento), el evaluador también pudo fusionar estos datos con los datos de la declaración fiscal facilitados por el Instituto Nacional de Seguridad Social (INPS, por sus siglas en italiano), lo que le permitió medir el impacto en términos de ingresos de una persona.

En Polonia, las evaluaciones utilizaron datos administrativos extraídos de los registros de desempleo de los SPE para especificar los grupos de control. Sin embargo, estos datos no contienen información sobre la situación o la vida laboral de las personas, y las evaluaciones tuvieron que basarse en un indicador (que anulase la situación de desempleo en los SPE) o recopilar información a través de encuestas de una muestra de personas, tanto tratadas como sin tratar.

En Letonia, el evaluador pudo fusionar dos de los principales conjuntos de datos administrativos. Los datos de la Agencia Estatal de Empleo de Letonia proporcionaron información sobre los participantes y no participantes inscritos como desempleados en fechas concretas, y los datos de la Administración Tributaria del Estado proporcionaron información sobre las condiciones laborales en diferentes fechas, así como sobre los ingresos de las personas. Con ellos, el evaluador pudo evaluar los efectos tanto en términos de la probabilidad de tener empleo en momentos diferentes como en términos de ingresos.

El único caso de aleatorización, el ejemplo sueco, utilizó tanto datos administrativos para medir variables de resultado, como una encuesta a las oficinas de empleo y a intermediarios para medir los tipos de apoyo prestado y la intensidad de dicho apoyo.

En las evaluaciones en el ámbito educativo se utilizaron datos administrativos para especificar el grupo de control y evaluar las variables de resultado. La excepción fue Polonia, que tuvo que basarse en una encuesta para medir las variables utilizadas como efectos directos.

En general, los ejemplos muestran la importancia de contar con datos administrativos adecuados para efectuar EIC, idealmente tanto para especificar a los grupos de control como para medir las variables de resultado. Desde este punto de vista, las AG que planifiquen EIC deben asegurarse de antemano de que disponen de los datos administrativos necesarios para llevar a cabo las evaluaciones, poniendo en marcha las acciones oportunas para afrontar posibles problemas de accesibilidad, integración de datos u otros aspectos problemáticos.

### Recuadro 7 Ejemplos de bases de datos integradas para EIC

#### El conjunto de datos longitudinales de los demandantes de empleo (JLD) en Irlanda

El conjunto de datos longitudinales de los demandantes de empleo (JLD, por sus siglas en inglés) es una base de datos administrativos gestionada por el Departamento de Protección Social.

El JLD es un intento ambicioso de adaptar los datos administrativos a fines de investigación. Empezó a desarrollarse hace unos diez años, después de que el Departamento de Protección Social encargara a la Escuela Superior de Dublín un estudio preliminar sobre la gestión del registro Live Register<sup>38</sup> y, en términos más generales, datos relativos al mercado laboral. El estudio aportó varias sugerencias para mejorar la

<sup>37</sup> Una experiencia similar es la del conjunto de datos longitudinales de los demandantes de empleo (JLD, por sus siglas en inglés) irlandés, que integró en una misma base de datos información sobre los pagos de asistencia social y seguros sociales, los programas del mercado de trabajo y las vidas laborales. El JLD se utilizó, entre otras, en la evaluación del programa de activación JobBridge financiado en el período de programación 2007-2013. Véase [Indecon, 2016](#).

<sup>38</sup> El Live Register contiene información sobre las personas que se registran en las oficinas locales del Departamento de Protección Social para recibir una prestación o un subsidio para demandantes de empleo o de otra índole [Jobseekers' Benefit (JB) o Jobseekers' Allowance (JA)].

recogida de datos y detectó algunas dificultades, tales como la duplicación de datos en diversos sistemas de datos o la falta de información.

El JLD integra varias fuentes de información: datos de pagos y administrativos del Departamento de Protección Social (p. ej., sobre los pagos de la asistencia social y los seguros sociales a la población en edad laboral incluida en el Live Register) y datos relacionados con los programas de política activa de trabajo gestionados por dicho departamento; datos sobre los programas del mercado laboral gestionados por SOLAS, el organismo nacional de educación y formación; y datos recopilados por las autoridades tributarias. Por lo que se refiere a las intervenciones del FSE, el JLD cubre parcialmente las intervenciones financiadas por el Fondo, pero más concretamente las financiadas por el Departamento de Protección Social y SOLAS<sup>39</sup>.

Los datos procedentes de las fuentes mencionadas se reorganizan como una serie de episodios. Cada episodio comienza cuando la persona inicia un período de desempleo y termina cuando consigue otro trabajo o inicia otra actividad o programa de formación; luego, cuando la situación de esta persona vuelve a cambiar, comienza un nuevo episodio. El inicio de un período de desempleo empieza a contarse cuando una persona solicita prestaciones y subsidios para demandantes de empleo<sup>40</sup>. En el JLD, una ventaja de esta estructura por episodios es que los períodos contiguos de prestaciones y subsidios para demandantes de empleo pueden concatenarse y representarse como un único episodio de desempleo. Cabe señalar que, como se ha visto anteriormente, los episodios pueden solaparse, y el investigador tiene el problema de dar prioridad a un episodio sobre otro.

El JLD ha hecho un seguimiento de las solicitudes de protección social, la activación, la formación y la vida laboral desde 2004 e incluye aproximadamente trece millones de episodios individuales de protección social y trabajo de alrededor de dos millones de personas. Cada episodio tiene una fecha de inicio y una fecha de finalización, así como un código operativo<sup>41</sup> que permite determinar la situación de una persona durante cada período. El JLD abarca un conjunto amplio de «variables»: género, edad, estado civil, ciudadanía, nivel educativo, ocupación anterior, historia de empleo y desempleo (duración y número de episodios) y características de los trabajos (sector, por ejemplo), historial de formación durante la situación de desempleo (tipo, duración y número de episodios), tipo de ayuda (prestación o subsidio para demandantes de empleo), número de hijos, personas dependientes a cargo, tipo de pago familiar (es decir, asignaciones por hijos y personas adultas a cargo, solo adultos, etc.), ingresos, y situación fiscal y geográfica<sup>42</sup>.

Además, identificadores individuales vinculan el JLD con otros datos administrativos. Por ejemplo, para la evaluación de JobsPlus, el JLD se combinó con una base de datos de seguimiento independiente, con información detallada sobre las fechas de inicio y finalización de JobsPlus, el número de días en el Live Register al comienzo de JobsPlus, y el tipo de tratamiento.

Se puede acceder a los datos del JLD principalmente a través de dos canales: contratando investigaciones, lo que ocurre cuando el Departamento de Protección Social inicia un proceso de licitación, o a petición de investigadores con las credenciales adecuadas. En este último caso, las solicitudes de acceso se estudian caso por caso y, de ser aceptadas, debe firmarse un acuerdo jurídicamente vinculante de intercambio de datos entre los investigadores o instituciones y el Departamento de Protección Social. Todos los datos están seudonimizados y solo se transmiten los campos de datos pertinentes para el tema de la investigación. Los investigadores o las instituciones de investigación tienen un acceso limitado a los datos, que deben suprimirse al final del proyecto de investigación o evaluación. Junto con los datos del JLD se

<sup>39</sup> Por ejemplo, los programas en el ámbito de la justicia (los proyectos de remisión juvenil y los proyectos de libertad vigilada para jóvenes, así como los proyectos para la integración y el empleo de los migrantes), la educación (para el acceso a la educación superior y la alfabetización de adultos), la defensa (el Sistema de Apoyo al Empleo de las Fuerzas de Defensa) o el sector comunitario (los Centros de Formación Comunitaria) no están incluidos en el JLD.

<sup>40</sup> La prestación para demandantes de empleo es un pago semanal que el Departamento de Protección Social realiza a las personas inactivas, totalmente desempleadas o que trabajan a tiempo parcial porque su empleador les ha reducido la jornada. Pueden percibirla los demandantes de empleo que hayan pagado un tipo adecuado de seguro social (en Irlanda, PRSI), hayan cotizado lo suficiente en el ejercicio fiscal correspondiente y hayan abonado, como mínimo, un total de 104 cotizaciones. El subsidio para demandantes de empleo es un pago sujeto a condiciones de recursos que se realiza a los demandantes de empleo que están desempleados y no reúnen los requisitos para percibir una prestación para demandantes de empleo, o cuya prestación para demandantes de empleo ha vencido.

<sup>41</sup> Representa una combinación de las actividades de activación y formación, las solicitudes de protección social, y el tiempo en situación de empleo.

<sup>42</sup> Conviene prestar atención al hecho de que algunos datos se refieren a períodos diferentes; por ejemplo, los ingresos se registran con periodicidad anual, mientras que los pagos de las prestaciones y los subsidios se actualizan cada semana.

facilita a los investigadores un libro de códigos, si bien es necesario mejorar sustancialmente la documentación.

El JLD se ha utilizado a lo largo de los años para diversas actividades de investigación y evaluaciones. El Subsidio para Reanudar la Educación (2015), el Programa de Activación JobBridge (2016), el Subsidio para la Reincorporación al Mercado Laboral como Autónomo (2017), JobPath (2019) y JobsPlus (2020) son algunos de los programas que han sido seleccionados para su evaluación.

La estructura y el contenido del JLD han mejorado desde las fases iniciales. No obstante, mantenerlo actualizado, mejorarlo y colmar las lagunas requiere un trabajo constante para, por ejemplo, incluir información más detallada sobre la educación, integrar datos sobre programas específicos (su contenido, finalización, etc.), y mejorar los datos relativos a los ingresos. Un futuro proyecto debe introducir mejoras en el JLD en términos de actualización periódica (por ejemplo, mensual) con la automatización y comprobación del circuito de datos, y adición o sustitución de fuentes de datos para garantizar que se cubra por completo el apoyo al empleo y la formación, otros programas de protección social y datos de los ingresos en tiempo real<sup>43</sup>.

### **La investigación del Instituto de Empleo en Alemania y las biografías laborales integradas (BLI)**

En Alemania, el acceso a datos con fines científicos mejoró tras las reformas del mercado laboral acometidas en 2003-2005, que hicieron hincapié en la evaluación de las políticas y las recomendaciones de la Comisión sobre la mejora de la infraestructura de información entre ciencia y estadística, con miras a crear un centro de datos de investigación en cada productor público de microdatos. En consecuencia, la Agencia Federal de Empleo creó un centro de datos de investigación dentro del Instituto de Investigación sobre el Empleo (Institut für Arbeitsmarkt- und Berufsforschung o IAB) en 2004. El IAB se encarga de extraer datos de procesos administrativos para preparar conjuntos de datos útiles para la investigación empírica.

El IAB gestiona más de quince conjuntos de datos, los cuales están a disposición de la comunidad científica. Los datos emanan de datos administrativos procedentes del sistema de seguridad social, de procesos internos de la Agencia Federal de Empleo y de encuestas realizadas por el IAB<sup>44</sup>.

En relación con los sistemas de seguridad social, todos los empleadores están obligados a notificar una serie de elementos y características de sus empleados, y estos datos proporcionan un amplio conjunto de información sobre la vida laboral de las personas. Algunos de los datos administrativos relativos a los procedimientos internos de la Agencia Federal de Empleo se refieren al seguro de desempleo obligatorio, las prestaciones por desempleo y los períodos de indemnización correspondientes, las reuniones de asesoramiento con personas desempleadas, las ofertas de trabajo y las medidas activas del mercado laboral. El IAB combina los datos en un único conjunto de datos exhaustivo, al que se conoce con el nombre de «biografías laborales integradas (BLI)». Estos datos administrativos empezaron a recabarse en 1975, aunque no se dispone de todas las variables para todo el período de observación. Las BLI pueden prepararse con diversos datos procedentes de encuestas realizadas por el IAB, tales como: el Panel de Establecimiento del IAB; la encuesta alemana sobre vacantes de empleo del IAB; las prácticas de gestión y organización alemanas; la encuesta panel sobre el mercado laboral y la seguridad social; el estudio «Working and Learning in a Changing World» («Trabajar y aprender en un mundo cambiante»); la encuesta para empleados sobre el pago de bonificaciones, los aumentos salariales y la equidad (BLoG); los datos vinculados sobre empleadores y empleados del IAB (LIAB); o la encuesta de empleados para el proyecto «Formación complementaria como parte del aprendizaje permanente» del grupo WeLL.

El IAB actualiza periódicamente sus productos de datos y ofrece diferentes muestras de estas abundantes fuentes de datos administrativos con fines de investigación. Para cada producto de datos, el IAB proporciona documentación detallada en alemán e inglés. La base jurídica para acceder a los datos aparece recogida en el Libro del Código Social alemán (varias versiones). En concreto, en él se prevén cuatro tipos de acceso a los datos para la comunidad científica: los archivos del campus (totalmente anonimizados y útiles únicamente con fines docentes); los archivos de uso científico, unos microdatos

<sup>43</sup> Nos gustaría agradecer a Frank Humphreys, Ciaran Judge, Saidhbhín Hardiman y Krzysztof Gigon, del Departamento de Protección Social, que nos hayan proporcionado información sobre el JLD.

<sup>44</sup> En 2011, se fundó en el IAB el Centro para la Vinculación de los Registros, un proyecto conjunto con la Universidad de Duisburg-Essen financiado por la Fundación Alemana de Investigación. El objetivo del Centro es simplificar la vinculación de los conjuntos de datos sin un identificador específico.

anonimizados que se facilitan a instituciones científicas de Alemania y de los Estados miembros de la UE para proyectos de investigación en el ámbito de la investigación del mercado laboral, pero no con fines docentes ni de investigación comercial. La institución científica que solicite los datos deberá garantizar la seguridad de estos. Además, a los datos poco anonimizados con información más detallada solo se puede acceder *in situ*. El IAB cuenta con zonas de trabajo separadas dentro de un entorno informático seguro en Núremberg y en diversos lugares de Alemania, Estados Unidos y el Reino Unido. Los investigadores tienen acceso directo a los datos, pero solo pueden obtener los productos de sus programas después de que el personal del IAB los revise y autorice su divulgación. La ejecución remota implica que los investigadores preparan sus programas con datos artificiales y luego los cargan en la aplicación de solicitud de trabajo (JosuA). En este proceso, los investigadores nunca ven los datos originales y solo reciben sus resultados. Se dispone de formularios de solicitud normalizados para todos los accesos a datos. Una vez que una solicitud ha sido aprobada, la institución del investigador y el IAB celebran un contrato que regula la utilización de los datos para un proyecto específico en un período determinado. El contrato especifica las normas de protección de datos, cuyo incumplimiento está sujeto a graves sanciones. Algunos de los conjuntos de datos gestionados por el IAB solo pueden usarse *in situ* (por ejemplo, los conjuntos de datos vinculados).

A partir de los datos generados por el IAB, las políticas activas del mercado de trabajo se evalúan periódicamente a través de los métodos empíricos más punteros y, en algunos casos, las conclusiones han dado lugar a un cambio en las políticas<sup>45</sup>. El IAB ofrece a los investigadores acceso a sus conjuntos de datos en Alemania y otros países, y el número de usuarios aumenta constantemente (por ejemplo, en 2016 casi un tercio de todos los acuerdos de uso de datos se concertaron con una instalación no alemana).

Para obtener información sobre el uso de datos administrativos para realizar EIC y las cuestiones prácticas conexas, véase también la Guía europea elaborada por investigadores del JRC: [Comisión Europea, 2020](#).

### ¿Qué problemas podrían surgir en relación con la protección de datos?

Pueden surgir dificultades a la hora de obtener datos que identifiquen a las personas o empresas que han participado en intervenciones financiadas por el FSE+<sup>46</sup>. Para llevar a cabo una EIC se necesitan microdatos, unos datos que contienen observaciones sobre cada unidad que integra el grupo de tratamiento y el grupo de control.

El anexo I del Reglamento (UE) 2021/1057 sobre el FSE+, indica que los datos sobre los participantes deben desglosarse por género, situación en el mercado laboral, grupo etario, nivel educativo y grupos vulnerables (migrantes, minorías, personas con discapacidad y otros grupos desfavorecidos). El RDC y el Reglamento sobre el FSE+ para el período 2021-2027 establecen la obligación jurídica de que las AG recopilen y traten datos personales en forma de registros individuales de los participantes. Además, en relación con el seguimiento y los indicadores, el artículo 17, apartado 6, del Reglamento (UE) 2021/1057 sobre el FSE+ especifica lo siguiente: «Cuando se disponga de datos en registros o fuentes equivalentes, los Estados miembros podrán permitir a las autoridades de gestión y otros organismos encargados de la recogida de datos necesaria para el seguimiento y la evaluación del apoyo general del capítulo del FSE+ en régimen de gestión compartida que obtengan datos a partir de dichos registros o fuentes equivalentes, de conformidad con el artículo 6, apartado 1, letras c) y e), del Reglamento (UE) 2016/679».

*Reglamento sobre el FSE+ en relación con la gestión de datos*

<sup>45</sup> Un ejemplo es una evaluación del acuerdo de integración obligatorio entre el demandante de empleo y el asistente social. Mediante un ensayo de campo aleatorizado y sobre la base de las biografías del mercado laboral de las personas incluidas en dicho ensayo, el IAB pudo demostrar que, para algunos grupos de personas desempleadas, la regulación obligatoria resulta contraproducente y debería ser sustituida por una gestión más flexible del instrumento (Van den Berg *et al.* 2016).

<sup>46</sup> Véase [Ismeri Europa, Ecorys e Instituto de Estudios de Empleo, 2019](#).

Estas normas, establecidas en el RDC y el Reglamento sobre el FSE+, facilitan el acceso a los datos personales necesarios para el seguimiento del FSE+ y, en el caso de las EIC, para definir el grupo tratado, así como el uso de dichos datos. No obstante, el acceso a los datos personales necesarios para componer los grupos de control y el tratamiento de los datos del grupo de tratamiento y el grupo de control deben estar en consonancia con el Reglamento (UE) 2016/679 (Reglamento General de Protección de Datos o RGPD), que se refiere a la transferencia y el uso generales de datos personales, incluidas categorías especiales de datos<sup>47</sup>, dentro de la UE. En el recuadro siguiente se ofrece una visión general de los principales contenidos y obligaciones regulados por el RGPD.

### **Recuadro 8 Marco regulador de la UE sobre el tratamiento de datos personales**

La legislación pertinente sobre el tratamiento de datos personales en Europa consiste, principalmente, en el Reglamento (UE) 2016/679 y en las directrices y medidas adoptadas por las autoridades competentes, como el Comité Europeo de Protección de Datos (CEPD). Esta legislación establece numerosas condiciones y limitaciones para el tratamiento de datos personales, a fin de proteger los derechos y libertades de los interesados. En cualquier caso, la necesidad de protección y salvaguardia que se desprende de las disposiciones y los reglamentos debe encontrar un equilibrio con la necesidad de no obstaculizar la investigación científica y actuar como un activo en su desarrollo. Por esta razón, y dentro de este marco jurídico, la Unión Europea ha previsto que, en determinadas condiciones, puedan hacerse excepciones para permitir actividades de investigación y la difusión de sus productos, siempre que se salvaguarde el derecho primordial de las personas afectadas, es decir, el derecho a la intimidad<sup>48</sup>.

#### **REGLAMENTO (UE) 2016/679: REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS**

El Reglamento General de Protección de Datos (RGPD) entró en vigor el 24 de mayo de 2016 y entró plenamente en vigor en todos los Estados miembros el 25 de mayo de 2018. El RGPD se aplica «al tratamiento total o parcialmente automatizado de datos personales, así como al tratamiento no automatizado de datos personales contenidos o destinados a ser incluidos en un fichero»<sup>49</sup>.

Por lo que se refiere al ámbito territorial, el Reglamento se aplica tanto al tratamiento de datos personales efectuado por personas situadas en el territorio de la Unión Europea como en los casos en que el tratamiento implica a interesados situados en la Unión Europea, incluso cuando el responsable o el encargado del tratamiento se encuentre fuera de la UE.

Desde un punto de vista sustantivo, y en comparación con la legislación anterior, el Reglamento refuerza los derechos de los interesados e impone una serie de obligaciones a los responsables del tratamiento con arreglo a una lógica basada en el análisis de riesgos y en el principio de responsabilidad. Además, el Reglamento establece una serie de requisitos que el responsable del tratamiento debe cumplir y que no estaban previstos en la legislación anterior. Algunos de ellos son la evaluación del impacto sobre la intimidad (artículos 35 y 36), la minimización de las operaciones de tratamiento con arreglo a los criterios para la protección de la intimidad desde el diseño y por defecto (artículo 25), la adopción del registro de las actividades de tratamiento (artículo 30) y el nombramiento del delegado de protección de datos (artículos 37 a 39).

Es importante destacar los principios fundamentales establecidos en el artículo 5 del Reglamento, que deben cumplir quienes tratan datos personales. Merecen mención especial los siguientes:

<sup>47</sup> El RGPD ya no utiliza el término «datos sensibles», que ha sido sustituido por «categorías especiales de datos». Según el artículo 9, estas categorías incluyen los datos «que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona física».

<sup>48</sup> Entre la legislación europea pertinente que contribuye al marco jurídico para el tratamiento de los datos personales figuran también el Reglamento (UE) n.º 557/2013, relativo a la estadística europea, en lo que respecta al acceso a datos confidenciales con fines científicos; y el Reglamento (UE) 2018/1725, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por las instituciones, órganos y organismos de la Unión, y a la libre circulación de esos datos.

<sup>49</sup> Artículo 2 del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo.

- a) Licitud, lealtad y transparencia: los datos personales se tratarán de manera lícita, leal y transparente.
- b) Limitación de la finalidad: los datos se recogerán con fines determinados, explícitos y legítimos, y no serán tratados ulteriormente de manera incompatible con dichos fines. En el caso de que se traten con fines estadísticos o de investigación científica, el responsable del tratamiento adoptará las garantías y medidas de protección necesarias.
- c) Minimización de datos: solo deben recogerse los datos estrictamente necesarios para alcanzar fines específicos.
- d) Exactitud: los datos recabados deberán ser exactos y, si fuera necesario, actualizados.
- e) Limitación del plazo de conservación: los datos personales serán mantenidos de forma que se permita la identificación de los interesados durante no más tiempo del necesario para los fines del tratamiento de los datos personales. A tal fin, los datos personales podrán conservarse durante períodos más largos siempre que se traten exclusivamente con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos, de conformidad con el artículo 89. En tales casos, los datos estarán sujetos a la aplicación de las medidas técnicas y organizativas apropiadas, a fin de proteger los derechos y libertades de los interesados.
- f) Integridad y confidencialidad: los datos personales serán tratados de tal manera que se garantice una seguridad adecuada de los datos personales, incluida la protección contra el tratamiento no autorizado o ilícito y contra su pérdida, destrucción o daño accidental, mediante la aplicación de medidas técnicas u organizativas apropiadas.
- g) Responsabilidad proactiva: es uno de los principios más importantes contemplados en el Reglamento. Establece que responsable del tratamiento es responsable del tratamiento de los datos y debe ser capaz de demostrar el cumplimiento de las medidas exigidas.

Este conjunto de principios constituye la estructura principal en la que se fundamenta el RGPD y determina una serie de obligaciones que deben cumplir el responsable y el encargado del tratamiento.

## TRATAMIENTO DE DATOS PERSONALES EN EL CONTEXTO DE LAS EIC: CONDICIONES, LÍMITES Y PRINCIPALES PROBLEMAS

Para realizar una evaluación de impacto contrafactual es necesario tratar grandes cantidades de datos, incluidos datos personales. En relación con el marco regulador descrito anteriormente, es importante conocer las condiciones, las limitaciones y los principales problemas a los que se enfrentan los Estados miembros y las AG cuando llevan a cabo EIC.

### **Base jurídica y fines del tratamiento**

Para llevar a cabo el tratamiento de datos personales, debe concurrir al menos una de las condiciones de licitud indicadas en el artículo 6 del RGPD<sup>50</sup>. Si, además, los datos que deben tratarse pertenecen a «categorías especiales de datos», también debe hacerse referencia al artículo 9 del RGPD.

En el contexto de las EIC, el **interés público** mencionado en el artículo 6, apartado 1, letra e), del RGPD («el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento») parece la base jurídica más adecuada para el tratamiento de datos. Esta base jurídica debe establecerse en el Derecho nacional o de la Unión, tal como se especifica en el artículo 6, apartado 3, del RGPD. El «interés público» representa claramente las obligaciones de las autoridades de gestión definidas en el artículo 17, apartado 6, del Reglamento (UE) 2021/1057 sobre el FSE+, que se refiere al uso de los datos recogidos «en registros o fuentes equivalentes». Además, el RDC establece en su artículo 4 que «los Estados miembros y la Comisión solo estarán autorizados a tratar datos personales cuando sea necesario para cumplir sus obligaciones respectivas en virtud del presente Reglamento, en particular para el seguimiento, la presentación de informes, la comunicación, la publicación, la evaluación, [...]». Las legislaciones nacionales también pueden conferir a los responsables del tratamiento competencias similares para la gestión y el tratamiento de datos en aras del interés público.

<sup>50</sup> Esas condiciones son las siguientes: a) el interesado dio su consentimiento para el tratamiento de sus datos personales para uno o varios fines específicos; b) el tratamiento es necesario para la ejecución de un contrato en el que el interesado es parte o para la aplicación a petición de este de medidas precontractuales; c) el tratamiento es necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento; d) el tratamiento es necesario para proteger intereses vitales del interesado o de otra persona física; e) el tratamiento es necesario para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento; f) el tratamiento es necesario para la satisfacción de intereses legítimos perseguidos por el responsable del tratamiento o por un tercero, siempre que sobre dichos intereses no prevalezcan los intereses o los derechos y libertades fundamentales del interesado que requieran la protección de datos personales, en particular cuando el interesado sea un niño.

Se podrá hacer referencia a otros fundamentos jurídicos para una EIC cuando se den otras de las condiciones previstas en el artículo 6, apartado 1, y en particular:

- El **consentimiento** del interesado [artículo 6, apartado 1, letra a), del RGPD]. El consentimiento, por ejemplo, puede ser una base jurídica adecuada cuando los datos necesarios para realizar una EIC se recaben mediante una encuesta y los interesados puedan dar fácilmente su consentimiento al tratamiento (véase el artículo 7 del RGPD, relativo al consentimiento). En general, el consentimiento es más complejo si no se planifica con suficiente antelación; podría considerarse una base jurídica «residual» para una EIC cuando no sean aplicables otros fundamentos jurídicos.
- El tratamiento es necesario para el **cumplimiento de una obligación legal** aplicable al responsable del tratamiento [artículo 6, apartado 1, letra c), del RGPD en relación con el artículo 17 del Reglamento sobre el FSE+]. Esta base jurídica debe establecerse por ley y podría implicar a entidades privadas o públicas. Por ejemplo, podría ocurrir que, debido a prescripciones legales específicas, una entidad privada o pública responsable de un conjunto de datos esté obligada a colaborar con la autoridad de gestión en la EIC.

Además, el responsable del tratamiento **podría utilizar los datos recogidos con otros fines si estos son compatibles con el fin inicial**. A este respecto, el responsable del tratamiento tendrá que evaluar las condiciones establecidas en el artículo 6, apartado 4, y el artículo 5, apartado 1, letra b), del RGPD. Estas disposiciones son especialmente pertinentes para el tratamiento de datos con fines de investigación científica o estadísticos, pues permiten utilizar datos administrativos para fines que difieren de los originales y no exigen un consentimiento específico para el nuevo uso, aunque se deben cumplir las normas de protección especificadas en el artículo 89, apartado 1, y principalmente la seudoanonimización (véase más adelante). Estas disposiciones, por ejemplo, podrían ser pertinentes en el caso de los datos del registro de desempleo que vayan a utilizarse en una EIC.

Habida cuenta de la especificidad de las EIC, el responsable del tratamiento que recabe los datos y la entidad que lleve a cabo la investigación pueden ser diferentes. En este caso, debe existir una condición que legitime la transferencia de datos y permita al destinatario proceder a la EIC.

#### ***Anonimización y seudonimización: tratamiento con fines estadísticos***

El artículo 89 del RGPD establece que el tratamiento de datos realizado con fines de interés público, fines de investigación científica o fines estadísticos estará sujeto a las garantías adecuadas para los derechos y las libertades de los interesados y respetará, en particular, el principio de «minimización», lo que implica utilizar técnicas de seudonimización<sup>51</sup>.

Cuando los fines previstos puedan alcanzarse mediante operaciones de tratamiento posteriores que no permitan, o ya no permitan, la identificación de los interesados, esos fines se alcanzarán mediante la anonimización de los datos para después tratarlos de forma agregada. En las EIC, por definición, los resultados se agregan y este riesgo no existe, a menos que los conjuntos de datos originales se publiquen por razones científicas. En este caso, los conjuntos de datos deben anonimizarse (véase el ejemplo del recuadro siguiente).

#### ***Conservación y tratamiento seguro de los datos***

Uno de los principales aspectos del tratamiento de datos es su conservación. La legislación no especifica cómo deben conservarse los datos, pero los principios antes mencionados exigen que la conservación y el tratamiento siempre guarden relación con la finalidad de la investigación. Cuando se logre la finalidad del tratamiento y deje de ser necesaria la conservación de los datos, esta cesará. Esta norma general debe especificarse en la declaración de privacidad que se proporciona a los interesados cuando son registrados en conjuntos de datos administrativos o, dada la imposibilidad probable de informar personalmente a todos los interesados, pueden encontrarse formas alternativas de facilitar la información (por ejemplo, mediante la publicación de una página informativa que contenga la política de privacidad aplicable a las actividades de investigación en el sitio web de la autoridad de gestión).

En particular, se considera que la finalidad se ha logrado una vez que se ha llevado a cabo la EIC y sus resultados se han puesto a disposición y se han agregado con fines estadísticos. Llegados a este punto,

<sup>51</sup> La seudonimización consiste en el tratamiento de datos personales de manera tal que ya no puedan atribuirse a una persona concreta sin utilizar información adicional. La anonimización se refiere al tratamiento de los datos personales de tal manera que sea imposible identificar a personas con ellos. En el siguiente enlace se puede consultar un resumen de las técnicas de seudonimización: <https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices>.

los datos deben anonimizarse (en caso de que vayan a utilizarse de nuevo en una fase posterior y para fines diferentes) o suprimirse. Si la legislación o los reglamentos lo permiten, la conservación podría ser por tiempo ilimitado, pero solo si las normas aplicables lo mencionan explícitamente.

La otra condición que exige el RGPD es que el tratamiento esté sujeto a medidas de seguridad adecuadas y apropiadas. El artículo 32 del Reglamento establece que el responsable o el encargado del tratamiento aplicarán medidas técnicas y organizativas apropiadas para garantizar un nivel de seguridad acorde al riesgo. Esto significa que no se prescriben medidas estándar, sino que estas deben definirse caso por caso atendiendo específicamente a los riesgos que plantea la operación de tratamiento en cuestión.

## OBLIGACIONES JURÍDICAS

### **Información a los interesados**

La primera labor del responsable del tratamiento consiste en informar a los interesados. Esta obligación se establece en el artículo 13 del RGPD cuando los datos se obtengan del interesado; cuando no sea este el caso, debe tomarse como referencia el artículo 14. Se debe informar al interesado de los fines y métodos del tratamiento, la base jurídica, los plazos de conservación y los derechos que puede ejercer. En el caso de las EIC, esta información se proporciona en el momento de recoger los datos que se utilizarán posteriormente para la investigación estadística y de evaluación, como sucede en el seguimiento del FSE o en los conjuntos de datos del registro de desempleo. Cuando esta comunicación sea prácticamente imposible, como se ha mencionado anteriormente, y en especial en el caso de las personas que figuran en los registros utilizados para EIC, se puede recurrir a sitios web u otras herramientas generales de información.

### **Gobernanza de la relación entre las distintas entidades que intervienen en el tratamiento**

Por lo general, en las actividades de investigación interviene más de un organismo y, en tales casos, las relaciones entre ellos deben regularse mediante acuerdos específicos denominados «acuerdos de tratamiento de datos»<sup>52</sup>. Dichos acuerdos deben definirse caso por caso para reflejar las contribuciones de las distintas partes interesadas. Estos son algunos ejemplos de posibles relaciones:

- La relación *responsable del tratamiento - responsable del tratamiento* se produce cuando las entidades colaboran en la ejecución de un proyecto, aunque en condiciones distintas y con tareas diferentes, y cada una efectúa el tratamiento con sus propios fines.
- La relación *responsable del tratamiento - encargado del tratamiento* se produce cuando una entidad (el responsable del tratamiento) determina los medios y fines del tratamiento y recurre a otra entidad (el encargado del tratamiento) para llevar a cabo determinadas actividades de tratamiento. En este caso, los términos de referencia y las obligaciones figuran en el artículo 28 del RGPD.
- *Corresponsables del tratamiento*: este tipo de relación se contempla en el artículo 26 del RGPD, que establece que «cuando dos o más responsables determinen conjuntamente los objetivos y los medios del tratamiento serán considerados corresponsables del tratamiento». En tales casos, las partes deben determinar sus respectivas responsabilidades en el cumplimiento de la normativa de manera transparente, y en particular en lo que se refiere a los derechos de los interesados.

### **Evaluación del impacto relativa a la protección de datos**

El artículo 35 del RGPD indica que, cuando sea probable que el tratamiento de datos personales entrañe un alto riesgo para los derechos y libertades de los interesados, el responsable del tratamiento debe realizar una evaluación del impacto<sup>53</sup> antes de proceder al tratamiento. En las EIC, una evaluación del impacto puede ser importante solo en el caso de que se traten a gran escala categorías especiales de datos. En esos casos, el responsable del tratamiento de la administración podrá llevar a cabo la evaluación del impacto de conformidad con las normas y los instrumentos que haya puesto a su disposición la autoridad nacional de protección de datos.

## PERSPECTIVAS DE CARA AL FUTURO: LA LEY DE GOBERNANZA DE DATOS

El marco antes expuesto refleja la situación actual y las normas en vigor en la UE. El RGPD se introdujo específicamente para favorecer un avance decisivo en la protección de los datos personales y para crear

<sup>52</sup> Así sucede, por ejemplo, cuando el responsable del tratamiento no coincide con quienes están llevando a cabo la investigación. En ese caso, es necesario alcanzar un acuerdo que regule las relaciones entre las partes para acceder a los datos y realizar la investigación.

<sup>53</sup> Para obtener más información, consúltense las directrices publicadas por el CEPD: <https://ec.europa.eu/newsroom/article29/items/611236>.

un sistema común en toda la Unión, de acuerdo con la evolución tecnológica que se está produciendo y la sociedad de datos actual. Otras reformas están en fase de preparación y podrían tener un impacto significativo en la recogida de datos personales en los próximos años.

La Ley de Gobernanza de Datos se encuentra en estos momentos en la fase de propuesta en la Comisión Europea<sup>54</sup>. El objetivo explícito de esta ley es promover la disponibilidad de datos utilizables reforzando la confianza en los intermediarios de datos y mejorando los mecanismos de intercambio de datos en toda la UE. Es probable que la nueva norma gire en torno a los datos personales, y su uso con fines estadísticos y de investigación podría ser más amplio gracias a la introducción de una nueva figura: el intermediario para el intercambio de datos personales. Esta medida podría contribuir a crear un entorno más propicio para las EIC, lo que facilitaría el acceso a una gran cantidad de información a gran escala.

Resulta evidente que las normas del RGPD no excluyen la realización de una EIC. Por consiguiente, en todos los Estados miembros deben respetarse algunos pasos procedimentales y operativos básicos para que las EIC sean seguras. Por ejemplo, en un caso normal en el que los datos de seguimiento (personas tratadas) se combinan con datos de un registro público (grupo de control) y un evaluador externo lleva a cabo la EIC, estos son los pasos más importantes:

*Principales  
medidas para  
cumplir mejor el  
RGPD*

1. La AG llega a un acuerdo con las administraciones responsables de los datos (p. ej., registro de desempleo, registro fiscal, etc.) necesarios para especificar el grupo de control y analizar los grupos de tratamiento y de control. La AG verifica que el uso de estos datos cumple lo dispuesto en el artículo 6, apartado 1, letra e), del RGPD, o en el artículo 6, apartado 4, y el artículo 5, apartado 1, letra b), de dicho Reglamento, si no se ha obtenido el consentimiento expreso.
2. La AG celebra acuerdos con las demás entidades (los propietarios de los datos y los evaluadores) para regular los flujos de información y las responsabilidades mutuas de conformidad con las normas del RGPD. El contrato de servicios entre el evaluador y la AG debe contener una cláusula específica relativa a la protección de datos. En el caso del resto de administraciones, un memorando de entendimiento o procedimientos nacionales específicos pueden regular la protección de datos cuando se lleven a cabo EIC.
3. Sobre la base del acuerdo anterior, el propietario de los datos se los transfiere a la AG en forma pseudoanonimizada y, de conformidad con el RGPD, dicha autoridad transfiere datos al evaluador para tratar el grupo tratado y el grupo de control<sup>55</sup>.
4. La conservación de los datos personales de los grupos de tratamiento y de control solo cumple las normas relativas a la conservación de datos en el transcurso de la investigación y para los fines de esta, y de conformidad con las normas básicas de seguridad, tal como se recoge en el RGPD. Estas normas son respetadas por todas las entidades que participan en la EIC.

Las prácticas nacionales suelen estar en consonancia con el RGPD, pero difieren de unos Estados miembros a otros. Además, los evaluadores señalan

---

<sup>54</sup> La propuesta de una Ley de Gobernanza de Datos fue aprobada por la Comisión a finales de 2020. La situación de la Ley de Gobernanza de Datos puede consultarse en el siguiente enlace: <https://eur-lex.europa.eu/legal-content/ES/HIS/?uri=CELEX:52020PC0767>.

<sup>55</sup> Si los grupos de tratamiento y de control se analizan mediante una encuesta, esta debe obtener el consentimiento de las personas encuestadas para utilizar los datos con fines de investigación.

que las normas nacionales de protección de datos siguen planteando grandes obstáculos para el uso de microdatos. El motivo puede ser el tiempo que se requiere para adaptar las normas y los hábitos nacionales a la última versión del RGPD. También puede deberse a que las distintas administraciones nacionales interpretan dicho Reglamento de manera diferente, a la sobrerregulación, al endurecimiento del RGPD en algunos países o a otras interpretaciones erróneas. En algunos países, las AG o las autoridades nacionales deberían emprender una iniciativa específica para superar estos obstáculos en el espíritu del RGPD y del Reglamento sobre el FSE+. En el recuadro siguiente se explica cómo afrontó la región italiana del Véneto algunas de las dificultades más comunes. Una buena práctica consiste en ponerse en contacto con la autoridad nacional de protección de datos para discutir los acuerdos propuestos antes de celebrarlos.

### Recuadro 9 Protección e intercambio de datos

Una evaluación contrafactual del *Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations* [«Estudio piloto y de viabilidad sobre la sostenibilidad y la eficacia de los resultados para los participantes del Fondo Social Europeo mediante evaluaciones de impacto contrafactual», documento en inglés]<sup>56</sup> llevada a cabo por la DG EMPL en 2019 constituye un buen ejemplo de acceso a datos personales anonimizados en un plazo relativamente breve.

La evaluación, uno de los cuatro casos que contiene el estudio, se centró en las intervenciones del FSE (PI 9.i) dirigidas a los parados de larga duración que se llevaron a cabo en la región del Véneto en 2015-2016. Los datos de seguimiento sobre los participantes y los tipos de medidas se combinaron con datos sobre los períodos de empleo antes y después de las intervenciones del FSE (denominados «Comunicazioni obbligatorie»). El proceso duró unos dos meses (desde finales de julio hasta finales de septiembre) desde la primera reunión sobre los requisitos en materia de datos con la autoridad de gestión y la institución titular de los datos, el Instituto Regional de Trabajo del Véneto, llamado «Veneto Lavoro». Veneto Lavoro era responsable de la anonimización de los datos, y era fácil vincular todos los conjuntos de datos entre sí<sup>57</sup> a través de un identificador común único facilitado por este instituto.

Si bien las organizaciones implicadas no encargaron directamente la evaluación, la experiencia positiva se debe a dos factores principales: la cooperación estrecha (aunque informal) entre las oficinas regionales de la autoridad de gestión y el instituto regional Veneto Lavoro, necesaria para afrontar y resolver los problemas de protección de datos, y la existencia de una base de datos (Mercurio) gestionada por Veneto Lavoro a la que se puede acceder como archivo de uso público con fines de investigación<sup>58</sup>.

Mercurio es una base de datos estadística que contiene todos los datos registrados por las oficinas de intercambio de trabajadores de la región del Véneto respecto de los períodos de empleo y desempleo en la población. Además, contiene información detallada sobre todos los trabajadores y empresas registrados. En el momento del análisis, Mercurio contenía información sobre más de 4 millones de trabajadores, 17,6 millones de períodos de empleo y 4 millones de períodos de desempleo<sup>59</sup>.

El conjunto de datos «se actualiza y limpia» periódicamente, gracias a lo cual fue relativamente sencillo cotejarlo con los datos de seguimiento de la AG. Una posible limitación es que el archivo de uso público no se actualice con la suficiente frecuencia (por ejemplo, la última versión actualmente disponible se actualizó en diciembre de 2020), ya que se requiere un gran esfuerzo para garantizar la buena calidad de

<sup>56</sup> [Ismeri Europa, Ecorys e Instituto de Estudios de Empleo, 2019.](#)

<sup>57</sup> Se facilitó un conjunto de datos adicional donde figuraban las fechas de inicio y finalización de cada Declaración de Disponibilidad Inmediata para Trabajar firmada por todas las personas, tanto tratadas como sin tratar.

<sup>58</sup> En el siguiente enlace se puede consultar una descripción de la base de datos Mercurio y de las normas que regulan el acceso a ella: <https://www.venetolavoro.it/public-use-file>.

<sup>59</sup> Se extrajeron las siguientes variables de interés de los datos administrativos: las fechas de inicio y fin de cada período de empleo y desempleo; el tipo de contrato de cada período de empleo (permanente, temporal, prácticas); la duración prevista de los contratos temporales; información sociodemográfica (género, fecha de nacimiento, nivel educativo, ciudadanía).

los datos<sup>60</sup>. La experiencia indica que la AG debe tomar medidas con antelación, así como preparar planes de evaluación para establecer posibles formas de cooperación con los organismos externos que gestionan los datos necesarios para un ejercicio contrafactual; hallar soluciones a los posibles obstáculos jurídicos; y aclarar las maneras de acceder a datos administrativos de forma anonimizada, gestionarlos y suministrarlos.

### 2.3.2. ¿Cómo se puede especificar el grupo «tratado»?

Para llevar a cabo una EIC, es esencial definir claramente qué significa ser tratado o haber participado en la intervención. Además, una vez que se tenga una idea clara de cuándo se considera que una persona o empresa ha sido tratada, es importante poder identificarla. Aquí se presentan los principales problemas relacionados con la definición de los grupos de tratamiento y de control, y el capítulo 3 ofrece explicaciones más detalladas de sus características metodológicas.

A primera vista, podría parecer sencillo definir en qué consiste la participación. No obstante, hay una serie de cuestiones que podrían no resultar evidentes de inmediato, pero que son cruciales y requieren una profunda reflexión. Por ejemplo, ¿se considera que los participantes en un programa de formación que abandonan la intervención han sido tratados? ¿A cuántas sesiones de un curso de formación debe asistir una persona para ser considerada participante? También se deben tener en cuenta ciertos efectos previsibles. En previsión de ser objeto de una intervención, algunos solicitantes de prestaciones de la seguridad social podrían darse de baja del sistema de prestaciones sociales para evitar medidas de activación. ¿Estas personas son tratadas a pesar de que, por ejemplo, nunca asisten físicamente a las citas que se les conciertan en una oficina del SPE?

*Definición del grupo de tratamiento*

A la hora de definir el grupo de tratamiento, es preciso distinguir entre la «intención de tratar» y el «tratamiento de las personas tratadas». Desde una perspectiva política, la cuestión clave suele ser si el interés reside en los efectos de ofrecer la posibilidad de participar en una intervención o en los efectos de participar realmente. En el primer caso, las personas a las que se ofrezca una intervención podrían participar en ella o no. En el último, cuando el interés reside en el efecto del tratamiento para las personas tratadas, el grupo tratado solo consta de quienes efectivamente participan<sup>61</sup>.

*Intención de tratar o tratamiento de las personas tratadas*

En un principio, los responsables políticos suelen asumir que están interesados en determinar los efectos netos del tratamiento en los participantes. Sin embargo, pensándolo bien, puede que el asunto no esté tan claro. Si se puede identificar a las personas a las que se ofrece tratamiento, quizá sea más útil, desde una perspectiva política, definir las como el grupo «tratado». Así ocurre, en particular, cuando la participación en una intervención no es obligatoria. Los responsables políticos no pueden obligar a participar en una intervención a las personas a las que se la ofrecen, por lo que la pregunta que ha de hacerse es la siguiente: ¿cuál es el impacto

*Oferta de tratamiento o tratamiento real*

<sup>60</sup> Esto significa que es necesario trabajar para incorporar información más actualizada de la que figura en la última versión disponible de Mercurio. En concreto, hay que redoblar los esfuerzos para integrar y depurar la información más reciente sobre los períodos de empleo.

<sup>61</sup> Cuando la participación en una intervención es obligatoria, en esencia no existe ninguna diferencia entre estas dos situaciones: todas las personas a las que se ofrece un tratamiento deben participar en él. No obstante, en la mayoría de los casos las intervenciones no son obligatorias, y así se asume en esta Guía.

en el empleo y los ingresos posteriores de aquellas personas a las que se ofreció la oportunidad de participar en un programa de formación?

Para estimar los efectos de la oferta de tratamiento en una serie de resultados, es necesario poder identificar a quienes reciben dicha oferta. En muchos casos, esto puede resultar difícil.

### **Dónde encontrar datos adecuados**

Una vez que se haya definido quiénes son las personas tratadas y qué se entiende por tratamiento, es importante considerar cómo se identificará a las personas que reciben tratamiento a efectos de la evaluación. Esto significa invariablemente encontrar una fuente de datos a partir de la cual todas las unidades tratadas, ya sean personas o empresas, puedan ser contabilizadas o incluidas en la muestra. Estos registros suelen extraerse de los sistemas de seguimiento del FSE y, cuando los hay, de otros registros de datos establecidos para la intervención concreta.

*Encontrar fuentes de datos para las personas tratadas*

Debido a los requisitos de seguimiento e información del FSE (y FSE+), las organizaciones beneficiarias deben registrar cuántas personas reciben servicios a través de una intervención y algunas características personales sobre ellas. A efectos de la EIC, las intervenciones tendrán que ir un paso más allá y proporcionar microdatos sobre quienes hayan participado. Los evaluadores solicitarán un registro de cada unidad tratada (empresa o persona) con datos sobre sus principales características (sexo, edad, nivel educativo, etc.). Aunque estos datos pueden anonimizarse/seudonimizarse por razones de privacidad, cuando sea necesario realizar una encuesta para efectuar la EIC, habrá que determinar si las condiciones exigidas por el RGPD permiten transmitir las identidades de estas unidades (nombres, direcciones, números de teléfono, etc.) para que puedan ser incluidas en la muestra. También es preciso asignar un identificador único a cada unidad para facilitar la vinculación de los registros entre las distintas fuentes de datos.

### **2.3.3. Factores que deben tenerse en cuenta para especificar un grupo de control**

Para obtener una estimación de la hipótesis contrafactual, normalmente será necesario especificar un grupo de control. Por lo general, la elección del grupo de control se ve limitada por dos factores: si la intervención es obligatoria para los participantes o no, y si la intervención se ejecuta universalmente dentro de una jurisdicción o se circunscribe a una zona concreta o a un período limitado. La elección de un grupo de control adecuado tiene tres aspectos: 1) analítico; 2) político; y 3) práctico.

*Tres aspectos que deben tenerse en cuenta a la hora de definir un grupo de control*

#### **Definir un grupo de control desde una perspectiva analítica**

El objetivo de una EIC es obtener estimaciones imparciales de los impactos de una intervención en una serie de resultados. Para lograrlo, es preciso hacer estimaciones de los resultados contrafactuales. Las estimaciones de los resultados contrafactuales se obtienen de un grupo de control (véase la sección 1.1). Como se muestra en los gráficos 6 y 7, el impacto se calcula restando una estimación del resultado contrafactual de un resultado observado en el grupo «tratado». El nivel de sesgo de un impacto dependerá de hasta qué punto el resultado contrafactual computado para el grupo de control representa el resultado que se habría producido en el grupo tratado

de no haber sido tratado, si todos los demás factores hubiesen permanecido inalterados.

Para encontrar un grupo de control equivalente en ausencia de aleatorización, este debe ser equivalente al grupo de tratamiento en todos los aspectos importantes, tanto en las dimensiones observables como en las no observables. Esto es así en todos los enfoques cuasi experimentales y, por consiguiente, es una condición necesaria para todas las EIC que no adopten un enfoque aleatorizado.

Casi todas las intervenciones del FSE son a) voluntarias (el grupo destinatario no está obligado a participar en una intervención), o b) limitadas de algún otro modo. Si se trata de intervenciones piloto o instrumentos circunscritos a una región o jurisdicción concreta, los evaluadores contarán con un conjunto de unidades que podrían ser seleccionadas para conformar el grupo de control. Será necesario cribar en cierta medida este posible grupo para afinar la elección final de las unidades de control, de modo que tengan una buena correspondencia con los participantes (el grupo tratado). En muchos casos, puede haber cuatro opciones<sup>62</sup>:

*Opciones para seleccionar los grupos de control*

- **Ubicación:** las unidades de control que son similares a las que participan en una intervención, pero se encuentran en zonas del Estado miembro donde la intervención no está disponible (en caso de haberlas). La diferencia en diferencias es a menudo el enfoque preferido cuando existen esos grupos de control y se dispone de los datos correctos. Las poblaciones de distintos lugares pueden ser muy similares entre sí, y estos grupos no habrán tenido la oportunidad de participar en la intervención o de rechazar la invitación para hacerlo, de modo que no existirá esta importante fuente de posibles sesgos. Sin embargo, las condiciones del mercado laboral serán diferentes para las poblaciones de distintos lugares. Ante tales variaciones, las técnicas de diferencia en diferencias funcionan bastante bien, ya que las divergencias en las condiciones del mercado de trabajo local tienden a fijarse razonablemente con el paso del tiempo. No obstante, es menos aconsejable extraer muestras de control de diferentes mercados de trabajo locales en aquellos casos en que se recurra al emparejamiento para calcular los impactos. Se ha demostrado que el sesgo asociado a la selección de muestras de control de diferentes mercados laborales puede ser mayor que el sesgo de selección<sup>63</sup>.
- **Tiempo:** las unidades de control son similares a los participantes, pero se observan en diferentes momentos, ya sea antes o después de la intervención. Suele ser necesario seleccionar a los grupos de control de este modo cuando la intervención es universal y obligatoria, es decir, cuando todos los miembros de los grupos destinatarios están obligados a participar y el programa se aplica en toda una jurisdicción. Los grupos de control formados de esta manera presentan una desventaja considerable: sus resultados se medirán en momentos diferentes a los del grupo de tratamiento, por lo que pueden ser susceptibles a fluctuaciones cíclicas, alteraciones en la composición y cambios en las tendencias macroeconómicas que podrían minar la capacidad de observar un resultado contrafactual imparcial. Solo se debería valorar recurrir a estas

<sup>62</sup> Esta sección se fundamenta en [Card, D.; Ibarra, P. y Villa, J.M., 2011.](#)

<sup>63</sup> Véase [Heckman, J.J.; Ichimura, H.; Smith, J., y Todd, P., 1998.](#)

unidades de control cuando los resultados apenas varíen a lo largo del tiempo y cuando no se disponga de un grupo de control contemporáneo.

- **Elegibilidad:** en este caso, las unidades de control se seleccionan entre grupos que se encuentran en el mismo lugar y momento, pero entre los candidatos que, por una u otra razón, no pudieron ser elegidos para participar. Se suele recurrir a estas unidades de control cuando una intervención es universal, las tasas de participación son elevadas o la participación es obligatoria, y cuando existen normas claras de elegibilidad, de modo que, por ejemplo, las unidades que «simplemente no son elegibles» pueden usarse a modo de control. El objetivo es encontrar grupos que sean similares a los tratados, pero que, por razones bien conocidas y fijas (que pueden cuantificarse en datos), no fueron elegibles para el tratamiento. El acceso a intervenciones en el marco del FSE+ se basa con frecuencia en normas de elegibilidad claras que pueden medirse fácilmente y no son susceptibles de manipulación (edad específica de los participantes, duración de la situación de desempleo, etc.). Por lo tanto, la selección de las unidades de control puede utilizar estos umbrales como factores discriminatorios para distribuir a las unidades tratadas y de control.
- **Decisión/conocimiento:** en esencia, tanto el grupo de tratamiento como el grupo de control (en lugar de solo el grupo de tratamiento) están sujetos a procesos de selección basados en decisiones motivadas por factores que podrían no observarse<sup>64</sup>. Se puede seleccionar a las unidades de control entre aquellas que eran elegibles, pero finalmente no participaron en la intervención. La ventaja de este método es que las unidades de control suelen pertenecer al mismo mercado laboral que las tratadas. Hay que analizar cuidadosamente estas unidades de control, sobre todo cuando la EIC se diseñe mediante el emparejamiento y cuando pueda utilizarse una gran cantidad de datos para fundamentar la decisión de selección. En otras circunstancias, por ejemplo, cuando se aplique el enfoque de la diferencia en diferencias, las unidades de control de decisión/conocimiento serán menos atractivas.

Hay otro aspecto que merece atención especial. Cuando se disponga de mediciones de los resultados previas al tratamiento en relación tanto con el grupo de tratamiento como con el grupo de control, es importante examinar las tendencias de las mediciones de los resultados para ambos grupos antes de la intervención. Al comprobar la hipótesis de las «tendencias comunes», se afronta el problema de la caída temporal de las tasas de empleo y los ingresos antes de la intervención que se habrá producido para algunas de las personas que pueden acogerse a las políticas activas del mercado de trabajo (de lo contrario no serían elegibles para recibir ayuda, un problema que se conoce con el nombre de «caída de Ashenfelter»). El evaluador busca tendencias temporales similares en las mediciones de los resultados del grupo de tratamiento y el grupo de control, de modo que la recuperación de la pérdida a corto plazo del empleo o los ingresos no se confunda con las mejoras relativas a largo plazo que se pretenden detectar con una EIC.

La selección de grupos de control adecuados es un ejercicio complejo desde el punto de vista técnico y metodológico. Durante el desarrollo de los sistemas

*Analizar las  
tendencias previas  
a la intervención*

<sup>64</sup> [Card, D.; Ibararan, P., y Villa, J.M., 2011](#) se refieren a esto como «sesgo de selección doble».

de evaluación, se recomienda que los funcionarios se familiaricen con los principales conceptos y tomen medidas tempranas para especificar posibles unidades de control. Es importante que quienes encarguen una evaluación involucren a expertos desde el inicio de la fase de diseño, a fin de que proporcionen apoyo y asesoramiento.

### ¿Cuáles son las consideraciones políticas de interés?

La selección de un grupo de control adecuado no es un proceso meramente técnico o analítico. Aunque para especificar las unidades de control adecuadas es fundamental tener en cuenta aspectos analíticos, también es importante que un grupo de control represente una alternativa pertinentes a la intervención cuando esta se considere desde la perspectiva de la formulación de políticas.

*Definir una alternativa a una intervención*

Las EIC pueden adoptar múltiples formas. Por ejemplo, pueden comparar los resultados de uno o varios grupos de tratamiento con un grupo de control que no recibe tratamiento, o pueden comparar un grupo de tratamiento con otro sin un grupo de control no tratado. En la elección del grupo de control influirá qué tipo de comparación es más pertinente desde el punto de vista político, y si es posible encontrar un grupo de control «sin tratamiento». El recuadro 10 presenta un ejemplo de una comparación entre dos tratamientos, sin un grupo de control no tratado. El objetivo es determinar con qué intervención conviene seguir adelante. También cabe señalar que la comparación entre un programa y otro, sin la ventaja que supone contar con un grupo de control no tratado, puede dar lugar a ambigüedades (esto se analiza en más detalle en el recuadro 11).

*Comparar un tratamiento con la ausencia de tratamiento o con una alternativa*

#### Recuadro 10 Preguntas políticas relacionadas con un programa de formación

Pongamos por ejemplo que un responsable de la formulación de políticas se propone introducir una nueva intervención en materia de formación que se financiará a través del FSE. Llamaremos a esta intervención la «intervención A». Además, supongamos que el Estado miembro ya cuenta con un programa de formación (la intervención B) dirigido a las mismas personas, pero financiado con fondos nacionales. En tal caso, podría plantearse la siguiente pregunta política: ¿los niveles de empleo y de ingresos de los participantes en la intervención A son superiores a los de los participantes en la intervención B después de la participación? Y, por extensión, ¿ofrece la intervención A una mejor relación calidad-precio? Si los ingresos de los participantes en la intervención A son mayores, la respuesta política evidente es interrumpir la intervención B en favor de la intervención A, si su realización también resulta rentable.

#### Recuadro 11 Interpretar los efectos netos

Un estudio podría no encontrar ninguna diferencia entre los ingresos de los participantes en la intervención A y los participantes en la intervención B. Es posible que la respuesta política a esta información no esté clara si, por ejemplo, la intervención B fue muy eficaz en comparación con las unidades que no recibieron ningún tratamiento. Esto significaría que ambas intervenciones son muy eficaces. Sin embargo, en algunos casos puede ocurrir que no haya pruebas de la eficacia de la intervención B en relación con la ausencia de tratamiento. Otra posibilidad es que ni la intervención A ni la intervención B sean eficaces, aunque una intervención podría parecer relativamente más eficaz que la otra. En circunstancias en las que determinados grupos de la población puedan ser los destinatarios de más de una intervención, aún puede ser más informativo encontrar un grupo adecuado de unidades no tratadas que sirva de comparación.

Téngase en cuenta que la diferencia en diferencias no puede utilizarse para comparar varios tratamientos en ausencia de un grupo de control no tratado.

### **¿Qué consideraciones prácticas deben tenerse en cuenta para seleccionar un grupo de control?**

Además de consideraciones analíticas y políticas, es preciso tener en cuenta los aspectos prácticos de la selección de los grupos de control. Para seleccionar o incluir en la muestra las unidades (personas o empresas) que servirán de control, resulta necesario encontrar un marco de muestreo adecuado. Asimismo, los marcos de muestreo deben contener unidades que se ajusten a los requisitos analíticos y políticos. Cuál es exactamente la mejor manera de hacerlo variará de una evaluación a otra, en función del contexto específico de la intervención que se esté comprobando.

Muchas veces, se suelen aprovechar dos fuentes de datos para especificar grupos de control adecuados. En tales casos, ambas fuentes deben conocer la identidad del grupo de tratamiento.

Los censos de población de diversos tipos pueden utilizarse para encontrar unidades de control. Por ejemplo, si una intervención activa en el mercado de trabajo se dirige a las personas de entre dieciocho y veinticuatro años que reciben prestaciones por desempleo, sus registros de prestaciones pueden utilizarse para definir la población destinataria. Además, si se sabe cuál es el grupo tratado y se puede asociar a sus integrantes con los datos sobre las prestaciones, se podrá encontrar a personas de entre dieciocho y veinticuatro años que no reciben tratamiento y, por tanto, podrían servir de control. De manera alternativa, supongamos que una intervención está dirigida a pequeñas y medianas empresas. Los registros nacionales de empresas, de haberlos, podrían usarse para definir la población destinataria, y los posibles grupos de control podrían encontrarse con la información disponible sobre qué empresas están siendo tratadas.

*Censos de población y registros fiscales de empresas*

Los registros de los solicitantes pueden utilizarse cuando la admisión en una intervención no sea universal; por ejemplo, cuando no todos los candidatos a un programa de formación sean aceptados (grupo de control de decisión/conocimiento). Del mismo modo, no todas las empresas que solicitan financiación la obtendrán. En algunos casos, las unidades que no sean aceptadas para recibir formación o financiación pueden servir de control (véase la explicación que se hizo anteriormente en esta misma sección sobre la prudencia que se debe tener al seleccionar grupos de control en estas circunstancias).

*Registros de solicitantes*

### **2.3.4. ¿Qué tipos de cuestiones relativas a los datos deben plantearse en el sistema de evaluación?**

#### **¿Qué tipos de datos se requieren y cómo se recogerán?**

Como se ha observado, para efectuar una EIC suele ser necesario acceder a cantidades considerables de microdatos (en algunos casos podrían

utilizarse datos agrupados, p. ej., datos regionales). Estos datos deben recopilarse, cotejarse y documentarse. Es preciso vincular entre sí los datos procedentes de diversas fuentes a partir de campos de identificación comunes. Los datos deben conservarse y transferirse de forma segura entre los responsables de gestionar y llevar a cabo la EIC. Además, deben elaborarse conjuntos de datos analíticos a partir de estas fuentes de datos para facilitar la estimación de los impactos.

*Gestionar las fuentes de datos*

Cuando se prepare un sistema de evaluación, es importante hacerse las siguientes preguntas en relación con los datos:

*Una lista de control básica para la gestión de datos*

#### *Disponibilidad*

- ¿Qué fuentes pueden utilizarse para obtener diversos tipos de datos agrupados o microdatos?
- ¿Se dispone de los datos individuales necesarios? ¿Esto se aplica también a los datos de «categorías especiales» (si procede)?
- Cuando sea necesario realizar una encuesta, ¿se especifica al grupo destinatario y el grupo de control de una manera que permita hacer un seguimiento mediante entrevistas? ¿Se dispone de sus datos de contacto y estos están actualizados?

#### *Coherencia*

- ¿Existe una única fuente de datos o es necesario vincular varias (p. ej., estadísticas sobre desempleo, prestaciones sociales, seguridad social, datos sobre empresas o la constitución de sociedades, etc.)?
- ¿Son coherentes las fuentes entre sí? ¿Es posible identificar a las personas que se encuentran en distintas fuentes de datos de forma coherente?

#### *Accesibilidad*

- ¿Es posible acceder a fuentes de datos nacionales sobre las trayectorias profesionales, los ingresos o las prestaciones sociales de cada persona para comparar a los participantes de programas financiados con cargo al FSE con un posible grupo de control?

#### *Acuerdos*

- ¿Hay en vigor acuerdos específicos en materia de disponibilidad de datos entre la AG y otros propietarios de datos? ¿Qué obstáculos jurídicos u organizativos deben debatirse?
- ¿Quién se encargará de negociar el acceso y llegar a un acuerdo para el uso de los datos?

#### *Almacenamiento*

- ¿Dónde se conservarán los datos? ¿Qué sistemas e infraestructuras informáticas se necesitarán?
- ¿Qué medidas se adoptarán para garantizar que los datos se conserven de forma segura y que el acceso esté reservado a quienes los necesiten para fines de evaluación?
- ¿Cómo se anonimizan los datos? ¿Es posible seguir a las personas a lo largo del tiempo y a través de las fuentes de datos vinculadas?

#### *Cumplimiento de la protección de datos*

- ¿Cumplen las soluciones anteriores las normas básicas del RGPD y del Reglamento sobre el FSE+? ¿El uso de diferentes bases jurídicas de

conformidad con el artículo 6, letras a), c) y e), del RGPD y el artículo 17 del Reglamento sobre el FSE+ y el posible uso de datos anonimizados justifican los obstáculos condicionales para la accesibilidad de los datos?

### ¿Cómo se tratarán los datos?

En muchos casos, las EIC requerirán microdatos, es decir, datos que contengan observaciones sobre unidades individuales (normalmente personas o empresas individuales) del grupo de tratamiento y el grupo de control (en ocasiones podrían utilizarse datos agrupados, p. ej., datos regionales o datos de las oficinas del SPE). Los tres tipos de datos principales necesarios son: a) registros del grupo de tratamiento y el grupo de control, b) registros de los resultados, y c) datos contextuales (datos utilizados para comprobar si podrían existir diferencias importantes entre los grupos de tratamiento y de control). Estos datos podrían proceder de una o varias fuentes. Las fuentes deben estructurarse para formar conjuntos de datos analíticos (o muestras analíticas) que se utilicen para estimar los impactos. En muchos casos, esta estructuración implicará vincular los registros individuales de personas o empresas de distintas fuentes. Para vincularlos, se requieren identificadores a nivel individual (por ejemplo, números individuales de identificación de la seguridad social) que permitan alinear el registro de una persona (por ejemplo, de datos fiscales) con los registros de participación, o suficientes datos para vincular los registros de distintas fuentes (por ejemplo, el nombre y la fecha de nacimiento deben figurar en todas las fuentes). Es importante sopesar qué fuentes de datos se utilizarán para la EIC que se esté planificando, pero también si será posible vincular los registros de distintas fuentes.

*Vincular los microdatos entre distintas fuentes*

### 2.3.5. ¿Cuáles son las principales limitaciones para analizar los datos y los resultados?

Como se ha señalado anteriormente, las EIC suelen determinar los impactos comparando los resultados del grupo de tratamiento con los del grupo de control, y esta comparación fundamental forma, en esencia, parte del enfoque de la EIC. La diferencia entre los grupos de tratamiento y de control es el impacto o efecto neto de la intervención. No obstante, la manera exacta en que se calcularán los impactos dependerá de cómo se haya diseñado la investigación.

Cuando se planifique una EIC, es importante valorar si la intervención es lo suficientemente grande y puede generar impactos estadísticamente detectables. De hecho, los métodos de EIC se basan en el principio de que el grupo destinatario y el grupo de control son una muestra de todas las poblaciones de tratamiento y de control potenciales y, para producir resultados apreciables, el número de unidades contenidas en los dos grupos debe ser estadísticamente significativo.

Cuando se valore si una muestra tiene un tamaño suficiente para el análisis, resulta útil el concepto de «efecto mínimo detectable»<sup>65</sup>. En pocas palabras, un efecto mínimo detectable es el menor impacto real que puede detectar

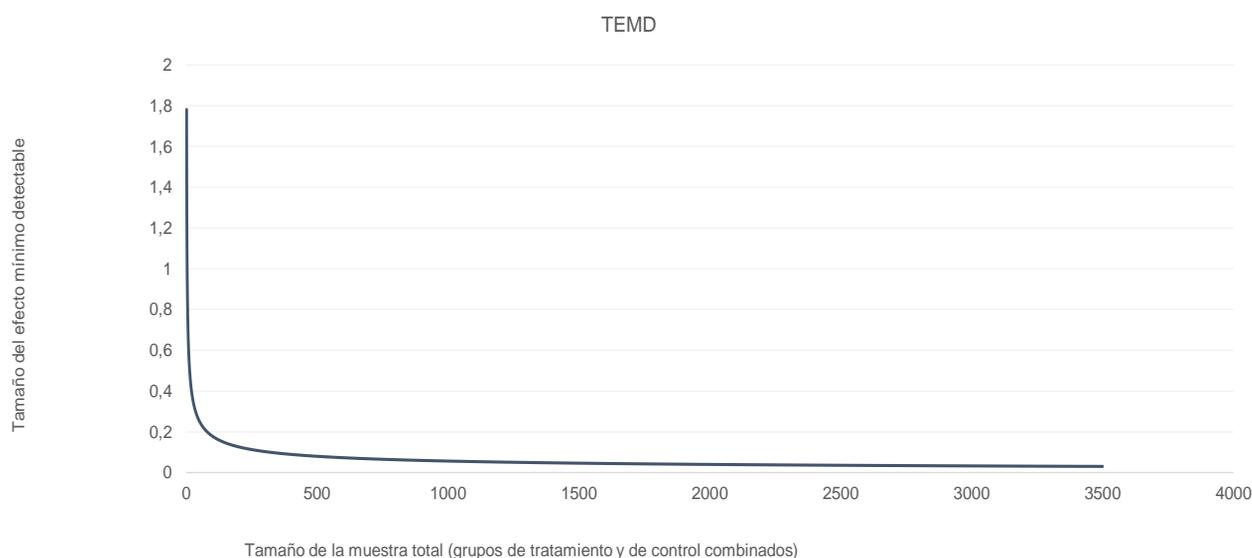
*Evaluar el tamaño de la muestra y del efecto*

<sup>65</sup> Bloom, H. S., 1995, proporciona orientaciones prácticas sobre cómo calcular los efectos mínimos detectables de diseños experimentales. En el caso de los enfoques cuasi experimentales, habrá que ajustar dichos cálculos. En términos generales, los enfoques cuasi experimentales requieren unas muestras de mayor tamaño del que se necesita para un diseño experimental.

una muestra de este tamaño con niveles estándar de confianza y potencia estadísticas<sup>66</sup>.

En el gráfico 4 se muestra cómo el tamaño del efecto mínimo detectable varía con el tamaño total de la muestra (los números totales de las muestras en los grupos de tratamiento y de control) a un nivel de confianza estadística del 95 % y a un nivel de potencia estadística del 80 %. De izquierda a derecha, el tamaño del efecto mínimo detectable disminuye rápidamente a medida que el tamaño de la muestra se aproxima a 500 (250 unidades de tratamiento y 250 unidades de control). Dicho de otro modo, a medida que aumenta el tamaño total de la muestra, el diseño de la EIC es más preciso y capaz de detectar impactos menores.

**Gráfico 4 Tamaños de los efectos mínimos detectables con muestras de distintos tamaños**



Nota: 95 % estadísticamente significativo y 80 % de potencia estadística. En el gráfico se asume que el diseño es aleatorizado.

Cuando se planifica una EIC, resulta útil estimar el tamaño probable de las muestras para definir el grupo tratado y el grupo de control. Esta estimación se basa en previsiones del número de unidades que serán tratadas, el diseño del método probable que se utilizará en la EIC y el tamaño de los correspondientes grupos de control derivados de los datos administrativos disponibles o de una encuesta adecuada. A continuación, esta información puede utilizarse, bajo determinadas hipótesis, para verificar si los efectos mínimos detectables resultantes son suficientemente significativos y si la EIC es una opción razonable. Un aspecto clave de este ejercicio es si es probable que la intervención en cuestión genere efectos de un tamaño equivalente a los efectos mínimos detectables estimados. A este respecto, conviene examinar la bibliografía existente y otras evaluaciones similares.

*El tamaño probable de la muestra*

<sup>66</sup> El nivel de confianza estadística es una medida estadística de la fiabilidad del procedimiento de estimación, mientras que el nivel de potencia estadística es la probabilidad de que una prueba detecte un efecto real. Estos dos parámetros dependen también del tamaño de la muestra y, a la inversa, pueden utilizarse para calcular el tamaño de una muestra con respecto a unos niveles adecuados de importancia. En general, la confianza estadística es del 95 % y la potencia estadística, del 80 %.

También cabe señalar que las EIC tienen a menudo la intención de examinar los efectos en las distintas subpoblaciones que participan en la intervención (hombres/mujeres, jóvenes/mayores, mayor/menor nivel educativo, etc.). Estos análisis reducen forzosamente el tamaño del grupo tratado y el grupo de control para cada subpoblación (hombres, mujeres, jóvenes, etc.) y, por consiguiente, disminuye la significación estadística. Cuando el estudio de los efectos en las subpoblaciones sea un componente importante de la evaluación, es necesario contar con un número suficiente de personas tratadas y de control en cada subpoblación.

Para definir el tamaño de la muestra se requieren unas capacidades técnicas que la AG no siempre posee. En este caso, el evaluador debe calcular el tamaño de la muestra y acordar con la AG la viabilidad y el alcance de la EIC. No es posible definir un umbral para el tamaño de la muestra por encima del cual siempre es posible realizar una EIC, ya que dicho umbral puede variar en función de los métodos adoptados, los análisis de las subpoblaciones previstos, los niveles de significación estadística aceptados y otros elementos. Sin embargo, en muchas evaluaciones se ha demostrado que un total de 2 000 personas para el grupo tratado y el grupo de control combinados basta para realizar un análisis razonablemente preciso (incluidos los principales subgrupos, como hombres/mujeres, etc.). Esta indicación no puede utilizarse como «umbral» científico, sino simplemente como ayuda para examinar la posible realización de una EIC.

El tamaño de la muestra afecta a la solidez de los resultados de la EIC. En el siguiente recuadro se explican algunos ejemplos de incertidumbres en la interpretación de los resultados y su relación con el tamaño de la muestra.

### **Recuadro 12 Incertidumbres en la interpretación de los resultados**

Entre los ejemplos de evaluaciones de intervenciones financiadas por el FSE, la evaluación de varias medidas dirigidas a los parados de larga duración en la región italiana de Las Marcas no arrojó resultados estadísticamente significativos en una de las cuatro medidas analizadas, los «bonos para formación». Probablemente esto se deba al hecho de que la muestra de personas tratadas objeto de análisis era pequeña en comparación con el gran número de beneficiarios reales. Algo similar ocurrió en la evaluación de las medidas para promover la formación profesional en la región de Podlaskie (Polonia). Aparte de la cuestión de la significación estadística, este ejemplo muestra que, al analizar los efectos, la dependencia de muestras demasiado pequeñas de los grupos de tratamiento puede dar lugar a problemas de generalización en las conclusiones.

La evaluación de la formación lingüística para inmigrantes en Alemania arroja resultados positivos, a diferencia de algunos estudios anteriores que evaluaron los impactos de programas en la materia. Una posible explicación de esta diferencia, según el autor, es que el programa alemán analizado también contenía un módulo de experiencia laboral. El autor no pudo comprobar esta hipótesis con los datos disponibles, debido a la imposibilidad de distinguir entre los componentes de formación lingüística y experiencia laboral de la intervención. No obstante, podría ser un aspecto interesante para futuras investigaciones. El ejemplo demuestra que, en ocasiones, los ejercicios contrafactuales deben repetirse con datos mejorados y más variados o complementarse con otros enfoques de evaluación, a fin de explicar los mecanismos que determinan los resultados de una intervención.

### 2.3.6. Una lista de comprobación para verificar la preparación y viabilidad de una EIC

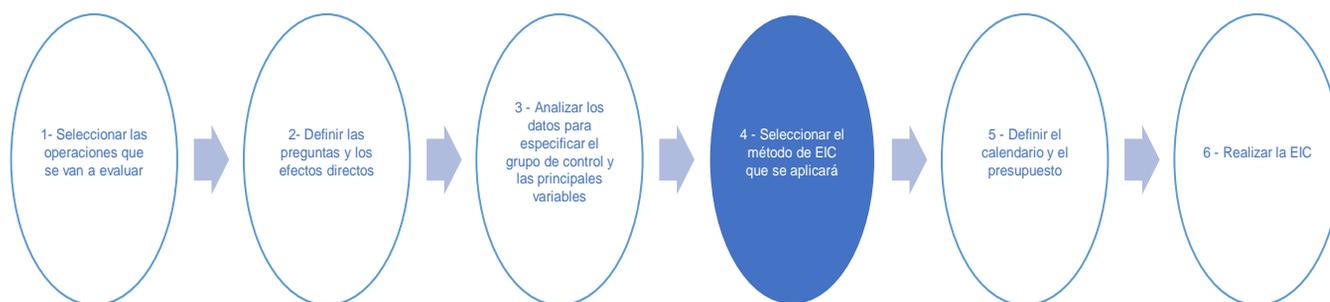
Llegados a este punto, resulta útil resumir en una lista de comprobación básica los principales factores que una AG debe sopesar y verificar cuando prepare una EIC:

*Una lista de los factores clave que deben tenerse en cuenta*

- ¿La intervención seleccionada para la evaluación es adecuada para una EIC?
  - o ¿Es la intervención específica, distintiva y suficientemente homogénea?
  - o ¿Sugiere la «teoría del cambio» de la intervención un mecanismo causal convincente de los efectos directos que se examinarán con una EIC?
  - o ¿Resulta fácil especificar un grupo de control de conformidad con las normas de la intervención?
- ¿Se pueden medir de forma cuantitativa los efectos sobre los participantes?
  - o ¿Los efectos directos de la intervención medidos son claros y coherentes con la teoría del cambio de la intervención?
  - o ¿Ha transcurrido tiempo suficiente desde el final del tratamiento para detectar los efectos directos en los participantes?
- ¿Las preguntas de la evaluación requieren una medición del impacto (los efectos netos) de la intervención?
- ¿Se dispone o puede disponer de datos adecuados (registros de los grupos de tratamiento y de control, registros de los efectos directos, datos contextuales)?
  - o ¿Se ajustan los datos disponibles al RGPD? En caso negativo, ¿pueden activarse los acuerdos interadministrativos y las soluciones técnicas necesarios para que los datos cumplan dicho Reglamento?
  - o ¿Es posible definir un grupo de control adecuado con los datos disponibles?
  - o ¿El tamaño de la muestra es lo suficientemente grande como para que los resultados de la EIC tengan la significación estadística necesaria?

Si se responde de manera afirmativa a todas estas preguntas, significa que es posible diseñar y llevar a cabo una EIC. Dar una respuesta negativa en algunas de ellas no descarta su realización, pero será necesario hacer comprobaciones adicionales de las condiciones de viabilidad y mejorar la disponibilidad de datos u otros factores clave.

## 2.4. Método de EIC que se aplicará



La selección del método que se utilizará en la EIC es un paso crucial en la preparación de la evaluación, ya que el método seleccionado afecta tanto a la calidad de los resultados como a la calidad global. En general, ningún método es superior a otro, pero uno puede ser más adecuado para aprovechar los datos disponibles, mientras que otro puede ser más adecuado para investigar un tipo concreto de intervención; determinar cuál es el más adecuado para una evaluación específica es un paso crucial en el diseño.

*Algunas condiciones fundamentales para determinar cuál es el método más adecuado*

Los aspectos técnicos y las características intrínsecas de cada método se examinan detalladamente en el **capítulo 3**, y en esta sección resulta útil reflexionar brevemente sobre las implicaciones prácticas de la selección.

En primer lugar, antes del inicio de la intervención debe establecerse un enfoque aleatorizado, que asigne aleatoriamente personas al grupo de tratamiento o de control. Esto asume tanto que el diseño de la evaluación se finalizará como que habrá una organización capaz de abordar la tarea. Aunque, como veremos, en las intervenciones del FSE rara vez se ha adoptado este enfoque, es posible en determinadas condiciones y muy eficaz desde el punto de vista metodológico.

En segundo lugar, la selección del método de EIC puede sobrepasar las capacidades de la AG, ya que requiere poseer unas capacidades técnicas avanzadas y experiencia específica. En tal caso, el evaluador seleccionará el método más adecuado, pero, evidentemente, solo después de haber sido designado. Esto puede tener la ventaja añadida de actuar como criterio para evaluar la capacidad de los evaluadores en el procedimiento que se siga para su selección. En cualquier caso, la AG debe confiar en que se podrá llevar a cabo una EIC tras haber respondido afirmativamente a todas o la mayoría de las preguntas de la lista de comprobación anterior y haber adquirido, ya sea interna o externamente, las capacidades necesarias para aprobar la elección del evaluador.

En tercer lugar, los métodos aleatorizados pueden diferir de los métodos cuasi experimentales en términos de duración y costes, y, en los ensayos de control aleatorizados, el evaluador debe participar desde el inicio de la intervención. Los métodos cuasi experimentales no difieren mucho entre sí. El evaluador podría empezar a trabajar una vez iniciada la intervención (pero con tiempo suficiente para organizar y preparar los datos necesarios), y la calidad de los datos disponibles podría suponer la diferencia en términos de tiempo y costes.

Por último, en ocasiones el tipo de intervención y la disponibilidad de datos permiten utilizar más de un método cuasi experimental. En este caso, puede

resultar útil aplicar más métodos, dado que confirmar los mismos efectos con distintos métodos mejora la solidez de los resultados.

## 2.5. Calendario y presupuesto



### 2.5.1. ¿De qué recursos se dispone?

La disponibilidad de recursos es una cuestión clave que debe tenerse en cuenta a la hora de diseñar el sistema de evaluación de una EIC. Estas consideraciones pueden referirse a numerosos aspectos, que aquí se agrupan en tres apartados: a) recursos de expertos, b) tiempo y c) recursos financieros.

#### ¿Qué expertos externos y personal interno se necesitan para realizar una EIC?

En la mayoría de los casos, la una evaluación de impacto se encargará a un proveedor externo. No obstante, el contrato deberá ser gestionado en el seno de la AG por personal que conozca los métodos de EIC. Estos conocimientos son necesarios para garantizar la calidad y establecer un contacto efectivo con expertos externos. La AG también podría tener que poseer otros conocimientos especializados, como conocimientos estadísticos y experiencia en la recopilación y gestión de datos. Es importante valorar de antemano si la AG tiene acceso a personal debidamente cualificado y formado, y si este personal tiene la capacidad de apoyar la evaluación.

*Personal interno*

Para encargar una EIC efectiva, es preciso recurrir a contratistas que posean las capacidades y la experiencia necesarias para llevarla a cabo. Además, los contratistas adecuados deberán comprender el contexto político y administrativo del Estado miembro en cuestión, estar familiarizados con las posibles fuentes de datos y dominar las lenguas oportunas. Conviene examinar qué se necesita para desarrollar una base de proveedores de EIC en un Estado miembro (en el capítulo 4 se analiza este asunto en más profundidad).

*Personal externo*

Para que una EIC sea efectiva, quienes gestionan el programa o la intervención objeto de evaluación deben cooperar entre sí. Por ejemplo, será necesario acceder a los registros mantenidos por los responsables de la intervención. Estos registros proporcionan información sobre las personas o empresas que participaron en una intervención.

*Personal que gestiona programas o intervenciones*

Los responsables de programas o intervenciones pueden proporcionar asesoramiento y orientación sobre estos tipos de datos. También se les

puede exigir que lleven un registro, además de lo que tendrían que hacer en ausencia de una evaluación del impacto.

Para resolver el problema de la recopilación de datos de diversas fuentes, quienes planifiquen una EIC tendrán que ponerse en contacto con el personal que se ocupa de las fuentes de datos oficiales (p. ej., registro de desempleo, datos de la seguridad social, oficinas estadísticas, etc.). De este modo, el suministro de datos se podrá planificar con mucha antelación.

*Conocimientos estadísticos*

### **¿Qué factores son pertinentes para programar una EIC?**

Para realizar una EIC se requieren las contribuciones de distintos recursos humanos. Además, estas evaluaciones se prolongan durante períodos considerables. Un sistema de evaluación debe incluir un calendario general con los hitos esenciales del proyecto, el cual será aplicable a quienes estén involucrados en la propia intervención y a los asociados a la evaluación. El calendario general tendrá que organizarse para todas las actividades de evaluación y de ejecución de la intervención, además de para los principales hitos políticos.

*Las EIC llevan tiempo*

Para crear un calendario general para una EIC que sea lógico y realista, es preciso encontrar un difícil equilibrio. Por una parte, la autoridad de gestión (o el OI) que planifica la evaluación debe tener en cuenta las fechas clave en las que se habrán tenido que tomar las decisiones que dependen de las conclusiones de la evaluación. Por otra parte, habrá limitaciones inevitables que afectarán al calendario de los informes. Algunos resultados tardarán años en aparecer, y los calendarios de recopilación de datos, análisis y presentación de información tendrán que reflejarlo en la medida de lo posible (véase la sección 2.5.2).

*Planificar un desfase que permita la aparición de impactos*

La evaluación debe llevarse a cabo con la suficiente antelación durante el período de programación para permitir que se introduzcan cambios y, en el tiempo restante, se saque el máximo provecho a las experiencias y lecciones aprendidas. En algunas circunstancias, las mismas intervenciones u otras similares podrían recibir apoyo en períodos de programación sucesivos. Los resultados de las EIC centradas en intervenciones de períodos de programación anteriores pueden ser sumamente útiles para orientar la ejecución y el diseño en períodos de programación posteriores.

*Centrarse en momentos concretos*

También es importante tomar en consideración el modo en que el calendario de una evaluación contrafactual puede estar vinculado al calendario de otros componentes de evaluación. Es probable que, antes de realizar una EIC, haya que efectuar una evaluación basada en la teoría. En el caso de las intervenciones innovadoras, es posible que se deba informar de los elementos esenciales de la evaluación del proceso bastante antes, a fin de favorecer la mejora del diseño de la EIC, o en un momento posterior para facilitar la investigación pormenorizada de sus causas y efectos. Tal es el caso, por ejemplo, de las intervenciones del FSE en 2014-2020 en apoyo de la infancia en la República Checa, los servicios para impulsar la inclusión social de las personas que reciben la nueva ayuda a la renta en Italia, o las diversas medidas de apoyo a las personas que teletrabajaron o estudiaron a distancia durante el confinamiento por la COVID-19 en muchos Estados miembros. Cuando se efectúe una EIC de una intervención que ya lleve

*Secuenciar varios tipos de evaluación*

tiempo en curso, sería más pertinente que la evaluación del proceso se realizase junto con la evaluación del impacto.

El calendario también se verá afectado por la disponibilidad de datos. Las fuentes de datos pueden tardar mucho tiempo en actualizarse, como suele ocurrir con los registros fiscales. Superar las barreras jurídicas e institucionales para la obtención de los datos necesarios también puede ser un proceso largo y costoso. Por otro lado, la utilización de datos procedentes de diversas fuentes, garantizando su compatibilidad, comprobando su calidad y moldeándolos de una forma que pueda servir para estimar los impactos, requiere un tiempo y un esfuerzo considerables.

*La recopilación de datos puede ser un proceso largo*

### ¿Cómo se pueden evaluar los costes?

Es importante establecer un presupuesto indicativo que refleje cuánto dinero puede gastar y está dispuesta a gastar una AG en una EIC. El presupuesto consta de dos componentes: costes internos y externos. Los primeros se refieren al esfuerzo necesario de los recursos internos para seguir la evaluación; en las administraciones, estos costes no suelen detallarse en términos financieros, pero deben tenerse en cuenta para asegurar un control y un seguimiento adecuados de la evaluación. Los costes de encargar la EIC a expertos externos deben calcularse cuidadosamente para garantizar una evaluación de gran calidad con un coste razonable. En los siguientes párrafos la atención se centrará en los costes externos.

Es necesario distinguir entre las evaluaciones de intervenciones rutinarias, en las que los gastos suelen ser inferiores, y las acciones innovadoras o piloto en las que la recopilación de una cantidad relativamente elevada de datos, el uso de nuevas fuentes de datos y la participación de nuevas partes interesadas podrían justificar un gasto mayor. Sin embargo, no se trata de una norma y es algo que se debe analizar en el contexto de las preguntas de evaluación, cuya complejidad y número pueden requerir una aportación financiera menor o mayor.

*Un posible método para evaluar los costes*

El enfoque de evaluación que se elija también supondrá una diferencia. Como se ha mencionado anteriormente, un enfoque aleatorizado requiere la presencia del evaluador para recopilar y verificar la información durante la intervención. Por otra parte, en los métodos cuasi experimentales el esfuerzo viene determinado en gran medida por el número de fuentes de datos, su calidad y su disponibilidad.

Un documento de orientación publicado por la Comisión<sup>67</sup> proporciona una indicación de cuánto cuesta la evaluación de un programa e indica que los programas a gran escala y rutinarios no deben dedicar más del 1 % de su presupuesto a la evaluación, mientras que las iniciativas innovadoras o piloto podrían destinar hasta el 10 %. Sin embargo, estas orientaciones no abordan explícitamente las necesidades de recursos de las EIC, y se puede asumir que esos porcentajes son el límite máximo de los costes de evaluación e incluyen una o varias EIC.

Como cabría imaginar, es imposible señalar el coste indicativo de una EIC para la elevada cantidad de intervenciones y condiciones operativas diferentes de los programas que el FSE+ lleva a cabo en los veintisiete Estados miembros. En este caso, resulta más útil proponer un método para calcular cuál sería un presupuesto razonable para las EIC en diferentes

<sup>67</sup> Véase [Comisión Europea, 2013](#).

contextos. En el siguiente cuadro se presenta un enfoque sencillo: en las filas figuran las principales tareas que deben realizarse, mientras que en las columnas se indican los costes principales.

La duración y la complejidad de cada actividad determinan el esfuerzo necesario en términos de días laborables del equipo de evaluación y, por consiguiente, su coste principal. Recopilar datos mediante encuestas cuantitativas a los participantes y a los miembros del grupo de control tiene un coste considerable. Cuando una EIC se base en las fuentes de datos administrativos existentes y estas sean fácilmente accesibles, los costes totales serán inferiores. No obstante, suele ser necesario someter los datos administrativos a una «preparación» esencial (corregir los datos faltantes o incorrectos, ajustar el formato de la base de datos, verificar las normas administrativas que rigen los datos, etc.), lo que puede aumentar los costes.

**Cuadro 3 Estructura de los principales costes de una EIC**

Actividades principales	Equipo de evaluación			Otros costes	Notas
	Responsable del programa	Expertos superiores	Expertos junior	(Equipo, materiales, suministros, viajes, etc.)	
Planificación y coordinación					El proceso de planificación abarca la viabilidad de la EIC, su organización y la ultimación de la metodología de acuerdo con la AG. En esta fase, se debe dedicar tiempo suficiente a analizar las posibles lagunas de datos. La coordinación abarca toda la duración de la EIC e incluye la organización del trabajo y las interacciones con la AG y otras partes interesadas pertinentes para concretar las preguntas de evaluación.
Revisión de la bibliografía					La revisión de la bibliografía respalda el esbozo de la teoría del cambio, la determinación de las variables que se utilizarán y la comprensión del contexto. También permite afinar las preguntas de evaluación, aprovechar resultados anteriores y formular hipótesis.
Recopilación y publicación de datos					El esfuerzo necesario para esta actividad varía considerablemente en función del método o métodos de recopilación de datos. La recopilación de datos originales, a través de encuestas, puede resultar larga y costosa. Otros métodos pueden requerir que se invierta en tecnología ( <i>software</i> o <i>hardware</i> ) o se llegue a acuerdos con los propietarios de los datos. Los costes pueden reducirse implicando a los propietarios de los datos en la preparación de estos.
Análisis de los datos					Para analizar los datos es preciso poseer una experiencia y unas capacidades avanzadas. El método, la cantidad de análisis y la calidad de los datos influyen en el tiempo que se dedica a esta actividad.
Preparación de informe(s)					El esfuerzo requerido para esta actividad varía en función del número y tipo de informes y otras herramientas de comunicación. Los costes pueden abarcar la impresión y el diseño gráfico, además del tiempo de preparación.
Reuniones de seguimiento					Las reuniones de seguimiento y otras actividades de difusión son un paso importante para dar a conocer las conclusiones. Los gastos varios podrían abarcar el alquiler de espacios y los alimentos.
<b>Total</b>					

A falta de otras referencias, el coste del personal puede calcularse en relación con los costes del FSE+ derivados de los servicios de formación o empleo adoptados en cada país para formadores o expertos experimentados y noveles. Por lo general, estas tasas son comparables a las de un investigador novel y experimentado y, aunque no tengan por qué ser equivalentes, son un parámetro útil que debe tenerse en cuenta. Un simple análisis de mercado y algunas entrevistas con unos pocos investigadores podrían ayudar a determinar el coste previsto de la EIC.

### **2.5.2. ¿Cuándo se debe evaluar la intervención?**

Es fundamental determinar en qué momento de una intervención es más adecuado llevar a cabo una evaluación del impacto, así como las cuestiones críticas de cuándo se deben medir los resultados y estimar los impactos.

#### **¿Cuándo se deben evaluar las intervenciones nuevas y en curso?**

El debate sobre en qué momento de una intervención conviene llevar a cabo una evaluación de impacto contrafactual dependerá de si la intervención es nueva o lleva tiempo en curso. Cuando la intervención es nueva, hace falta más tiempo para que madure y alcance cierta estabilidad. Una EIC que se realice antes de este punto será prematura y podría aportar pruebas engañosas. En el caso de las nuevas intervenciones, efectuar una evaluación preliminar del proceso antes de la EIC suele ser útil para detectar problemas iniciales e indicar soluciones.

*El calendario difiere para las intervenciones nuevas y en curso*

A la hora de determinar el calendario óptimo de una EIC de una intervención nueva, deben tenerse en cuenta otros factores, tales como las medidas para garantizar la disponibilidad de fuentes de datos adecuadas, la creación de un equipo interno del proyecto compuesto por personal debidamente formado y la designación de un contratista externo. Además, las necesidades del proceso decisorio al que se dirige en última instancia la evaluación constituirán una limitación crítica.

*Cuestiones en materia de recursos*

El calendario de una evaluación del impacto de una intervención en curso se regirá principalmente por requisitos prácticos y relacionados con las políticas. La intervención ya se debería haber asentado y alcanzado cierto grado de madurez, con lo que sería adecuado realizar una EIC. Otra cuestión que se debe tener en cuenta es la presencia de otras reformas paralelas a la intervención evaluada. Los efectos de estas reformas podrían influir en el impacto de la intervención considerada. Los responsables políticos tendrán que valorar si la presencia de otras reformas en el panorama político es pertinente para las decisiones políticas que se basarán en los resultados de la EIC en cuestión.

Las evaluaciones del FSE suelen centrarse en un período de programación. No obstante, y en especial en el caso de aquellas intervenciones estables que ya formaban parte del programa del FSE en el período anterior, puede resultar útil combinar una evaluación retrospectiva del período anterior con una evaluación continua del período actual, con vistas a cubrir un lapso más largo de una intervención.

## Cuándo medir los resultados y calcular los impactos

*Múltiples períodos de programación*

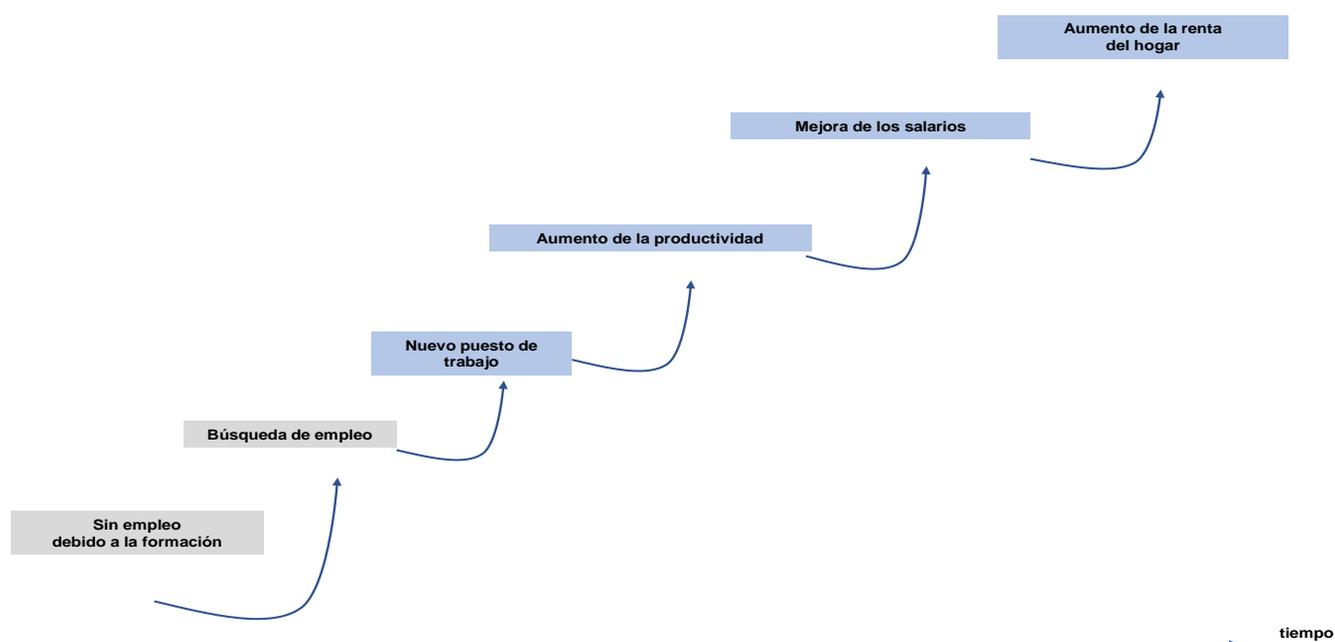
La segunda cuestión clave asociada al calendario de una evaluación es cuándo se deberían medir y estimar los impactos o, más concretamente, cuándo se prevé que aparecerán los impactos tras una intervención.

Al examinar una intervención de formación dirigida a la población desempleada, la cuestión es en qué plazo se podría esperar un aumento de las tasas de empleo. Se ha comprobado que, cuando se lleva a cabo un programa de formación, la tasa de empleo de sus participantes suele disminuir a corto plazo debido al denominado «efecto de bloqueo». Cuando se lleva a cabo una intervención de formación, es habitual que las personas desempleadas que participan en ellas abandonen la búsqueda de empleo para asistir a los cursos. En cambio, los integrantes del grupo de control tienen la determinación de encontrar un trabajo. Si los impactos se calculan demasiado pronto, podrían ser negativos o subestimarse. Al planificar una EIC, es importante ser realista en cuanto al momento en que se materializarán los impactos y cuándo es probable que sean mensurables. En el gráfico 5 se presenta un modelo simplificado de impactos posteriores.

Cuando se reflexione sobre cuál es el mejor momento para medir los resultados y estimar los impactos, habrá que tener en cuenta que los responsables políticos exigen recibir cierta información en determinados plazos. En el caso de las intervenciones destinadas a mejorar la empleabilidad a largo plazo, desde una perspectiva analítica podría tener sentido hacer un seguimiento de los participantes dos años o más después de haber estado expuestos al tratamiento, para comprobar si sus ingresos y tasas de empleo son superiores a las de algún grupo equivalente de personas no tratadas. Por el contrario, los responsables de programas a menudo necesitan conclusiones rápidamente, en cuyo caso hay que encontrar un equilibrio entre lo que es un intervalo de seguimiento razonable desde la perspectiva de la intervención y la necesidad de las autoridades decisorias de recibir pruebas con puntualidad.

*Entre lo razonable y lo viable*

**Gráfico 5 Calendario simplificado de resultados de un programa de formación**



Si las mediciones de los resultados se obtienen de fuentes administrativas (p. ej., registros de la seguridad social con detalles sobre el empleo y los ingresos), resultará práctico hacer un seguimiento repetido de los resultados a lo largo de un período prolongado de tiempo y estimar los impactos (incluso con periodicidad mensual). El riesgo es que la naturaleza de las conclusiones pueda cambiar con el tiempo. Si es necesario recopilar datos primarios para medir los resultados mediante encuestas por muestreo, la estimación de los impactos en intervalos regulares sería muy costosa, a menos que sea viable recabar datos retrospectivos sobre los resultados. No obstante, no se debe subestimar el coste de extraer datos de múltiples sistemas administrativos y de crear un único conjunto de datos analíticos.

*Centrarse en momentos concretos*

Como se ha expuesto en la sección 1.5, la articulación de una teoría del cambio (o lógica de intervención) puede ayudar a determinar cuándo se deben estimar los impactos.

Una alternativa para quienes planifican una EIC sin contar con una teoría del cambio, pero que también puede ser útil para quienes sí puedan basarse en una teoría del cambio clara, es llevar a cabo un breve análisis de estudios anteriores que evalúen intervenciones similares a las consideradas. Considerar detenidamente los resultados de estudios anteriores puede proporcionar una buena indicación de cuál es la forma adecuada de medir los resultados y calcular los impactos.

*...o analizar otros estudios recientes*

## 2.6. Realizar la EIC



La definición del calendario y el presupuesto constituyen el último paso en la planificación de la EIC, tras lo cual se puede proceder a realizar la evaluación. El último paso de esta Guía pretende ilustrar las principales actividades que se llevan a cabo al realizar la evaluación y cómo la AG puede superar los obstáculos con los que podría toparse. En particular, esta sección aplica las lecciones aprendidas por el Servicio de Asistencia para la Evaluación, a través del cual la CE y la DG EMPL prestan asistencia a las AG de todos los Estados miembros. La realización de la evaluación se examina en términos generales en *EVALSED: The Resource for the Evaluation of Socio-Economic Development*<sup>68</sup> [«EVALSED: el recurso para evaluar el desarrollo socioeconómico», documento en inglés]. La presente Guía se centra en cuatro actividades principales, concretamente desde el punto de vista de la EIC:

1. seleccionar el evaluador;
2. supervisar la realización de la EIC;

<sup>68</sup> Disponible en: [https://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/guide/guide\\_evalsed.pdf](https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf).

3. presentar información;
4. utilizar los resultados.

### 2.6.1. Seleccionar el evaluador

A la hora de seleccionar un evaluador, lo primero que hay que decidir es si se quiere recurrir a un candidato interno o externo. En este último caso, la selección del evaluador tiene lugar mediante un procedimiento de contratación pública.

*La independencia del evaluador*

El artículo 44, apartado 3, del RDC reconoce la existencia de un problema de independencia y establece que: «Las evaluaciones se confiarán a expertos internos o externos que sean funcionalmente independientes». Por lo tanto, el primer criterio para seleccionar un evaluador es su verdadera independencia respecto de la AG y los procesos decisorios del programa del FSE+. El segundo criterio es si internamente se cuenta o no con las capacidades y los recursos necesarios para llevar a cabo una EIC. Probablemente este sea el obstáculo más frecuente para las evaluaciones internas, pues para realizar una EIC se precisan unos conocimientos técnicos muy específicos que la mayoría de las administraciones no suelen poseer.

Una vez se ha decidido entre evaluadores internos o externos, la administración debe enumerar las especificaciones técnicas de la evaluación. En la selección de un evaluador externo, las especificaciones formarán parte del procedimiento de contratación pública. En los términos de referencia de una EIC, merecen especial atención los siguientes aspectos:

*Principales elementos de los términos de referencia*

- El **objetivo de la evaluación** y las preguntas de investigación deben ser claros y dar lugar sistemáticamente a una evaluación de impacto y a una EIC, en la medida de lo posible.
- Deben especificarse los **datos disponibles**. Esta es una de las principales cuestiones a las que se debe hacer frente en el diseño y la realización de una EIC, pues hace que los evaluadores sean conscientes de los problemas que existen en lo tocante a la disponibilidad de datos y les permite formular propuestas realistas y precisas.
- La **solicitud de llevar a cabo una EIC** puede manifestarse explícitamente en los términos de referencia, pero el evaluador debe ser quien establezca los métodos concretos que se utilizarán. De este modo, podrá proponer soluciones originales y demostrar su capacidad.
- El **equipo propuesto** debe poseer una combinación de conocimientos especializados en varios ámbitos, y en particular en los campos de la evaluación y la EIC (saber cómo diseñar y aplicar una evaluación), la econometría (utilizar las metodologías adecuadas) y el ámbito político de la intervención evaluada (reconstruir la teoría del cambio e interpretar las conclusiones en su conjunto).
- Los **criterios de adjudicación** deben dar prioridad a la calidad frente al coste de la propuesta. La calidad de una evaluación depende en gran medida de la capacidad del evaluador para diseñarla y ejecutarla. A menudo, los marcos jurídicos aplicables a la contratación pública no responden de manera perfecta a las necesidades de las licitaciones públicas de evaluaciones, pero quien vaya a encargar una debe adaptar, en la medida de lo posible, las normas existentes para recompensar la calidad y la capacidad técnica.

- En el **proceso de selección** se debe incluir a expertos en evaluación y EIC en el comité de selección, a fin de valorar plenamente la calidad de las propuestas y evaluar correctamente sus diversas soluciones metodológicas y organizativas.

Algunos Estados miembros todavía tienen una experiencia ínfima en la realización de EIC. Esto puede suponer un problema en el proceso de selección, dado que el mercado de este servicio es demasiado restringido. Es posible que muy pocas empresas o expertos puedan cumplir las condiciones técnicas de los términos de referencia y presentar una propuesta de buena calidad. En estos casos, los seminarios abiertos de la AG con académicos y empresas interesadas en las evaluaciones podrían ayudar a preparar la EIC y la consiguiente convocatoria de propuestas. En estos seminarios, expertos nacionales presentarían sus propuestas y tendrían tiempo para organizar su participación en la licitación. A fin de respetar la transparencia y evitar conflictos de intereses, los seminarios deben ser públicos, y la información básica también debe ponerse a disposición de quienes no puedan asistir.

### **2.6.2. Supervisar la EIC**

Una vez que se ha elegido al evaluador y se ha puesto en marcha la EIC, la AG tiene que supervisar el proceso de evaluación. Esto implica, principalmente, controlar y validar los resultados del evaluador, organizar reuniones de progreso con el evaluador y las partes interesadas y, en caso necesario, llegar a acuerdos con los propietarios de los datos.

La AG debe tener la capacidad necesaria para llevar a cabo estas tareas. Esto implica también disponer de recursos para supervisar todo el proceso y poseer las capacidades necesarias para comprobar las soluciones técnicas de la EIC. Por lo general, la AG destina miembros específicos del personal a las evaluaciones y, cuando estos no tengan suficiente experiencia con las EIC, expertos externos independientes deben brindarles apoyo. Estos expertos podrían prestar sus servicios en el marco de contratos a corto plazo o, cuando procedan de instituciones públicas, los servicios serán gratuitos. Probablemente tendrán que dedicar entre seis y doce días laborables, lo que debería bastar para revisar los informes y asistir a algunas reuniones con la AG, el evaluador y las principales partes interesadas.

*Capacidades  
internas y externas  
en el proceso de  
supervisión*

La creación de un grupo director es una forma habitual de supervisar los procesos de evaluación de un programa. Este grupo suele estar integrado por la AG, funcionarios que representan a otros departamentos encargados de la ejecución, las principales partes interesadas (interlocutores sociales, ONG, otras administraciones) y algunos expertos del mundo académico o de instituciones públicas (p. ej., el instituto nacional de estadística). El grupo director participa en la definición de las preguntas de evaluación y examina las conclusiones de la evaluación, su difusión y su utilización en los procesos de formulación de políticas. Un subgrupo más flexible y especializado también podría supervisar los resultados de la evaluación con la AG.

### **2.6.3. Presentar información**

La presentación de información es una actividad clave de la evaluación y constituye el principal medio para comunicar los resultados. Los informes de evaluación deben ser claros, concisos, comprensibles para personas sin conocimientos técnicos y transparentes en sus valoraciones y recomendaciones políticas. Asimismo, deben demostrar que la evaluación es

fiable mediante la especificación de la metodología y los datos utilizados, así como la determinación de los posibles límites del análisis.

En general, en cada EIC se elaborarán tres informes:

- El **informe preliminar** suele presentarse poco después de la firma del contrato y expone la metodología terminada de acuerdo con los debates mantenidos con la AG y el examen inicial de los datos disponibles.
- El **informe intermedio** se presenta en una fase intermedia de la evaluación y en una EIC puede servir para mostrar los datos recogidos, la composición de los grupos de tratamiento y de control, y el muestreo.
- El **informe final** contiene los análisis y las conclusiones de la EIC. Incorpora detalles relativos a la metodología y los datos, y pone de relieve las implicaciones políticas que se deriven de las pruebas. La estructura del informe final de una EIC debe comprender algunas secciones clave: las preguntas de la evaluación, la teoría del cambio (o lógica de intervención) y la determinación de los efectos directos que deben evaluarse, la metodología adoptada, los datos utilizados y las características de los grupos de tratamiento y de control, los efectos estimados, y las respuestas a las preguntas de la evaluación y sus implicaciones políticas.

*Informes distintos para hacer un seguimiento de los avances y utilizar los resultados*

La persona que ha encargado la evaluación revisa todos los informes con el apoyo de expertos externos, cuando sea necesario, y analiza sus comentarios con el evaluador, que tramitará las aclaraciones y mejoras solicitadas antes de la aceptación oficial del informe.

El informe de una EIC no debe ser excesivamente técnico y debe exponer resultados claros y legibles para las partes interesadas y los no expertos. Los pasos más técnicos del análisis podrían explicarse en anexos específicos, a fin de garantizar que el informe principal sea claro y fluido. No obstante, los detalles metodológicos son cruciales para demostrar la fiabilidad de la EIC y también deben ser transparentes. En el informe se deben especificar las variables de resultado, la calidad de los datos, la composición de los grupos de tratamiento y de control y el tamaño de la muestra, el método de estimación de los efectos y las pruebas estadísticas utilizadas. Además, en consonancia con las prácticas académicas, los datos anonimizados usados y los métodos de tratamiento de datos podrían ponerse a disposición de los investigadores que tengan la intención de repetir y verificar los análisis.

El informe final de todas las EIC debe contener un conjunto normalizado de información mínima. Este conjunto mínimo de información tiene un doble objetivo: 1) facilitar la comparación de las conclusiones entre distintos PO y países; y 2) recabar información que estará disponible para futuros meta análisis.

*El informe final de una EIC debe contener un conjunto mínimo de información*

El modelo para recoger un conjunto normalizado de información mínima sobre cada evaluación del FSE sigue las recomendaciones formuladas en el «Estudio piloto y de viabilidad sobre la sostenibilidad y la eficacia de los resultados para los participantes del Fondo Social Europeo mediante evaluaciones de impacto contrafactuales». Desde un punto de vista operativo, esta información podría exponerse en un anexo del informe final y también podría remitirse a la CE. En el

Cuadro 4 se presenta un posible modelo para recopilar el conjunto mínimo de información necesario.

**Cuadro 4 Información básica que debe contener la ficha de presentación de la EIC**

Categoría de información	Información específica que debe recopilarse y presentarse
1. Información sobre la intervención	<ul style="list-style-type: none"> <li>- Nombre de la intervención evaluada</li> <li>- Características de la intervención (formación, servicios de empleo, servicios para la inclusión social, etc.)</li> <li>- Enfoque del programa del FSE+ (ejes prioritarios, objetivo específico, acción)</li> <li>- Autoridad de gestión u otros organismos de ejecución</li> <li>- Gastos</li> <li>- Duración/intensidad del tratamiento proporcionado por la intervención</li> <li>- Grupo participante por edad, género y situación de elegibilidad (beneficiarios del seguro de desempleo, parados de larga duración, ninis)</li> <li>- Ámbito territorial de la intervención</li> </ul>
2. Medición de la eficacia de la intervención	<ul style="list-style-type: none"> <li>- Indicador de si la EIC estima que el tratamiento tendrá un efecto positivo y estadísticamente significativo en los efectos directos pertinentes</li> <li>- La magnitud real de dicho efecto</li> <li>- Las dos medidas anteriores para el conjunto de efectos directos pertinentes y comparables: p. ej., tasa de empleo, ingresos, etc.</li> <li>- Indicadores de la «relación calidad-precio» (si se estima)</li> </ul>
3. Datos y metodología	<ul style="list-style-type: none"> <li>- Fuente de los datos (encuesta, datos administrativos)</li> <li>- período de observación</li> <li>- Método econométrico de la EIC</li> <li>- Horizonte temporal (generalmente, efectos a corto plazo: doce meses o menos desde el final de la intervención; a medio plazo: entre doce y veinticuatro meses; a largo plazo: más de veinticuatro meses)</li> <li>- Tamaño de la muestra de los grupos de tratamiento y de control</li> <li>- Disponibilidad/incorporación de datos previos al programa</li> </ul>

**2.6.4. Utilizar los resultados**

El uso de los resultados es importante porque, si no se difunden de manera eficaz y llegan al público previsto, la evaluación tendrá poca repercusión.

*Es crucial difundir las conclusiones de la EIC*

La difusión de las conclusiones y de los productos de la evaluación suele conllevar:

- al menos un informe de evaluación escrito, que contenga un *abstract* y un resumen;
- al menos una presentación verbal de las conclusiones, respaldada por diapositivas o herramientas similares;
- una sección técnica (o anexo) del informe en la que se expongan exhaustivamente la metodología utilizada, las hipótesis clave formuladas y el enfoque adoptado para los análisis estadísticos.

Todos los informes de evaluación deben hacerse públicos, tal y como establece el Reglamento sobre Disposiciones Comunes para el período de programación 2021-2027<sup>69</sup>. Además, las conclusiones de la evaluación deben presentarse y debatirse en el comité de seguimiento del programa. Este trámite es obligatorio, pero no basta por sí solo para promover un debate en profundidad, puesto que en las reuniones de los comités de seguimiento se suele hablar de las prioridades operativas. En el recuadro que figura a

<sup>69</sup> Artículo 44, apartado 7, del RDC.

continuación se expone un ejemplo de práctica nacional de presentación y difusión de las conclusiones de una evaluación.

No existe una única política de comunicación eficaz para difundir y generar un debate sobre las lecciones aprendidas; cada contexto requiere su propia estrategia. Por lo tanto, es importante elaborar una estrategia eficaz y, en particular, velar por que las partes interesadas, aparte de la AG, conozcan las conclusiones. Es posible dividir el «público» de la EIC en grupos principales y determinar los principales tipos de comunicación; por ejemplo:

- Responsables políticos de alto nivel y representantes políticos: están acostumbrados a ver documentos breves (informes políticos, resúmenes) con los principales resultados y recomendaciones.
- Funcionarios de la administración que participan en la ejecución de la intervención o políticas similares: los seminarios o talleres pueden ser un instrumento eficaz para presentar los resultados de la EIC, adaptando el nivel técnico de la presentación al nivel de interacción con los asistentes.
- Partes interesadas políticas (interlocutores sociales, ONG, beneficiarios, etc.): el resumen de los informes y las reuniones anuales sobre las conclusiones de la evaluación podrían ser el instrumento combinado para implicar a estos agentes en el debate. En particular, cuando estos agentes hayan contribuido a la definición de las preguntas de evaluación, la comunicación de los resultados de la evaluación debe vincularse a dichas preguntas.
- Expertos y académicos: las conferencias o seminarios científicos pueden ser los mejores lugares para presentar los resultados de la evaluación. Estos agentes validan los resultados de la evaluación desde un punto de vista científico, con lo que refuerzan su fiabilidad.

*Múltiples públicos y múltiples estrategias de comunicación*

La meta evaluación<sup>70</sup> es un valioso instrumento adicional para difundir los resultados de las EIC. Permite confirmar y generalizar las principales conclusiones que no se pueden derivar de una única EIC. Para llevar a cabo una meta evaluación, es necesario disponer de un buen número de EIC fiables y de buena calidad sobre políticas comparables. En la actualidad, las meta evaluaciones siguen siendo poco comunes y no se centran únicamente en las intervenciones del FSE. Para que las meta evaluaciones se utilizasen de forma habitual, sería necesario generalizar el uso de las EIC y planificarlas por adelantado, a fin de facilitar su posterior comparación. En el período de programación 2021-2027 del FSE+, se podrían efectuar meta evaluaciones en países con un número significativo de programas o mediante comparaciones entre países.

*Meta evaluación para reforzar la difusión y generalización*

<sup>70</sup> Se trata de una comparación sistemática, incluso con el apoyo de herramientas estadísticas, de numerosas EIC de intervenciones similares. Una meta evaluación examina los efectos de medidas similares en múltiples contextos distintos con miras a proporcionar indicaciones de la eficacia de un tipo de intervención. Sin embargo, no pueden analizar en detalle cada intervención utilizada en la comparación, y sus implicaciones políticas se refieren a orientaciones políticas generales, no a detalles particulares de cada intervención.

**Recuadro 13 La experiencia polaca con las conferencias de evaluación**

La Conferencia Internacional de Evaluación, organizada por el Ministerio de Fondos de Desarrollo y Política Regional y la Agencia de Desarrollo Empresarial polaca, es uno de los elementos clave del panorama de la evaluación de las intervenciones públicas en Polonia. Desde su creación, se ha convertido en una plataforma consolidada para que las partes interesadas responsables de la formulación y ejecución de políticas y la evaluación de los efectos debatan entre sí.

La conferencia, que se organizó por primera vez en 2005, se celebró anualmente durante diez años, y dos veces al año desde 2015, con lo que en 2021 tuvo lugar su 14.<sup>a</sup> edición. En la actualidad, se organiza como un acto de dos días de duración. Un día se dedica a las conclusiones de la investigación y los efectos de las intervenciones, así como a hablar sobre las políticas. El segundo día se centra en cuestiones metodológicas y en la evaluación como proceso.

A lo largo de las catorce ediciones de la conferencia, los métodos contrafactuales han sido objeto de mesas redondas o presentaciones de expertos en varias ocasiones, y se han abordado desde distintas perspectivas en ambos días. Por ejemplo, en 2017 se presentaron los efectos estimados en sesiones relativas al impacto de las intervenciones, y se celebraron por separado debates sobre las ventajas y los inconvenientes de las técnicas de estimación.

La conferencia también es importante debido a la diversidad del público que asiste a ella, ya que se dirige no solo a investigadores o especialistas en materia de metodología, sino también a quienes diseñan y ejecutan las intervenciones. El acto tiene dos objetivos: desarrollar las capacidades del sistema de evaluación mejorando las capacidades y los conocimientos de sus miembros, y crear redes al proporcionar a las partes interesadas un espacio para intercambiar experiencias entre socios nacionales, regionales y extranjeros. De hecho, cada edición congrega aproximadamente a entre trescientos y cuatrocientos representantes de diversos sectores: la administración pública, el mundo académico, empresas consultoras y ONG. Se trata de un número considerable, dado que el enfoque empírico de la política todavía no está plenamente arraigado en Polonia.

La Conferencia Internacional de Evaluación forma parte del sistema de evaluación de la administración pública. Las formaciones impartidas por el Centro Común de Investigación y el Banco Mundial, o financiadas en el marco de la asistencia técnica en programas operativos, contribuyeron a un importante desarrollo de capacidades. Se ofreció formación en materia de evaluación, y también sobre el enfoque contrafactual, a los empleados de las unidades de evaluación de las autoridades de gestión y a instituciones como las oficinas de empleo locales o regionales. La conferencia brinda la oportunidad de reunir, debatir y compartir los avances actuales y las conclusiones<sup>71</sup>.

---

<sup>71</sup> Se puede consultar información sobre la conferencia en el siguiente enlace: <https://www.ewaluacja.gov.pl/strony/xiv-miedzynarodowa-konferencja-ewaluacyjna/>. Quisiéramos dar las gracias a Piotr Strzęboszewski y otros funcionarios del Ministerio de Fondos de Desarrollo y Política Regional por la información que nos han facilitado acerca del sistema de las conferencias.

# Capítulo 3. Cómo seleccionar la metodología adecuada para llevar a cabo una EIC

En este capítulo se presentan métodos empíricos para la evaluación de impacto contrafactual. En concreto, se analiza el enfoque experimental (ensayo aleatorio controlado), así como los métodos cuasi experimentales más comunes: el emparejamiento (por puntuaciones de propensión), la diferencia en diferencias, el diseño de regresión discontinua y las variables instrumentales. Cada método se caracteriza por generar de una forma específica el grupo de control que responderá a la pregunta contrafactual: «¿Qué le habría ocurrido al grupo de tratamiento en términos de efectos directos si no hubiera participado en la intervención o no hubiera estado expuesto a ella?». En la práctica, el diseño elegido puede adaptarse al contexto, conforme al tipo de intervención y los datos disponibles o que puedan recopilarse<sup>72</sup>.

No es posible proporcionar orientaciones detalladas sobre la elección del diseño de evaluación más adecuado en las circunstancias muy diversas a las que se enfrentan las AG. Cuando elijan el enfoque más útil para efectuar una EIC en unas circunstancias particulares, las AG deben valorar qué ha funcionado bien en evaluaciones anteriores, tanto en el seno de la propia AG como en el Estado miembro donde se encuentra y en otros Estados miembros. Las AG pueden aprender de los logros conseguidos en el pasado con su programa y en otros lugares bajo circunstancias similares. Los foros para el intercambio de las lecciones aprendidas en el diseño y la realización de evaluaciones pueden ser fuentes de información útiles en este sentido. La búsqueda de bibliografía sobre evaluaciones de intervenciones similares también puede ser una fuente importante de información para facilitar el proceso de diseño. Los expertos designados por la AG también tendrán opiniones sobre la mejor manera de enfocar el diseño de una evaluación. Conviene recordar que las AG podrían poseer unos conocimientos especializados y una experiencia considerables que pueden aprovecharse.

*Seleccionar el enfoque adecuado*

## 3.1. Aleatorización: el enfoque experimental

### *Características fundamentales de la aleatorización*

- A las personas que reúnen los requisitos para participar se las asigna aleatoriamente a un grupo de tratamiento o un grupo de control.
- La aleatorización garantiza que ambos grupos sean idénticos (en promedio) en todas las características pertinentes
- Por lo tanto, el grupo de control responde a la pregunta contrafactual y la diferencia en los efectos directos entre el grupo de tratamiento y el grupo de control indica el efecto causal de la intervención.

<sup>72</sup> Para conocer otros aspectos metodológicos técnicos, véase también [Comisión Europea, 2019](#).

Las evaluaciones aleatorizadas se pueden diseñar de múltiples formas. En esta Guía, la atención se centra en un enfoque sencillo de dos grupos —un grupo de tratamiento y otro de control—, con vistas a aclarar los principios fundamentales. En el gráfico 1 se muestra un diseño aleatorizado simple.

La clave es que la aleatorización asegura que los dos grupos sean estadísticamente equivalentes en todos los aspectos en el momento de la aleatorización. Posteriormente, se expone al grupo de tratamiento a la intervención que es objeto de la evaluación y cuyos impactos o efectos van a medirse.

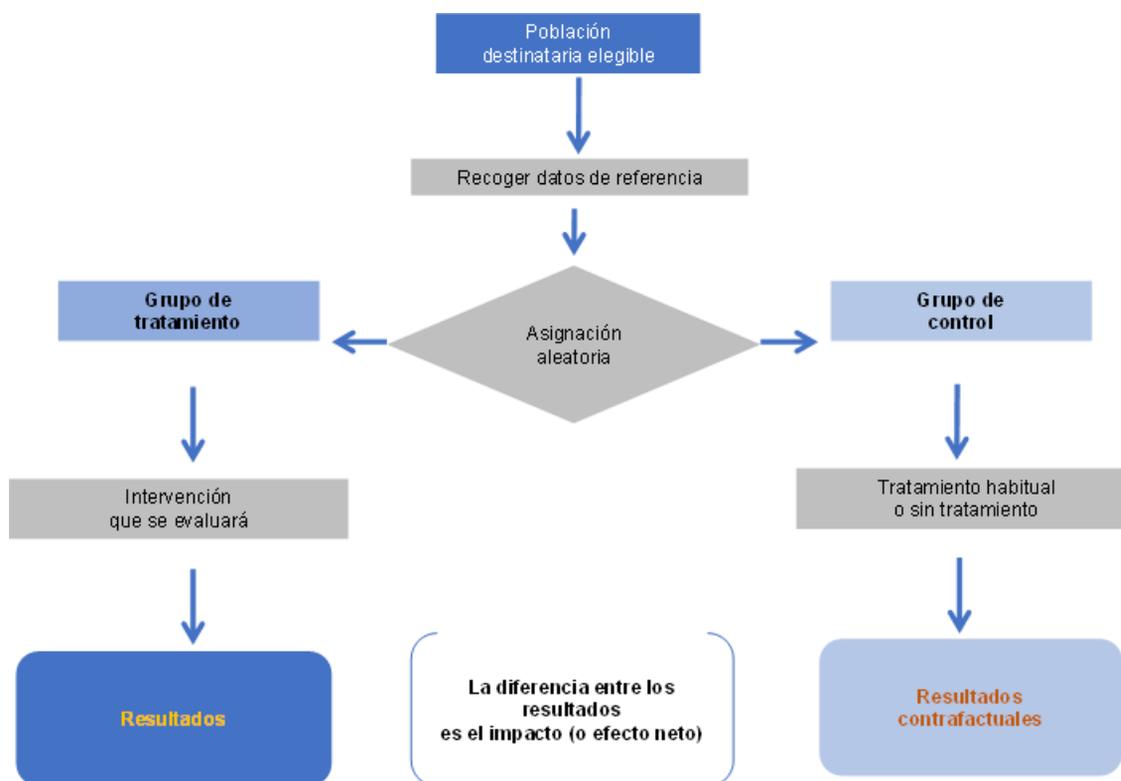
*Grupos estadísticamente equivalentes*

Dependiendo de la cuestión política de interés central, se puede decidir que el grupo de control no reciba ningún tratamiento, se puede comparar al grupo de tratamiento con un grupo expuesto a otro tratamiento (podría considerarse que representa el tratamiento habitual), o puede haber varios grupos de tratamiento asociados a un grupo de control. Por ejemplo, quizás interese comparar los efectos de un programa de formación financiado por el FSE con otra formación dotada de financiación nacional, o con la prestación de otros servicios a los mismos grupos.

*Grupo de control sin tratamiento o con un tratamiento distinto*

Dado que los grupos de tratamiento y de control son estadísticamente equivalentes en el momento de la aleatorización y que se controla la exposición a tratamientos posteriores, las diferencias en los resultados pueden atribuirse a la intervención evaluada (sujeta a incertidumbres estadísticas estándar) y se puede descartar que explicaciones alternativas sean las causantes de las diferencias observadas (el recuadro que figura más abajo contiene un ejemplo de esto).

**Gráfico 6 Control aleatorizado con dos grupos**



Debido a sus características de diseño intrínsecas, si se aplican correctamente, los diseños aleatorizados ofrecen la posibilidad de aportar pruebas sólidas de los efectos de una intervención. Por eso hay una clara

*Pruebas sólidas...*

preferencia por ellos. Sin embargo, requieren una planificación temprana y detallada y pueden resultar difíciles de diseñar y gestionar. Además, los responsables de programas se enfrentan a grandes dificultades para su correcta aplicación. Por ejemplo, la presencia misma del proceso de aleatorización podría alterar la composición de los participantes en una intervención: es decir, la idea de la aleatorización podría disuadir a algunos participantes potenciales, que podrían negarse a participar. Del mismo modo, es posible que las personas sujetas a la aleatorización no siempre respeten el papel que se les asigne, y podría ser necesario hacer frente a otras dificultades. En algunas circunstancias, puede resultar costoso poner en práctica diseños de ensayos aleatorios controlados.

Por estas y otras razones, parece poco probable que las evaluaciones de los instrumentos e intervenciones financiados por el FSE se lleven a cabo utilizando un enfoque aleatorizado. Sin embargo, la presente Guía recomienda evitar el impulso de descartar la aleatorización en todos los casos sin estudiarla debidamente. Este enfoque se ha utilizado ampliamente. Además del ejemplo del Reino Unido que se analiza en el recuadro siguiente, cabe mencionar otros, como los experimentos GAIN de Estados Unidos<sup>73</sup> (hay muchos otros ejemplos de América del Norte), los experimentos realizados en Suecia<sup>74</sup> con las intervenciones del FSE de 2014-2020 (véase el siguiente recuadro), así como un estudio llevado a cabo en Alemania para evaluar los efectos de los servicios activos del mercado de trabajo prestados por proveedores privados en comparación con los prestados a través del Servicio Público de Empleo<sup>75</sup>, entre muchos otros.

*...pero diseño complicado*

#### Recuadro 14 Ejemplo de ensayo aleatorio de un proyecto del FSE destinado a los jóvenes

##### Evaluación del proyecto Ung framtid (Futuro Joven) en Suecia

El proyecto sueco Ung framtid (Futuro Joven) es un proyecto financiado por el FSE que se llevó a cabo entre 2015 y 2018. Su objetivo era ayudar a jóvenes de entre dieciocho y veinticuatro años demandantes de empleo a incorporarse al mercado laboral en las regiones de Norrland Medio, Suecia Centro-Septentrional y Suecia Meridional. El proyecto, que fue llevado a cabo por el Servicio Público de Empleo sueco (el Arbetsförmedlingen), intensificó y personalizó las actividades de apoyo dirigidas a jóvenes demandantes de empleo. En concreto, consistió en realizar una planificación individual, proporcionar información, y brindar asesoramiento y apoyo específico en términos de adecuación al mercado laboral, formación y otras posibles actividades. En total, el proyecto llegó a casi 17 000 jóvenes.

El Consejo Sueco del FSE encargó la evaluación al Arbetsförmedlingen como parte del marco de evaluación desarrollado en un proyecto del FSE 2014-2020, cuya finalidad era mejorar la capacidad de evaluación de los SPE (proyecto con base empírica del fondo de la UE para 2014-2020; véase el recuadro que figura en el último capítulo de la Guía).

En un experimento realizado entre junio de 2017 y enero de 2018 en el que participaron ocho oficinas locales de empleo de las noventa que participaron en el proyecto, se distribuyó aleatoriamente a los jóvenes elegibles entre el grupo de tratamiento (las personas que recibieron un apoyo reforzado) y los grupos de control (las personas que recibieron el apoyo habitual)<sup>76</sup>. En el experimento, se repartió aleatoriamente a 4 689 entre los grupos de tratamiento y de control; de ellos, 2 972 fueron asignados al grupo de control. El proceso de asignación aleatoria dio lugar a grupos de tratamiento y de control muy similares en el momento de la distribución. Por consiguiente, cualquier diferencia entre los dos grupos en lo que respecta a las variables de resultado (fin de la situación de desempleo, proporción de desempleados

<sup>73</sup> Véase [Riccio, J.; Friedlander, D. y Freedman, S., 1994.](#)

<sup>74</sup> Véase [Hagglund, P., 2006.](#)

<sup>75</sup> Véase [Krug, G. y Stephan, G., 2011.](#)

<sup>76</sup> Entre abril y mayo de 2017, tuvo lugar una fase piloto en la que se pusieron a prueba varias actividades, como la gestión de la aleatorización, la implicación de los jóvenes participantes y la presentación de información.

y promedio de días de desempleo), medida tras acceder a la intervención del FSE, puede atribuirse a las actividades de Futuro Joven.

Las conclusiones del estudio muestran que el proyecto Futuro Joven tuvo impactos positivos para las mujeres, pero no para los hombres. En el caso de las mujeres, se observaron efectos positivos en todas las variables de resultado consideradas: el fin de la situación de desempleo de las mujeres tratadas fue aproximadamente un 7 % superior en los dos primeros meses tras el acceso al proyecto, la proporción de mujeres desempleadas disminuyó más en el grupo tratado que en el grupo de control y, por último, las mujeres tratadas estuvieron desempleadas menos días que las del grupo de control<sup>77</sup>.

Los diseños aleatorizados pueden distinguirse de otros enfoques principalmente porque hacen un gran hincapié en controlar el posible sesgo entre el grupo de tratamiento y el grupo de control mediante el diseño de la investigación. Este hincapié en el diseño hace que el enfoque sea muy intuitivo, pero requiere una planificación previa. A menudo, los diseños aleatorizados se aplican mejor para evaluar nuevas intervenciones piloto, en lugar de intervenciones ya existentes. Esto se debe a que requieren cierto grado de control sobre cómo se capta a los participantes para la intervención evaluada. Con frecuencia, cuesta más conseguir este «control» en los programas existentes que en las intervenciones nuevas que están abiertas a nuevas ideas.

*Aleatorización mediante el diseño de la investigación*

La aplicación de un diseño controlado aleatorizado implica que a una parte de la población destinataria elegible no se le ofrece la posibilidad de participar en la intervención y, en su lugar, es asignada a un grupo de control. Esta asignación es totalmente aleatoria y no depende ni del solicitante ni de los administradores de la intervención. Por este motivo, es posible que los responsables políticos tiendan a oponerse a los EAC por razones éticas antes de considerar si son viables desde un punto de vista práctico y analítico.

*Objeciones éticas*

Sin embargo, hay argumentos sólidos a favor de los diseños aleatorizados. Si la aleatorización proporciona la mejor calidad, la prueba más fiable de la eficacia de las intervenciones financiadas con fondos públicos, es importante que se utilicen más ampliamente para evaluar los impactos de las intervenciones. Además, si *a priori* no se conocen los impactos de una intervención concreta, no es poco ético excluir a determinadas personas, ya que no se puede asumir que de lo contrario habrían salido beneficiadas. Es más, estos enfoques se utilizan ampliamente en la medicina y otros ámbitos de estudio, como la investigación educativa. Por último, en algunas circunstancias en las que los servicios y el apoyo prestados por una intervención tienen un exceso de demanda (es decir, hay más personas elegibles de las que realmente pueden ser atendidas por la intervención), es posible que la forma más ética de distribuir los escasos recursos sea elegir aleatoriamente a las personas que participarán en la intervención entre todas las que reúnen los requisitos para ello.

*...pero también argumentos sólidos para utilizar un diseño aleatorizado*

### 3.2. Diseños no aleatorizados o cuasi experimentales

Si la aleatorización no es viable, la caja de herramientas para las EIC contempla una serie de métodos alternativos que aplican diferentes enfoques para generar un grupo de control que responda a la pregunta

*Cuando la aleatorización no es posible*

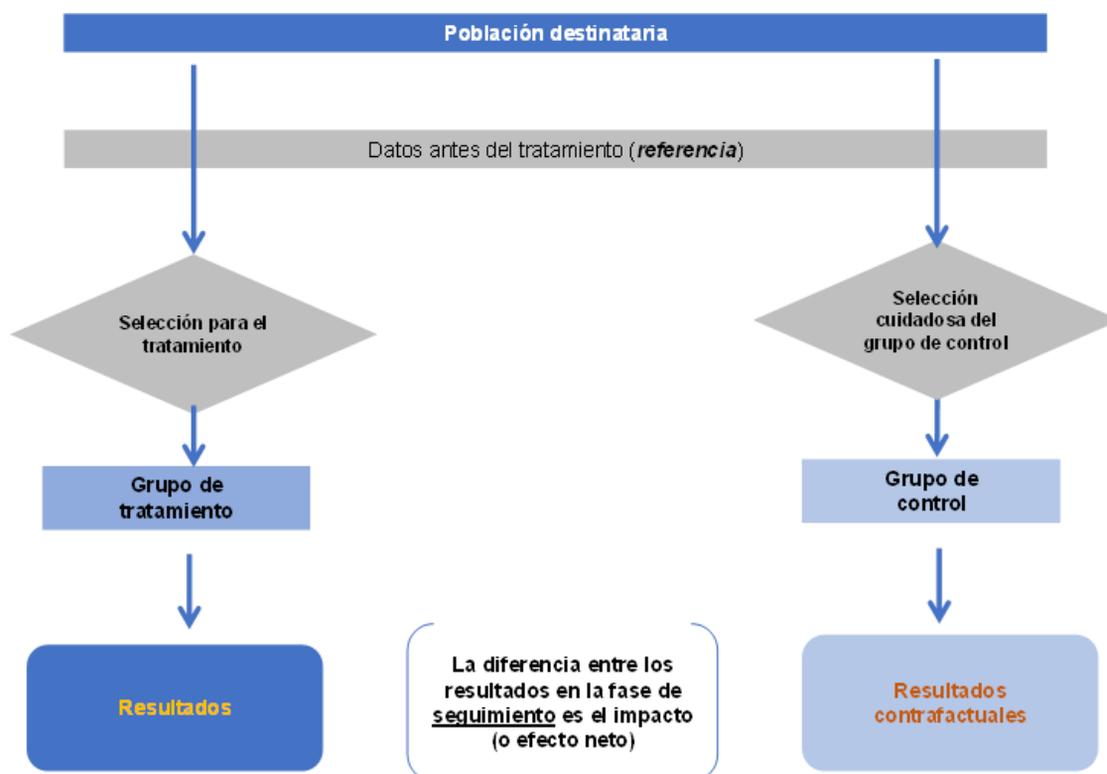
<sup>77</sup> Axdorph, E; Egebark, J.; Lundström, T. y Özcan, G., 2019.

contrafactual. En esta subsección se expone, en primer lugar, la principal dificultad de los métodos cuasi experimentales en general y, a continuación, se presentan cuatro enfoques que se utilizan habitualmente en la práctica empírica de la EIC.

### 3.2.1. Grupo destinatario y grupo de control sin aleatorización

En los diseños cuasi experimentales, los grupos destinatarios que reciben la intervención se comparan con un grupo de control<sup>78</sup> de destinatarios asignados de forma no aleatoria o destinatarios potenciales que no reciben la intervención. Al igual que en un experimento, el objetivo es obtener una estimación imparcial del cambio en los efectos directos que ha producido la intervención en cuestión. Puesto que los grupos de tratamiento y de control no se forman al azar, en los diseños cuasi experimentales es preciso prestar mucha más atención a los métodos que analizan las posibles diferencias entre los miembros del grupo de tratamiento y las posibles unidades de control que es probable que afecten a la decisión de participar y, por tanto, a los resultados. La clave reside en seleccionar un grupo de control plausible. La incapacidad de seleccionar un grupo de control adecuado y de explicar las diferencias restantes entre ambos grupos en el análisis resta credibilidad a las estimaciones y puede desbaratar los intentos de descartar explicaciones alternativas para cualquier efecto observado.

**Gráfico 7 Diseño cuasi experimental tipo con grupos de tratamiento y de control**



Por lo que se refiere a las intervenciones cofinanciadas por el FSE, el diseño cuasi experimental de la evaluación que se suele aplicar con mayor frecuencia es un diseño de referencia/seguimiento con dos grupos. Estos *Los grupos de control y de*

<sup>78</sup> Estrictamente hablando, «grupo de control» es el término usado en los diseños experimentales; en los diseños cuasi experimentales se suele hablar de «grupos de comparación». No obstante, en la práctica estos dos términos se suelen usar de forma intercambiable. En la presente Guía, por tanto, se emplea el término «grupo de control» para referirse al grupo utilizado para estimar la hipótesis contrafactual, con independencia del diseño de la EIC.

diseños incluyen un grupo de control y un grupo de tratamiento, como en la aleatorización, con la salvedad de que el grupo de control se crea (sin recurrir a la aleatorización) a partir de los grupos no participantes existentes, lo que lo asemeja lo más posible al grupo tratado. *tratamiento deben ser similares entre sí*

Una buena estrategia para encontrar un grupo de control válido en un entorno cuasi experimental consiste en seleccionar personas de control que hayan sido excluidas del tratamiento por factores no relacionados con sus características y resultados potenciales. En algunas circunstancias, puede haber motivos para creer que, aunque los grupos de control no se hayan creado explícitamente al azar, es posible encontrar *a posteriori* personas o empresas cuya no exposición al tratamiento resulte ser aleatoria con respecto a los resultados potenciales. Si se dan estas circunstancias, se aproximan a la situación ideal en el contexto de un enfoque cuasi experimental. Por ejemplo, algunos miembros del grupo destinatario de una intervención podrían quedar excluidos de participar en la intervención a causa de una omisión o un error administrativo. Por lo tanto, nunca se recalcará lo suficiente que entender el proceso de selección para participar en el tratamiento es sumamente importante de cara a crear un grupo de control válido.

Hay varias maneras de crear un grupo de control creíble. En primer lugar, se puede adoptar un enfoque de emparejamiento estadístico en el que se recaben datos tanto de las personas tratadas como de una muestra (normalmente muy amplia) de personas no tratadas. A continuación, entre las personas no tratadas se elige a las que son más similares a las que integran el grupo de tratamiento para crear el grupo de control. La «similitud» se refiere a un conjunto de características sociodemográficas —p. ej., edad, género, nivel educativo, empleo y desempleo, etc.— medidas en un momento anterior a que el grupo de tratamiento entrase en el programa. Como consecuencia, las personas no tratadas se «emparejan» de manera efectiva con las personas tratadas. En la práctica, la lista potencialmente larga de características sociodemográficas puede resumirse utilizando el valor de la «puntuación de propensión», que facilita la aplicación.

*Emparejar personas tratadas y no tratadas*

### 3.2.2. Emparejamiento por puntuaciones de propensión

#### *Características fundamentales del emparejamiento por puntuaciones de propensión*

- El emparejamiento imita a la aleatorización en tanto en cuanto crea *ex post* un grupo de control que se asemeja lo máximo posible al grupo de tratamiento en todas las características pertinentes.
- A diferencia de la aleatorización, solo pueden emparejarse las características observables (edad, género, nivel educativo, etc.), mientras que no es posible tener en cuenta aquellas que no pueden observarse (p. ej., la motivación).
- La validez del enfoque depende en gran medida de la disponibilidad de datos.

El **emparejamiento por puntuaciones de propensión** (EPP) implica estimar un modelo estadístico para toda la muestra (unidades de tratamiento y posibles unidades de control), lo cual da lugar a una propensión estimada a participar para cada persona o empresa, independientemente de si

*La propensión a participar como forma de definir el grupo tratado y el grupo de control*

realmente participan o no<sup>79</sup>. A continuación, se empareja a las personas o empresas tratadas con una o varias personas o empresas no tratadas, atendiendo a la puntuación de propensión<sup>80</sup>. Este procedimiento especifica un grupo de control que posteriormente puede utilizarse para obtener una estimación de la hipótesis contrafactual. Este emparejamiento garantiza que las estimaciones del impacto tengan en cuenta las diferencias observables entre el grupo tratado y las unidades de control y, suponiendo que todas las diferencias pertinentes antes del emparejamiento sean observables, permite obtener una estimación imparcial de los efectos de la intervención. Sin embargo, si la selección para poder acceder al tratamiento se basa en factores no observados, seguirá habiendo dudas sobre la adecuación del emparejamiento en cuanto a su capacidad para eliminar sesgos. La hipótesis crucial subyacente al enfoque de emparejamiento es que el proceso de selección se caracteriza por datos observables.

En el gráfico que figura a continuación se ilustra de manera intuitiva y simplificada el enfoque de emparejamiento por puntuaciones de propensión. El eje Y representa el número de personas en el grupo tratado y el grupo no tratado, ordenadas en función de la puntuación de propensión en el eje X. Normalmente, las personas tratadas suelen tener puntuaciones de propensión relativamente más altas, mientras que las personas no tratadas tienden a tener puntuaciones de propensión más bajas. La zona en la que se solapan las puntuaciones de propensión de los dos grupos se denomina «región de apoyo común».<sup>81</sup> Los casos tratados se emparejan con casos no tratados en esta región. El diagrama comprende dos ejemplos, pero el proceso se repite básicamente hasta que cada caso tratado se empareja con un caso no tratado dentro de la región de apoyo común. En el gráfico esto se hace utilizando el emparejamiento con «el vecino más cercano». El «vecino más cercano» de cualquier miembro del grupo de tratamiento es el miembro del grupo de control con la puntuación de propensión más similar a la suya. Una vez que se han formado dos grupos, sus resultados medios pueden compararse para obtener una estimación del impacto. En la práctica, el emparejamiento de las puntuaciones de propensión puede convertirse en un proceso sumamente complejo en el que hay que tener en cuenta una serie de cuestiones. Aquí se han ignorado muchas de estas cuestiones para que los principios fundamentales queden claros. En el recuadro siguiente se presenta un ejemplo práctico en el que se aplicó un enfoque de emparejamiento en una evaluación del FSE.

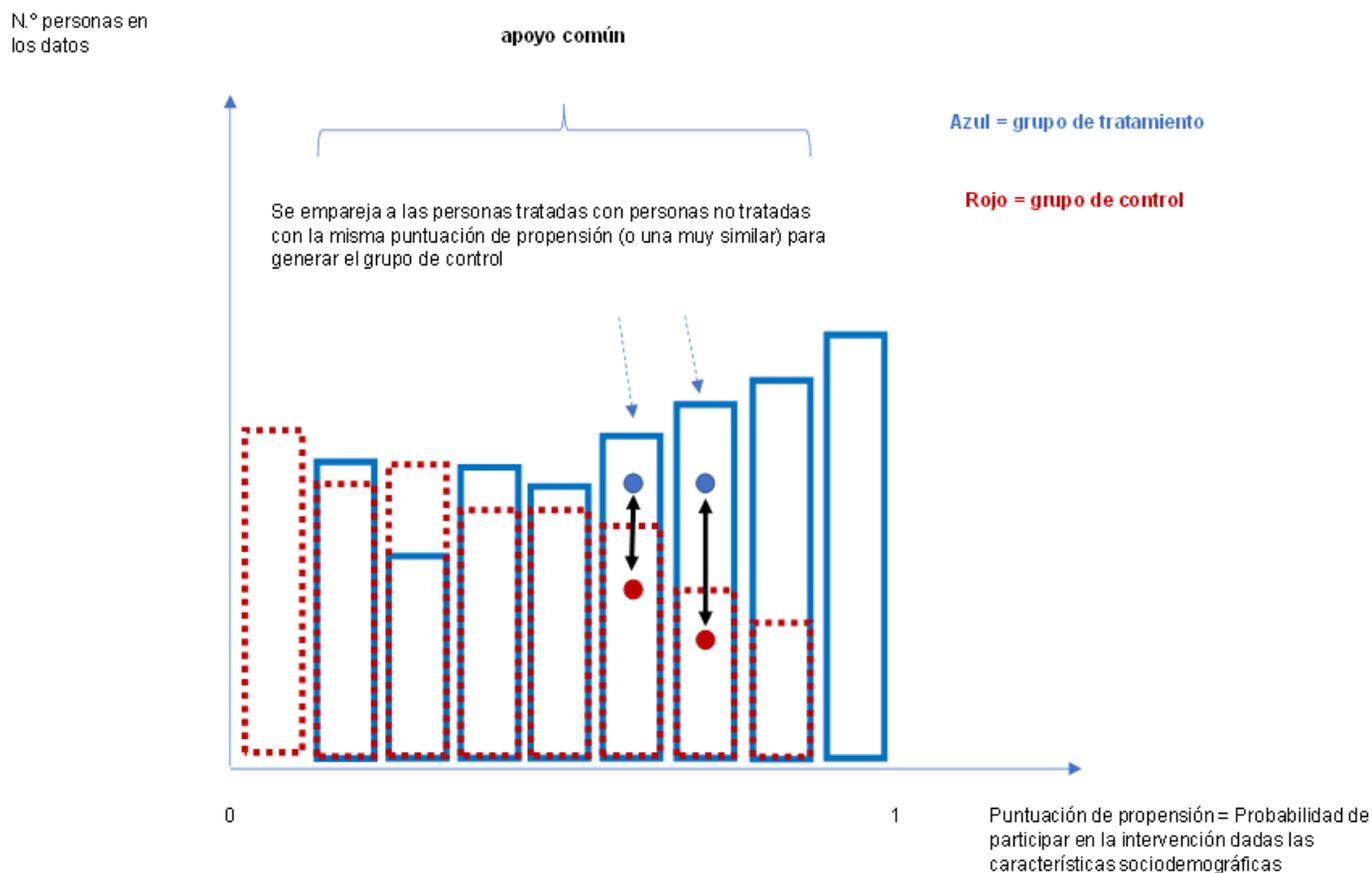
---

<sup>79</sup> Para simplificar este análisis, se asume que los responsables políticos quieren conocer el efecto del tratamiento en quienes sí recibieron los servicios del programa. En muchos casos, se trata de un subconjunto del grupo destinatario al que se ofreció dicha oportunidad. Esto se conoce como análisis del «tratamiento en los tratados».

<sup>80</sup> El emparejamiento por puntuaciones de propensión puede abordarse desde diversos enfoques. Para consultar un resumen, véase [Caliendo, M. y Kopeinig, S., 2008](#).

<sup>81</sup> El alcance de la región de apoyo común influye en el tamaño de la muestra y en la utilidad de los resultados para las políticas, en particular cuando una gran cantidad de casos tratados no entran en dicha región.

## Gráfico 8 Ilustración del enfoque de las puntuaciones de propensión



La viabilidad del enfoque de las puntuaciones de propensión se basa, entre otras, en la hipótesis de que la selección para el tratamiento se caracteriza plenamente por los datos observables. En otras palabras, la hipótesis es que no hay diferencias no observadas entre el grupo de tratamiento y el grupo de control que estén relacionadas con los resultados o con la decisión de participar en la intervención. La viabilidad de esta hipótesis se ve reforzada por la incorporación en la estimación de las puntuaciones de propensión de una amplia gama de variables, cuya selección se fundamenta en teorías y conocimientos previos. Concretamente, en el contexto de las intervenciones en el mercado de trabajo, se recomienda encarecidamente incorporar las anteriores vidas laborales individuales cuando se busquen posibles diferencias no observadas<sup>82</sup>.

*Selección basada en datos observables*

### Recuadro 15 Ejemplo de evaluación con un enfoque de emparejamiento<sup>83</sup>

#### Impacto de las intervenciones del FSE financiadas en 2014-2020 para parados de larga duración en la región de Las Marcas<sup>84</sup>

Se utilizó un enfoque de emparejamiento para evaluar el impacto de una serie de intervenciones del FSE (prácticas, becas de empleo, experiencia laboral en los municipios y vales de formación) en los parados de larga duración que se financiaron en la región italiana de Las Marcas. Las intervenciones no estaban específicamente dirigidas a los parados de larga duración, pero estos constituían la mayoría de los participantes en todas las intervenciones del FSE analizadas.

El objetivo de la evaluación era determinar el impacto de las medidas del FSE en la probabilidad de tener un empleo seis, nueve, doce, quince y dieciocho meses después del inicio de las intervenciones. Se

<sup>82</sup> Véanse [Caliendo, Mahlstedt y Mitnik, 2017](#), y [Kluve, Lehmann y Schmidt, 2008](#).

<sup>83</sup> Se pueden consultar otros ejemplos de la aplicación práctica de los distintos métodos (EPP, DRD y diferencia en diferencias) en [Comisión Europea, 2020](#).

<sup>84</sup> Véase [Pompili, M. y Giorgetti, I., 2020a](#).

utilizaron diversas variables de resultado: la probabilidad de tener un empleo en un determinado momento después de las intervenciones, la probabilidad de tener un empleo con un contrato indefinido y el número de días trabajados durante un período posterior a las intervenciones.

Se adoptó un enfoque de puntuaciones de propensión (emparejamiento con el vecino más cercano) para detectar las diferencias entre los grupos tratados y los no tratados, y las cuatro intervenciones del FSE mencionadas anteriormente se analizaron por separado. Para cada intervención, el grupo tratado estaba compuesto por participantes que empezaron en la intervención antes de finales de agosto de 2019 (526 para prácticas, 1 058 para becas de empleo, 236 para experiencia laboral en municipios y 241 para vales de formación). El grupo de control lo integraban desempleados que estaban inscritos en los SPE en el período 2016-2018 y llevaban al menos doce meses sin trabajar (77 255 registros).

Para calcular el EPP, se tuvieron en cuenta diversas variables derivadas de las características sociodemográficas de las personas tratadas y de las no tratadas, tales como su sexo, edad, ciudadanía, nivel educativo, lugar de residencia, fecha de entrada en la intervención y vida laboral treinta y seis meses antes de la intervención.

Para medir las variables de resultado de los grupos tratados y de control (antes y después de las intervenciones) se utilizaron datos del COB, el archivo de notificación obligatoria por parte de las empresas de los contratos laborales enviados a los SPE.

El estudio obtuvo resultados dispares: se constataron impactos positivos en el caso de las prácticas y las becas de empleo, mientras que se observaron efectos negativos en la experiencia laboral en los municipios y ningún efecto significativo para los vales de formación<sup>85</sup>.

### 3.2.3. Diferencia en diferencias

#### *Características fundamentales de la diferencia en diferencias*

- La diferencia en diferencias es un enfoque intuitivo que compara la diferencia de los efectos directos del grupo de tratamiento antes y después de la intervención con los efectos directos del grupo de control.
- Puesto que el cambio a lo largo del tiempo en el grupo de control mide qué habría ocurrido con el grupo de tratamiento en ausencia de intervención (la hipótesis contrafactual), cualquier diferencia adicional en los efectos directos del grupo tratado antes y después de la intervención muestra el efecto causal de esta.
- Se trata de un método sencillo que es viable en muchos casos.

Para realizar estimaciones de la diferencia en diferencias (DED), se pueden utilizar, por separado o en combinación con el emparejamiento, mediciones de referencia (o previas al tratamiento) de las variables de resultados. En este caso, la diferencia en un resultado antes y después del tratamiento en un grupo de control se deduce de la misma diferencia observada en un grupo tratado para obtener una estimación del impacto de una intervención. Una vez más, es esencial seleccionar un grupo de control plausible. Los impactos calculados sobre la base de la diferencia en diferencias suelen observarse dentro de un marco de regresión, que también tiene en cuenta otras diferencias observadas entre el grupo de tratamiento y el grupo de control. Además, este enfoque verifica las diferencias no observadas entre los dos grupos que permanecen fijas a lo largo del tiempo, así como las que varían con el tiempo, pero que afectan por igual al grupo de control y el grupo de tratamiento (por ejemplo, factores que afectan a la economía en su conjunto).

*Antes y después  
del tratamiento*

<sup>85</sup> Como se destaca en el informe de evaluación, en el caso de los vales de formación las conclusiones solo pueden considerarse preliminares, dado que la muestra tomada es demasiado pequeña en relación con el número total de participantes.

Debido a esta capacidad de comprobar algunos aspectos de las diferencias no observadas entre las unidades tratadas y las de control, un enfoque de la diferencia en diferencias representa, en la mayoría de los casos, una mejora con respecto a una estrategia de emparejamiento aplicada de manera transversal. En el gráfico 9 se presenta una representación visual del enfoque de la diferencia en diferencias.

El eje X representa el paso del tiempo y el eje Y, una escala en la que se registran los resultados. En este caso, los resultados podrían ser los salarios. Los salarios medios del grupo de tratamiento en el período anterior al tratamiento son YT1, mientras que para el grupo de control son YC1. En el período posterior al tratamiento, los salarios son YT2 e YC2 para el grupo de tratamiento y el grupo de control, respectivamente. Así pues, la línea superior continua representa el cambio en los salarios en el grupo de tratamiento, mientras que la línea inferior continua corresponde al grupo de control.

Podría obtenerse una estimación aproximada del impacto de la intervención mediante la comparación de los salarios del grupo de tratamiento y el grupo de control en el período posterior al tratamiento, es decir,  $YT2 - YC2$ . Sin embargo, esto sería incorrecto, ya que ignoraría las diferencias en los salarios antes del tratamiento. Se puede considerar que el estimador de la diferencia en diferencias consiste en restar una estimación previa al tratamiento del sesgo de la diferencia en los resultados después del tratamiento. Como consecuencia, la diferencia entre los salarios tras el tratamiento ( $YT2 - YC2$ ) se ajusta al restar de ella la diferencia entre los salarios antes del tratamiento ( $YT1 - YC1$ ) y, por tanto, el estimador del impacto de la diferencia en diferencias puede expresarse, de una forma muy simple, del siguiente modo:

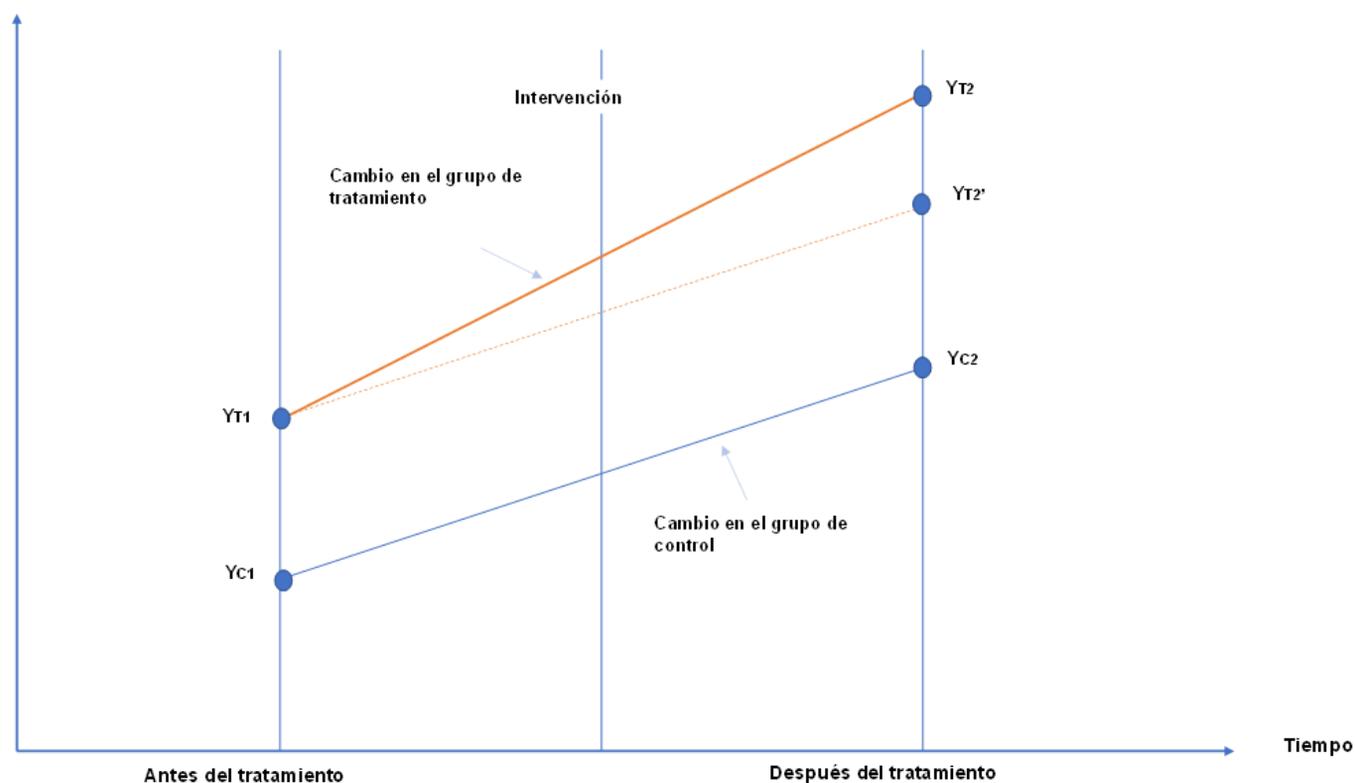
$$(YT2 - YC2) - (YT1 - YC1).$$

*La lógica de la  
doble diferencia*

Si las diferencias en los salarios después del tratamiento no se ajustan para tener en cuenta las diferencias que ya existían entre el grupo de tratamiento y el grupo de control, podrían obtenerse unas estimaciones sesgadas. De manera alternativa, como se ha mencionado anteriormente, puede considerarse que el enfoque de la diferencia en diferencias resta el cambio en los resultados entre el grupo de control de ese cambio observado en el grupo de tratamiento. Se considera que el cambio observado en el grupo de control es el que se habría producido en el grupo de tratamiento en ausencia de la intervención.

## Gráfico 9 Ilustración del enfoque de la diferencia en diferencias

Resultado (Y)



En el caso más sencillo, la hipótesis principal en la que se fundamenta el enfoque de la diferencia en diferencias es la de las tendencias comunes: es decir, en ausencia de tratamiento, las tendencias temporales en los resultados del grupo de tratamiento y el grupo de control son equivalentes. Esta hipótesis no puede comprobarse directamente, aunque respalda la plausibilidad del enfoque en los casos en que se dispone de múltiples mediciones de la variable de resultados previas al tratamiento tanto para el grupo de tratamiento como para el grupo de control, y estas mediciones muestran tendencias paralelas. En el siguiente recuadro se presenta un ejemplo del enfoque de la diferencia en diferencias.

*La hipótesis principal: las tendencias comunes son equivalentes en ausencia de tratamiento.*

### Recuadro 16 Ejemplo de una evaluación con un enfoque de diferencia en diferencias

#### Evaluación de las medidas dirigidas a los trabajadores de más edad en Lubelskie financiadas por el FSE en el período de programación 2007-2013<sup>86</sup>

El objetivo principal del estudio era evaluar las intervenciones del FSE para ampliar la actividad económica de las personas mayores en el contexto de una situación demográfica desfavorable en la región polaca de Lubelskie. Las intervenciones se financiaron en el marco del Programa Operativo de Capital Humano para 2007-2013.

El grupo tratado, una muestra de 145 personas, estaba formado por personas desempleadas en el momento de acceder a las intervenciones. El grupo de control se extrajo de los datos anonimizados del registro facilitados por las oficinas de empleo, que contienen datos sociodemográficos y el historial de eventos registrales relacionados con la situación laboral de determinadas personas (los datos se obtuvieron de ocho oficinas de empleo y contenían información sobre 67 102 personas). La variable de resultado era la «situación de estar inscrito en el registro de desempleo»<sup>87</sup>, lo que pudo observarse tanto antes como

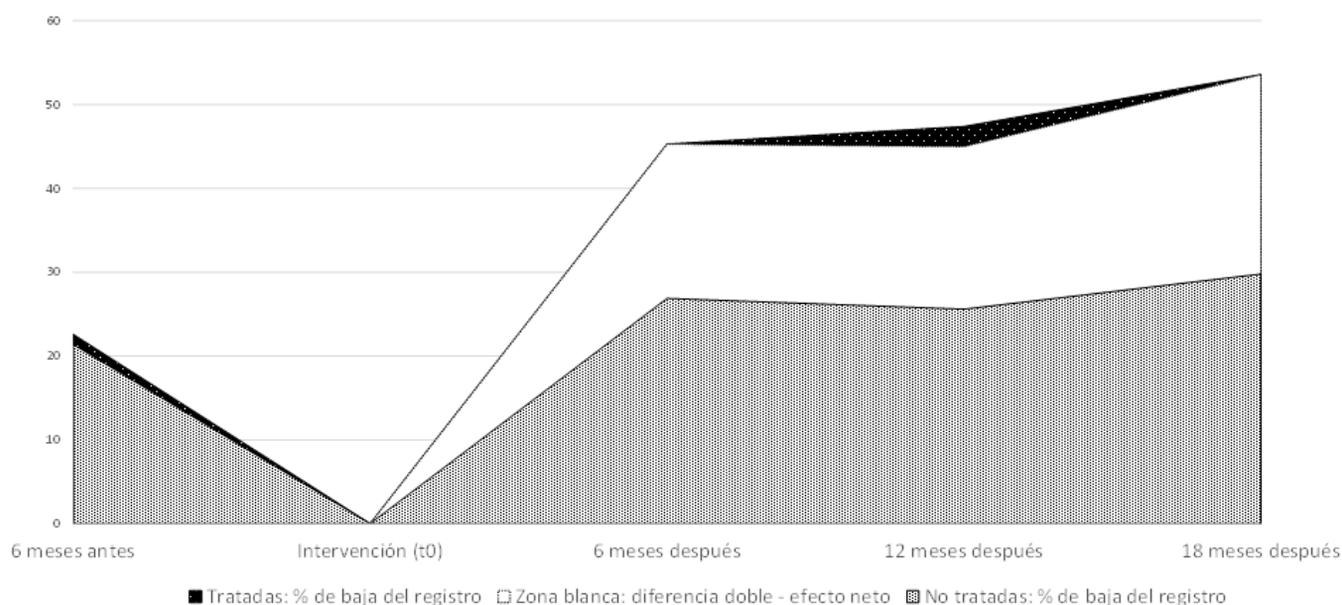
<sup>86</sup> Véase [Re-source Pracownia Badań i Doradztwa, 2015](#).

<sup>87</sup> Aunque no estar inscrito como «desempleado» no es exactamente lo mismo que estar «empleado», era necesario hacer esa

después de la intervención, lo que permitió la aplicación de un enfoque de diferencia en diferencias. La situación laboral se midió en cinco momentos: seis meses antes de la participación en el proyecto, durante dicha participación, y seis, doce y dieciocho meses después de la finalización del proyecto.

En este enfoque de diferencia en diferencias, se mejoró la elección del grupo de control utilizando métodos de emparejamiento para encontrar a una persona no tratada para cada una de las tratadas, sobre la base de una serie de covariantes observables como el género, la edad, el nivel educativo y la tasa de desempleo en el lugar de residencia.

El análisis mostró que el apoyo a los desempleados mayores de cincuenta años produjo un efecto neto positivo en el empleo: los desempleados de más de cincuenta años tenían una probabilidad más de 2,5 veces mayor de darse de baja como desempleados que una persona no tratada. En el gráfico que figura a continuación se muestra la evolución de la variable de resultado para el grupo tratado y el grupo de control y el efecto neto estimado tras considerar la hipótesis paralela del enfoque de la diferencia en diferencias.



### 3.2.4. Diseño de regresión discontinua

#### Características fundamentales de la regresión discontinua

- Un diseño de regresión discontinua se basa en la idea de que un valor específico («punto de corte») de una puntuación o calificación determina si una persona participa o no en la intervención.
- Las personas cercanas a ese valor se consideran comparables, con la única diferencia de que solo las de un lado del punto de corte participan en la intervención (el grupo de tratamiento) y las del otro lado, no (el grupo de control).
- La comparación de estos grupos suele proporcionar una medición precisa e intuitiva del efecto de la intervención, pero el método solo es aplicable si existe un proceso de selección basado en una puntuación o calificación

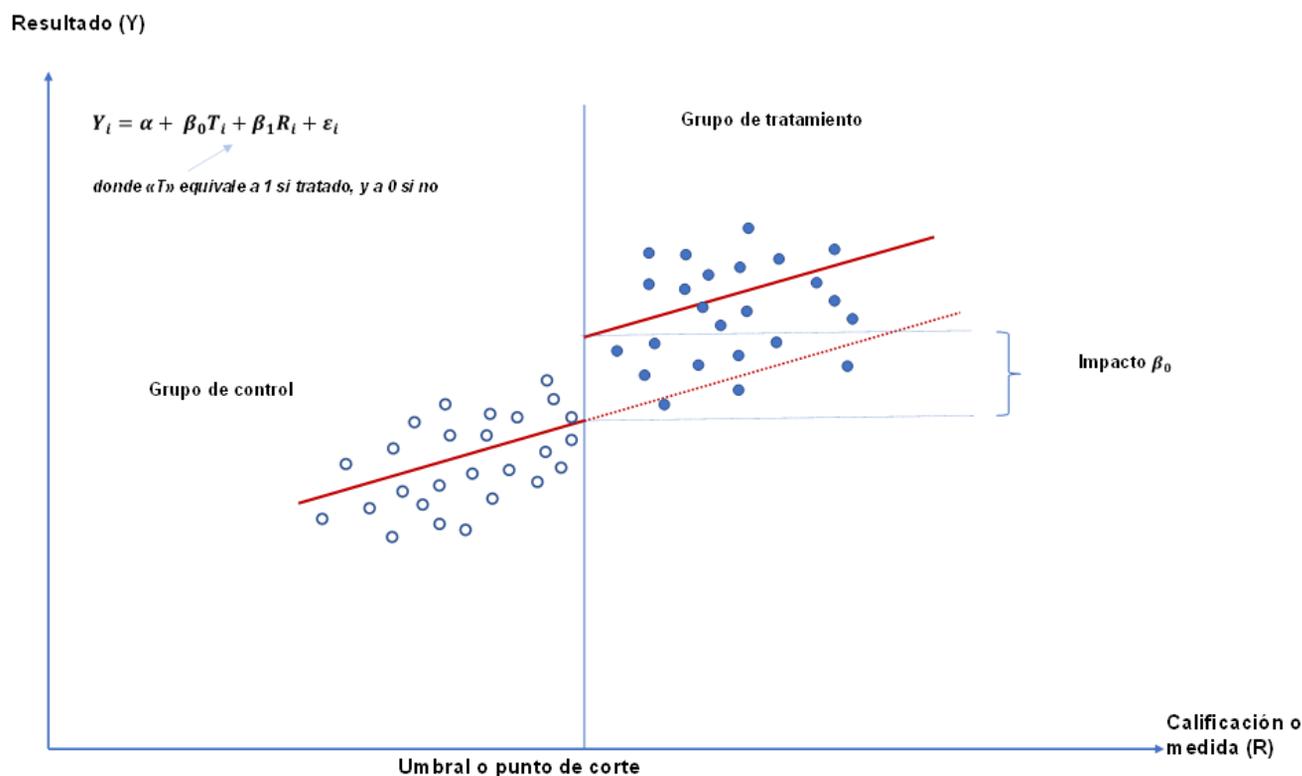
Podrá adoptarse un enfoque de **regresión discontinua** cuando el acceso a una intervención venga determinado por un punto de corte a lo largo de una calificación, escala o medida continua. Por ejemplo, el acceso a una formación podría estar supeditado a los resultados obtenidos en una prueba de aptitud: quienes obtengan una puntuación por encima de un determinado umbral (o corte) recibirán la formación, mientras que quienes obtengan una

*Un punto de corte divide a las personas tratadas y las personas de control*

suposición para poder llevar a cabo el estudio.

puntuación inferior al umbral no la recibirán. Para que el enfoque sea válido, el punto de corte debe determinarse sin conocer las puntuaciones de las personas que optan a la formación. Los candidatos que se encuentren muy cerca de la línea de corte serán muy similares entre sí, pero los que se encuentren justo por encima de ella participarán en la intervención y los que estén justo debajo, no. Los resultados de quienes se encuentren encima y debajo del corte pueden compararse para obtener una estimación del impacto de la intervención en el punto de corte.

### Gráfico 10 Ilustración del enfoque de la regresión discontinua



Se puede aplicar un diseño de regresión discontinua (DRD), o bien cuando el punto de corte determina por completo al grupo de tratamiento (con adhesión plena), en cuyo caso se obtiene una discontinuidad *nítida*, o bien cuando, en determinadas condiciones, no todas las personas a uno de los lados del punto de corte se adhieren estrictamente al tratamiento asignado (una discontinuidad *difusa*).

*Discontinuidad  
nítida o difusa*

En el gráfico 10 se presenta un ejemplo tipo de un diseño de regresión discontinua. Este es el ejemplo más sencillo de una discontinuidad nítida. La intervención produce efectos constantes en cada valor de la calificación y los impactos se calculan utilizando un modelo de regresión lineal (no hay problemas relativos a la forma funcional de la regresión del impacto). En la práctica, el análisis tendrá que ser siempre mucho más sofisticado que el presentado en el gráfico 10.

Los puntos del gráfico 10 representan unidades individuales; por ejemplo, personas que optan a una formación. El eje X registra la calificación o medida utilizada para asignar a los participantes vacantes en el curso de formación. Las personas con una puntuación a la derecha de la línea vertical continua que indica el umbral de esta calificación o medida (por ejemplo, una prueba de aptitud) entran en la formación e integran el «grupo de tratamiento». Los

candidatos que obtengan una puntuación inferior al umbral en la calificación o medida no acceden a la formación y conforman el grupo de control.

La clave es que la calificación utilizada para asignar a los destinatarios al grupo de tratamiento o al grupo de control es una variable cuantitativa continua medida antes del tratamiento, y una persona es admitida en el programa de formación en función de si su puntuación es superior o inferior a un punto de corte o umbral predefinido.

El resultado se representa en el eje Y. Esencialmente, el impacto del tratamiento se determina mediante la estimación de un modelo de regresión lineal sobre los datos (dadas las hipótesis anteriores), es decir, haciendo la regresión de la variable de resultados frente a la medición de la calificación junto con una variable ficticia (un indicador del tratamiento) que define si una puntuación es inferior o superior al punto de corte (es decir, si se asigna a la persona al grupo de tratamiento o al grupo de control).

Esta ecuación de regresión del impacto se ilustra en el gráfico 10. En nuestro ejemplo, el efecto o impacto de la formación se obtiene a partir del coeficiente del indicador del tratamiento, es decir,  $\beta_0$ <sup>88</sup>. Esto muestra de manera efectiva si hay una ruptura o discontinuidad en torno al punto de corte, que en el gráfico 10 se ha indicado colocando la línea de regresión por encima del umbral o punto de corte. En este ejemplo, se observa que la formación tuvo un impacto positivo en el resultado.

Otra forma de entender la estimación del impacto es considerar la línea de puntos que prolonga la línea del grupo de control representada en el gráfico 10. Esto puede interpretarse como una estimación contrafactual para el grupo de tratamiento —la relación entre la calificación y la medición del resultado que habría prevalecido de no haberse intervenido—; la diferencia entre esta línea de puntos y la línea de tendencia del grupo de tratamiento representa el efecto o impacto del tratamiento. Téngase en cuenta que, en ausencia de tratamiento, no hay discontinuidad en la línea y se asume que, en tal caso, el resultado varía continuamente con la calificación o medida.

El enfoque de regresión discontinua funciona porque las observaciones en los grupos de tratamiento y de control próximos al punto de corte son similares salvo por el hecho de que, en este ejemplo, quienes se encuentran por encima del punto de corte reciben formación y quienes están por debajo, no. Por lo tanto, la situación de las observaciones cercanas al punto de corte no es diferente de la aleatorización. Sin embargo, existe una limitación considerable. En la mayoría de las aplicaciones, los impactos estimados utilizando un enfoque de DRD solo pueden informar al responsable político sobre los efectos para las personas cercanas al punto de corte o umbral. El grado en que pueden hacerse generalizaciones sobre las más alejadas del umbral puede ser limitado.

*Buena validez interna, pero la generalización podría ser limitada*

El DRD puede ser un enfoque útil cuando se asigne a las personas a una intervención en función de la necesidad, medida a partir de una calificación o puntuación continua. No obstante, el análisis puede complicarse cuando el punto de corte es difuso y hay falta de adhesión, y cuando existen problemas de forma funcional en el modelo de regresión del impacto. En la práctica, es necesario recurrir a una serie de hipótesis cuya validez no siempre puede confirmarse.

<sup>88</sup> En un caso simple, este sería el efecto de la intención de tratar en el punto de corte (véase [Bloom, H. S., 2009](#)).

El siguiente recuadro muestra un ejemplo práctico en el que se utilizó un enfoque de regresión discontinua en una evaluación de un programa de Garantía Juvenil.

#### Recuadro 17 Ejemplo de una evaluación con un enfoque de regresión discontinua

***Vocational Training for Unemployed Youth in Latvia: Evidence from a Regression discontinuity design*** [«Formación profesional para jóvenes desempleados en Letonia: pruebas procedentes de un diseño de regresión discontinua», documento en inglés]<sup>89</sup>

Los investigadores del Centro Común de Investigación de la Comisión Europea utilizaron un enfoque de regresión discontinua para evaluar el impacto de un programa de formación profesional en los efectos directos en el mercado laboral de los jóvenes desempleados en Letonia. Dicho programa se financió en el marco del sistema Garantía Juvenil para el período 2014-2020 y estaba dirigido a jóvenes de entre quince y veintinueve años que ni estudiaban, ni trabajaban, ni recibían formación (ninis).

Los datos utilizados se obtuvieron de la Agencia Estatal de Empleo de Letonia, que facilitó información sobre los participantes y los no participantes inscritos como desempleados en fechas concretas, y se cotejaron con los datos de la Administración Tributaria del Estado, que posee información sobre los ingresos de cada persona en fechas concretas anteriores y posteriores al programa (enero de 2012 y junio de 2017). Tras depurar los conjuntos de datos, la muestra final del grupo tratado ascendía a 898 personas y la del grupo de control, a 10 717.

Los evaluadores tomaron como fundamento una norma del Gobierno letón que prioriza la participación de los desempleados menores de veinticinco años en el programa de formación profesional. Por lo tanto, la edad es la variable de asignación que determina la probabilidad de participar en el programa de formación profesional, y veinticinco años es el umbral por debajo del cual la participación en el programa aumenta drásticamente debido a la norma de prioridad. Puesto que las personas no tienen ningún control sobre su edad, puede considerarse que la asignación al programa de formación profesional es aleatoria alrededor del umbral. La norma de prioridad y el hecho de que la participación fuera voluntaria obligaron a los investigadores a aplicar un diseño de regresión discontinua difusa. La estrategia de identificación se basó en el hecho de que las personas cercanas al punto de corte (veinticinco años) eran similares entre sí salvo por el hecho de participar o no en el programa de formación profesional.

Las conclusiones mostraron que el impacto en el empleo probable en el futuro y en los ingresos mensuales hasta tres años y medio después de acceder al programa es positivo, pero irrelevante desde el punto de vista estadístico. Pese a ello, se observó que la norma de prioridad tuvo un efecto positivo en la participación en el programa. Dado que la validez del diseño de regresión discontinua difusa se basa en el hecho de que los participantes potenciales no pueden controlar la variable de asignación (edad), se realizaron varias pruebas para confirmar este efecto.

<sup>89</sup> Véase [Bratti, M. et al., 2018](#).

### 3.2.5. Variables instrumentales

#### Características fundamentales de las variables instrumentales

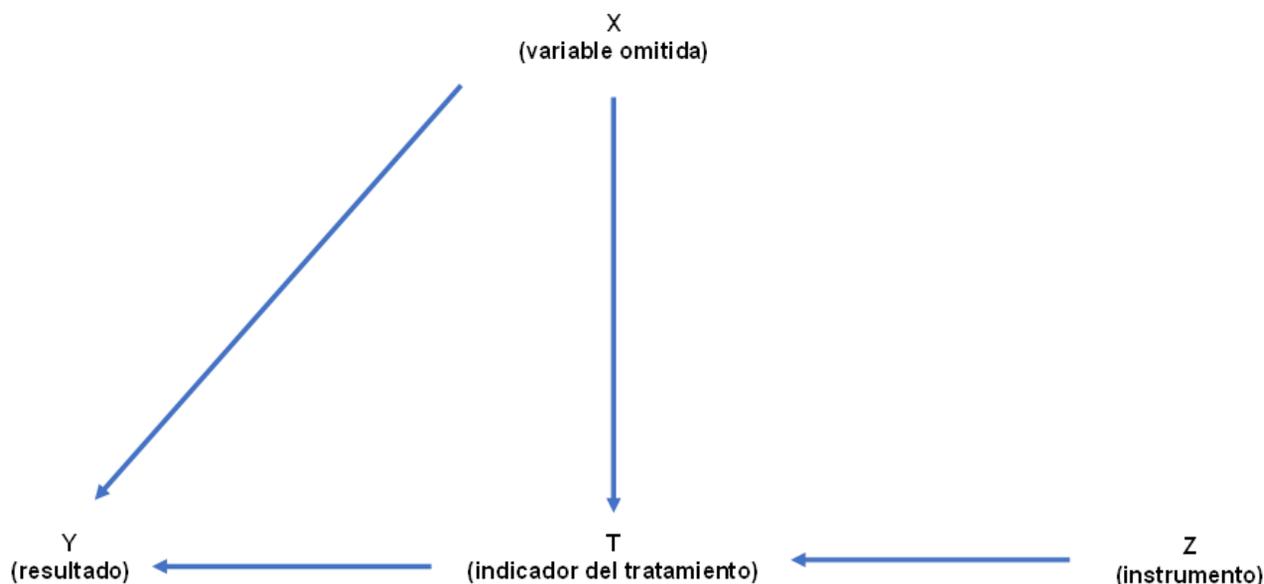
- La idea de las variables instrumentales es que algún proceso (predeterminado/exógeno) determina la participación, pero no es el procedimiento de selección real.
- El proceso predeterminado y sus principales variables pueden utilizarse para tener en cuenta cualquier autoselección en la decisión de las personas de participar o no.

Para el enfoque de las **variables instrumentales** (VI), la selección para acceder al tratamiento debe estar determinada, al menos en parte, por un factor exógeno que no guarde relación con resultados distintos del tratamiento. Como tal, el factor exógeno influye en la participación, pero no directamente en los resultados. Por lo general, estos factores exógenos pueden ser omisiones o errores administrativos u otras variaciones aleatorias en el tratamiento recibido.

*Cómo afectan los factores exógenos a la participación*

En el gráfico 11 se ilustra el enfoque de variables instrumentales. Cuatro variables se muestran en un sistema causal muy simplificado. Las variables representan datos recogidos de una población a la que hipotéticamente se dirige una intervención de formación (tanto las personas que reciben la formación como las que funcionan como control).

#### Gráfico 11 Ilustración de un enfoque de variables instrumentales



«Y» representa el resultado considerado. En el caso de una intervención de formación, podrían ser los ingresos. «T» es un indicador que revela si una persona ha asistido a una formación<sup>90</sup>.

«X» es una variable omitida que no se observa, pero que está relacionada tanto con «Y» (el resultado) como con «T» (el indicador de tratamiento) y amplía la idea de un programa de formación; puede ser una medida de referencia de la capacidad, por ejemplo. En este caso, la capacidad guarda relación tanto con la participación en la formación como con los ingresos. Por

<sup>90</sup> En otras palabras, la adhesión es plena, y todas las personas que integran el grupo de tratamiento participan en la formación.

ejemplo, los miembros más capacitados del grupo destinatario podrían optar por recibir la formación y percibir salarios más elevados.

La existencia de «X» motiva la búsqueda de un instrumento y hace que el impacto de la formación en los ingresos se vea frustrada por su existencia. En otras palabras, la estimación está sesgada debido a la existencia de X y al hecho de que es un factor que no se ha observado ni puede contabilizarse directamente en el análisis.

Por último, la variable «Z» es un instrumento que, según Morgan y Winship<sup>91</sup>, puede considerarse una alteración de «T» que es independiente de «X». Por ese motivo, en el gráfico 11 no hay ninguna línea que una Z con X. Además, Z solo afecta a Y a través de T; no existe ninguna otra vía por la que Z afecte a Y. Esto significa que Z puede usarse para generar una variación en T (el tratamiento) que no guarde ninguna relación con la variable de confusión X. En consecuencia, con esta variación se puede obtener una medición imparcial del efecto de T en Y<sup>92</sup>.

*Es fundamental especificar un instrumento válido*

A continuación se describen las circunstancias más simples en las que podría adoptarse un enfoque de VI, dejando de lado muchas de las complejidades que conlleva. En la práctica, a menudo resulta difícil encontrar un instrumento convincente. La plausibilidad de los distintos instrumentos potenciales está estrechamente ligada al contexto y, por lo general, no es posible comprobar estadísticamente las hipótesis de identificación subyacentes (en realidad, la correlación requerida entre Z y T puede comprobarse estadísticamente, pero la «independencia» de Z y X no). Por ejemplo, una estrategia podría consistir en utilizar la distancia entre los centros donde se imparte la formación (la ubicación física del curso de formación) y el domicilio de los participantes como instrumento para estimar el efecto de la formación en los ingresos netos de las personas que asisten a ella. Tal vez se observe que las personas que viven más cerca de los centros de formación tienen más probabilidades de participar en una intervención de formación y que la distancia entre el domicilio y el centro de formación no guarda relación con otros factores determinantes de los ingresos netos y de la participación en la formación (por ejemplo, mediciones relativas al capital humano). Por lo tanto, la única vía por la que esta medición de la distancia podría afectar a los ingresos netos es a través de su efecto en la formación<sup>93</sup>.

Las variables instrumentales pueden utilizarse en una gran variedad de contextos. Las estimaciones pueden obtenerse mediante diversos enfoques de estimación, en función de la variable de respuesta. Hasta ahora, este enfoque se ha utilizado poco en las evaluaciones del FSE. En el recuadro 18 se muestra un ejemplo del análisis de los efectos causales de una política mediante un enfoque de VI.

<sup>91</sup> Morgan, S.L. y Winship, C., 2014.

<sup>92</sup> El efecto causal de T en Y se calcula en presencia de un instrumento mediante una estimación de la relación entre Z e Y, tras lo cual dicha estimación se divide por la relación estimada entre Z y T.

<sup>93</sup> La interpretación de las conclusiones de tal análisis podría complicarse por el hecho de si el instrumento está relacionado con la variación en los efectos del tratamiento (véase [Bryson et al., 2002](#)).

**Recuadro 18 Ejemplo de un estudio mediante un enfoque de variables instrumentales****Efectos en el empleo de la formación lingüística para inmigrantes desempleados<sup>94</sup>**

La adquisición de competencias en la lengua del país de acogida es un factor importante para que los inmigrantes logren un alto nivel de integración en el mercado laboral.

En un artículo publicado recientemente (Lang, J., 2021), la autora utiliza un enfoque de variables instrumentales para estimar el efecto causal de un programa de formación en lengua alemana con fines profesionales sobre los resultados en materia de empleo de los inmigrantes que participan en los cursos dos años después de su matriculación. El programa lo llevó a cabo la Oficina Federal Alemana de Migración y Refugiados (BAMF, por sus siglas en alemán) con financiación del FSE, y estuvo operativo hasta finales de 2017. Los cursos tenían varios componentes: formación en lengua alemana, desarrollo de capacidades profesionales y prácticas laborales. Eran elegibles las personas de origen migrante y activas en el mercado laboral alemán, independientemente de su nacionalidad y la fecha en que inmigraron.

Para abordar la heterogeneidad no observada en términos de capacidades lingüísticas, la autora se sirve de la variación exógena en la intensidad de la formación en la lengua local a nivel de centro de trabajo. Esto se considera un instrumento válido, ya que los centros de empleo tienen un amplio margen de apreciación en la aplicación de diferentes combinaciones de programas en función de sus «estilos políticos», y esta variable es exógena a los efectos directos en materia de empleo de los demandantes de empleo. Se analizaron los cursos impartidos en 2014.

La evaluación se basa en un extenso conjunto de datos administrativos, las biografías laborales integradas (BLI), una base de datos fusionada de información administrativa de la Agencia Federal de Trabajo alemana. Las BLI incluyen información sobre la participación en la formación lingüística, información detallada sobre la vida laboral (excepto sobre el trabajo por cuenta propia), las búsquedas de empleo, los pagos realizados durante el período de desempleo y los datos personales de las personas. También se utilizaron otros datos (WGH, Werdegangshistorie) para disponer de información detallada sobre la participación en los cursos de alemán impartidos por la BAMF y para subsanar, en la medida de lo posible, las lagunas de las BLI con información sobre el trabajo por cuenta propia y los episodios de permiso parental.

Las conclusiones muestran que, tras un período de efectos de bloqueo con una duración general algo superior a un año, dos años después de la intervención los inmigrantes que participan en el programa de formación tienen un 9 % más de probabilidades de tener trabajo que el grupo de control, y esta probabilidad no se limita a empleos no cualificados. También se constató que impartir formación lingüística de forma temprana (poco después de llegar a Alemania) es beneficioso para la integración de los inmigrantes en el mercado laboral.

<sup>94</sup> Véase [Lang, J., 2021](#).

**Cuadro 5 Comparación de las características fundamentales de los principales enfoques de EIC**

Enfoque	Características fundamentales	Ventajas	Requisitos en materia de datos	Limitaciones
Aleatorización: enfoque experimental	<ul style="list-style-type: none"> <li>- A las personas que reúnen los requisitos para participar se las asigna aleatoriamente a un grupo de tratamiento o un grupo de control</li> <li>- La aleatorización garantiza que ambos grupos sean idénticos (en general) en todas las características pertinentes</li> <li>- Por lo tanto, el grupo de control responde a la pregunta contrafactual y la diferencia de efectos directos entre el grupo de tratamiento y el grupo de control indica el efecto causal de la intervención</li> </ul>	<ul style="list-style-type: none"> <li>- Si se aplica correctamente, las estimaciones del impacto son «imparciales»</li> <li>- Los resultados son transparentes y fáciles de entender</li> <li>- Las conclusiones son menos susceptibles a matizaciones y dudas</li> <li>- Hay múltiples variantes de diseño para lidiar con diversos contextos políticos y circunstancias de intervención</li> </ul>	<ul style="list-style-type: none"> <li>- Requisito básico para seleccionar el grupo de control en la intervención mediante aleatorización</li> <li>- Un registro de a quién se ha asignado a cada grupo</li> <li>- La recopilación de datos de referencia es esencial</li> <li>- Deben registrarse mediciones de los resultados obtenidos por el grupo de tratamiento y el grupo de control</li> </ul>	<ul style="list-style-type: none"> <li>- A menudo, el grupo tratado y el grupo de control no cumplen las normas de asignación al tratamiento</li> <li>- Se suele necesitar el consentimiento de los participantes</li> <li>- La aleatorización puede influir en la selección de las personas que participarán o solicitan participar en una intervención</li> <li>- Ser conscientes de a qué grupo se les ha asignado podría alterar el comportamiento de los participantes e influir en los resultados</li> <li>- Preocupaciones éticas</li> <li>- Requisitos considerables en materia de planificación y diseño</li> <li>- Puede ser costoso (aunque no necesariamente)</li> </ul>
Emparejamiento (por puntuaciones de propensión)	<ul style="list-style-type: none"> <li>- El emparejamiento imita a la aleatorización en tanto en cuanto crea <i>ex post</i> un grupo de control que se asemeja lo máximo posible (en general) al grupo de tratamiento en todas las características pertinentes</li> <li>- A diferencia de la aleatorización, solo pueden emparejarse las características observables (edad, género, nivel educativo, etc.), mientras que no es posible tener en cuenta aquellas que no pueden observarse (p. ej., la motivación)</li> <li>- Por tanto, la validez del enfoque depende significativamente de la disponibilidad de los datos</li> </ul>	<ul style="list-style-type: none"> <li>- Exige conocer en profundidad los procesos de selección, pero no requiere tener un control directo de la selección para participar en la intervención</li> <li>- Puede aplicarse retroactivamente, si se dispone de los datos correctos y en diversos contextos</li> <li>- Técnicamente es un método de estimación semiparamétrico; requiere menos hipótesis paramétricas (por ejemplo, no es necesaria una hipótesis de regresión estándar)</li> <li>- Puede utilizarse para estimar múltiples efectos de un tratamiento</li> </ul>	<ul style="list-style-type: none"> <li>- Identificación precisa de los participantes en la intervención</li> <li>- Fuentes de datos a partir de las cuales extraer la muestra</li> <li>- Concepto claro de participación y buena comprensión de la selección para acceder al tratamiento</li> <li>- Datos variados, que idealmente se recabarán de una base de referencia de la que se seleccionarán las unidades de emparejamiento</li> <li>- Medición de los resultados de la intervención para los participantes y las unidades de control</li> </ul>	<ul style="list-style-type: none"> <li>- Requiere una gran cantidad de datos que permitan una caracterización completa del proceso de selección</li> <li>- La validez depende de la calidad de las unidades de control, de su cuidadosa selección y del grado de apoyo común</li> <li>- Se basa en el supuesto de que la selección para acceder a la intervención puede caracterizarse adecuadamente mediante datos observables</li> <li>- Es preciso analizar con cuidado los distintos enfoques de emparejamiento disponibles</li> <li>- Los resultados pueden ser complejos</li> </ul>

Enfoque	Características fundamentales	Ventajas	Requisitos en materia de datos	Limitaciones
				de explicar e interpretar, y potencialmente ambiguos
Diferencia en diferencias (DED)	<ul style="list-style-type: none"> <li>- La DED es un enfoque intuitivo que compara la diferencia de los efectos directos en el grupo de tratamiento antes y después de la intervención con la diferencia en el grupo de control antes y después de la intervención</li> <li>- Puesto que el cambio a lo largo del tiempo en el grupo de control mide qué le habría ocurrido al grupo de tratamiento en ausencia de la intervención (la hipótesis contrafactual), cualquier diferencia adicional en los efectos directos del grupo tratado antes y después de la intervención muestra el efecto causal de esta</li> <li>- Se trata de un método sencillo que en la práctica es aplicable en muchos casos</li> </ul>	<ul style="list-style-type: none"> <li>- Comprueba algunos aspectos de diferencias no observadas entre los participantes y las unidades de control</li> <li>- Puede utilizarse en combinación con el emparejamiento</li> <li>- Funciona con datos de antes y después de la intervención, como datos de panel (datos individuales a lo largo del tiempo) o datos transversales repetidos (datos sobre personas recogidos en diferentes momentos)</li> </ul>	<ul style="list-style-type: none"> <li>- Los requisitos en materia de datos son similares a los de otros enfoques, pero con el requisito adicional de medir los resultados antes de la intervención</li> <li>- Para comprobar las principales hipótesis, antes del tratamiento es necesario hacer múltiples observaciones de los resultados para el grupo de tratamiento y el grupo de control</li> </ul>	<ul style="list-style-type: none"> <li>- Requiere invocar la hipótesis de que existen tendencias comunes en los resultados de los participantes y las unidades de control</li> <li>- El análisis puede llegar a ser bastante complejo y prestarse a interpretaciones erróneas</li> <li>- Se precisan datos variados anteriores al tratamiento sobre los resultados para comprobar la hipótesis de las tendencias comunes</li> <li>- No puede utilizarse para estimar múltiples efectos de un tratamiento<sup>95</sup></li> </ul>
Diseño de regresión discontinua (DRD)	<ul style="list-style-type: none"> <li>- El DRD se basa en la idea de que un valor específico («punto de corte») de una puntuación o calificación determina si una persona participa o no en la intervención</li> <li>- Las personas cercanas a ese valor se consideran comparables, con la única diferencia de que solo las de un lado del punto corte participan en la intervención (el grupo de tratamiento) y las del otro lado, no (el grupo de control)</li> <li>- La comparación de estos grupos suele dar una medición precisa e intuitiva del efecto de la intervención, pero el método solo es aplicable si existe un proceso de selección basado en una puntuación o calificación</li> </ul>	<ul style="list-style-type: none"> <li>- Existen enfoques de DRD tanto nítidos como difusos</li> <li>- Puede producir estimaciones imparciales de los efectos del tratamiento en determinadas condiciones</li> </ul>	<ul style="list-style-type: none"> <li>- La elección del punto de corte debe ser independiente de los valores de la calificación dada a cada miembro de un grupo destinatario</li> <li>- Se requieren datos sobre las personas que integran el grupo de tratamiento y el grupo de control en términos de su calificación o medición, el umbral o punto corte, y los resultados</li> </ul>	<ul style="list-style-type: none"> <li>- Este enfoque no es válido sin una medida o calificación continua que determine el tratamiento</li> <li>- Los análisis pueden llegar a ser complejos e inciertos cuando las cuestiones relativas a la forma funcional de la regresión del impacto se hacen patentes, cuando no hay adhesión o cuando el tamaño de la muestra en torno al punto de corte es limitado</li> <li>- Puede haber peligros a la hora de interpretar las conclusiones y extrapolar generalizaciones</li> </ul>
Variables instrumentales (VI)	<ul style="list-style-type: none"> <li>- La idea de las VI es que existe algún proceso (predeterminado/exógeno) que determina la participación, pero no es el proceso de</li> </ul>	<ul style="list-style-type: none"> <li>- Puede proporcionar estimaciones de gran calidad de los efectos causales</li> </ul>	<ul style="list-style-type: none"> <li>- Necesita datos de referencia y datos sobre los resultados y la participación en la intervención, así como que se</li> </ul>	<ul style="list-style-type: none"> <li>- Puede resultar difícil encontrar un instrumento plausible</li> </ul>

<sup>95</sup> Véase [Frolich, M., 2004.](#)

Enfoque	Características fundamentales	Ventajas	Requisitos en materia de datos	Limitaciones
	<p>selección real</p> <ul style="list-style-type: none"> <li>- Luego, ese proceso predeterminado puede utilizarse para tener en cuenta cualquier autoselección en la decisión de las personas de participar o no</li> </ul>	<p>o pruebas de la existencia de estos</p> <ul style="list-style-type: none"> <li>- Resuelve el problema del sesgo de la variable omitida (o sesgo de selección)</li> <li>- Puede aplicarse retroactivamente</li> </ul>	<p>pueda especificar un instrumento</p> <ul style="list-style-type: none"> <li>- Un instrumento debe estar relacionado con la participación en la intervención y solo afecta a los resultados sobre esta base El instrumento no debe estar relacionado con ningún otro factor determinante de los resultados</li> </ul>	<ul style="list-style-type: none"> <li>- Puede ser difícil de explicar a personas que no sean expertas</li> <li>- La interpretación de los resultados no es sencilla; existen limitaciones para comprobar las hipótesis de identificación</li> </ul>

# Capítulo 4. Impulsar el programa de EIC

La presente Guía tiene por objeto animar a las AG a que lleven a cabo EIC más generalizadas y de calidad, así como apoyarlas en su realización. Proporciona orientación a los responsables de planificar y encargar evaluaciones del impacto de las intervenciones cofinanciadas por el FSE+. Hasta ahora, la atención se ha centrado en la planificación y realización de una EIC, y se han analizado diversas cuestiones clave que es necesario valorar. Sin embargo, el período de programación 2014-2020 puso de relieve una serie de «cuestiones más amplias» y dificultades a las que hay que hacer frente para reforzar las evaluaciones del FSE+ y el uso de las EIC.

Esta sección de la Guía contiene algunas sugerencias para abordar estas cuestiones «más amplias». En particular, se analizan las medidas para:

- Subsancar la falta de conocimiento de los enfoques de EIC en las AG y entre la comunidad responsable de la formulación de políticas de los Estados miembros en general.
- Remediar la inexistencia en los Estados miembros de contratistas externos, debidamente cualificados y con suficiente experiencia capaces de llevar a cabo EIC.
- Afrontar las barreras jurídicas presentes de forma genérica en todas las EIC.
- Avanzar hacia una mayor planificación de EIC.
- Ampliar el alcance de las EIC.

*Cuestiones más amplias a las que hay que hacer frente*

## 4.1. Mejorar el nivel de conocimiento entre las partes interesadas

Para el período de programación 2021-2027, el artículo 44, apartado 1, del RDC estipula que la AG «llevará a cabo evaluaciones de los programas en relación con uno o varios de los siguientes criterios: eficacia, eficiencia, pertinencia, coherencia y valor añadido de la Unión, con el fin de mejorar la calidad del diseño y la ejecución de los programas. Las evaluaciones podrán también llevarse a cabo en relación con otros criterios pertinentes, como la inclusividad, la no discriminación y la visibilidad, y referirse a más de un programa». Como se ha explicado en secciones anteriores, esto significa que la AG que contrate la CEI y los evaluadores que se postulen para el contrato deben poseer la capacidad de evaluación adecuada.

En algunos casos, el Servicio de Asistencia para las Evaluaciones constató que algunas administraciones carecían de la capacidad para llevar a cabo EIC, es decir, para presentar ofertas y acompañar su ejecución. Esto dificultó la realización de EIC por parte de los evaluadores, ya que no se habían definido preguntas de evaluación claras, no se había determinado la disponibilidad de datos básicos ni se había previsto una planificación bien fundamentada con antelación.

En ocasiones, y en especial en los países pequeños, el mercado de consultoría no tiene la capacidad de llevar a cabo EIC ni cuenta con los conocimientos técnicos necesarios. Por consiguiente, existe una necesidad generalizada de estimular tanto la demanda como la oferta de EIC. La oferta puede mejorar a medida que las AG y los Estados miembros empiecen a encargar EIC o hagan públicas sus solicitudes de tales estudios. La velocidad con que se responda al aumento de la demanda de EIC dependerá de las capacidades y la experiencia que existan previamente y de si los Estados miembros cuentan con instituciones capaces de aplicar estos enfoques. Sin embargo, la demanda también se puede estimular, en parte, a través de la mejora del conocimiento y la comprensión de los métodos de EIC entre los trabajadores de las AG.

*Estimular la oferta y la demanda de EIC*

Una solución a este problema es que las AG impartan cursos de formación sobre los métodos de EIC para su personal, los cuales deben centrarse en los beneficios de adoptar métodos de EIC para las AG. Por otra parte, se debe hacer hincapié en los problemas de rendición de cuentas y en aprender qué funciona. En el anexo 2 figura una propuesta para el esquema del curso.

*Desarrollar formación en métodos de EIC*

## 4.2. Desarrollo de capacidades

Otra cuestión que surgió durante el período 2014-2020, y que se menciona en las secciones anteriores, fue la necesidad de desarrollar la capacidad de llevar a cabo EIC en las comunidades investigadoras, académicas y consultoras de los Estados miembros. En algunos casos, fue evidente que los Estados miembros disponían de las capacidades necesarias para ello, pero que quienes poseían dichas capacidades se habían encontrado con obstáculos para ponerlas en práctica en un contexto de evaluación (p. ej., acceso limitado a datos utilizables o problemas para especificar un grupo de control razonable).

Es posible adoptar una serie de medidas para fomentar la prestación de servicios de evaluación. Muchas de las cuestiones planteadas se aplican tanto a las EIC como a las evaluaciones en general. Para mejorar la oferta de evaluaciones se suelen adoptar tres medidas:

*Fortalecer las instituciones y crear comunidades de prácticas*

- establecer relaciones con instituciones académicas, y en particular universidades;
- desarrollar y reforzar una comunidad independiente de consultores;
- apoyar la creación de una comunidad de evaluación profesional.

### Universidades

El establecimiento de vínculos con universidades es importante por dos razones. En primer lugar, es posible que el personal académico de las universidades posea las capacidades y los conocimientos necesarios para llevar a cabo EIC. Por ejemplo, muchos microeconomistas, expertos en econometría, sociólogos cuantitativos o psicólogos cuentan con los tipos de capacidades necesarias. Es posible que en muchos Estados miembros ya se disponga de las capacidades necesarias, pero que quienes las posean no se hubieran planteado antes aplicarlas a la evaluación de las intervenciones. Podrían no tener incentivos para hacerlo, en cuyo caso habría que solucionarlo.

*Desarrollar las capacidades académicas*

En algunos Estados miembros, es habitual que los investigadores académicos participen activamente en la investigación política aplicada. En este contexto, los académicos estarán acostumbrados a trabajar con el Gobierno y las AG. En otros Estados miembros donde las universidades y los académicos no participan en el trabajo aplicado, tal vez sea necesario un cambio de cultura. Un método útil para crear una base de proveedores dentro del sector universitario es que las autoridades de los Estados miembros y las AG aporten financiación básica para cubrir los costes de los centros de investigación dedicados a los métodos de EIC.

Las universidades y la comunidad académica también pueden desempeñar un papel en la formación de la próxima generación de evaluadores. Si se trabaja en estrecha colaboración con las universidades, tal vez se las pueda animar a incorporar métodos de evaluación de programas en sus planes de estudios y, como parte de este avance, garantizar que los programas de enseñanza incluyan los métodos de EIC. En algunos Estados miembros, las universidades también podrían participar en la organización de cursos de desarrollo profesional continuo relativos a los métodos de evaluación del impacto y EIC. Dichos cursos pueden dirigirse a responsables políticos, a especialistas técnicos de las AG y a otros posibles proveedores, como consultores independientes. Los Estados miembros podrían considerar la posibilidad de proporcionar financiación para dicha formación.

*Formar a la próxima generación*

### **Consultores independientes**

Existe un mercado internacional para ciertas evaluaciones a gran escala, como es sin duda el caso de las grandes EIC. Sin embargo, muchos Estados miembros querrán desarrollar su capacidad nacional para llevar a cabo EIC. Una estrategia para lograrlo puede ser el establecimiento de alianzas estratégicas entre los posibles proveedores nacionales y las consultorías internacionales.

*Desarrollar el mercado*

A continuación se presentan varias sugerencias para crear una base de proveedores nacionales para la realización de EIC que las AG (u otros organismos) que encargan EIC podrían poner en práctica:

- insistir en que las ofertas para establecer consorcios o alianzas incluyan siempre algunos consultores locales;
- baremar los contratos de evaluación de tal forma que los operadores nacionales recién llegados al mercado de la evaluación puedan llevar a cabo evaluaciones relativamente pequeñas y de bajo riesgo;
- garantizar que los requisitos técnicos y financieros asociados a las ofertas para las evaluaciones no sean demasiado restrictivos y permitan la participación de nuevos operadores;
- hacer hincapié en los criterios técnicos y prácticos en la selección, en lugar de en procedimientos administrativos complejos con los que los consultores menos experimentados podrían no estar familiarizados;
- organizar reuniones informativas con consultores potenciales para resolver sus dudas y fomentar las ofertas en un entorno competitivo;
- favorecer que consultores de evaluación relativamente aislados trabajen en red, a fin de fomentar la creación de equipos y la formación de consorcios y otras redes y asociaciones profesionales, también a escala internacional;

- conseguir que quienes encarguen una evaluación reconozcan que tal vez tengan que adoptar un enfoque de gestión más práctico para los nuevos contratistas, con miras a que adquieran conocimientos y experiencia más rápido.

### Recuadro 19 Ejemplo de un proyecto destinado a reforzar la cultura de las EIC y la capacidad para llevarlas a cabo

En el período de programación 2014-2020, se llevó a cabo un proyecto del FSE en Suecia destinado a mejorar la cultura y la capacidad de evaluación de los Servicios Públicos de Empleo (SPE). El proyecto, llamado «Proyectos con base empírica financiados por la UE», tuvo lugar entre 2016 y 2018 y contó con la participación de responsables y otros empleados de los SPE en talleres y otras acciones formativas en el ámbito de la evaluación.

Las actividades no se centraron en cuestiones técnicas, sino en la importancia de una evaluación, las principales etapas de una evaluación del impacto, la importancia de disponer de datos de calidad y el aseguramiento de la calidad de una evaluación. Se creó un sitio web donde se subió material educativo.

Los principales resultados del proyecto fueron la mejora de los conocimientos de los responsables de los SPE en el ámbito de las evaluaciones, su apertura a la práctica de la evaluación y su actitud cooperativa. En concreto, según la persona entrevistada, «los efectos de los proyectos en el número de evaluaciones del impacto realizadas» todavía no son elevados, pero es evidente que se ha avanzado, habida cuenta de que en el anterior período de programación (2007-2013) no se realizaron EIC en Suecia.

La evaluación del proyecto Ung framtid (Futuro Joven), que adoptó un enfoque aleatorizado, constituye un ejemplo práctico concreto en el que los responsables de los SPE intervinieron en el proceso de evaluación<sup>96</sup>.

### Comunidad profesional

Es importante desarrollar en los Estados miembros comunidades de evaluación profesional que permitan debatir los métodos de EIC y compartir experiencias. Crear comunidades profesionales es importante para el apoyo y el aprendizaje mutuos, así como para el mantenimiento de las normas de calidad. Una estrategia útil podría consistir en desarrollar vínculos con las sociedades nacionales de evaluación pertinentes y animarlas a promover las EIC de intervenciones del FSE+ en actos de formación, conferencias o seminarios específicos, o sesiones de sensibilización.

*Crear comunidades profesionales*

### Compartir experiencias

La CE aspira a que se lleven a cabo evaluaciones del impacto del FSE+ más rigurosas, y las EIC se recomendaron ampliamente en el período 2014-2020<sup>97</sup>. En la actualidad, es posible afirmar que el uso de EIC ha aumentado, pero siguen siendo limitadas en una serie de AG y Estados miembros. Compartir experiencias respecto de los métodos de EIC es uno de los principales medios para desarrollar las capacidades y respaldar la difusión de las EIC en toda la UE. A tal fin, deben utilizarse los foros de aprendizaje mutuo existentes, como las revisiones inter pares de las políticas de empleo e inclusión social, y las comunidades de prácticas en el marco del FSE+. Iniciativas de AG o Estados miembros concretos, como conferencias o

*Utilizar los foros existentes*

<sup>96</sup> La información se basa en una entrevista con un funcionario gubernamental del SPE sueco, el Arbetsförmedlingen.

<sup>97</sup> El anexo XI del RDC para el período 2014-2020 [Reglamento (UE) n.º 1303/2013] solicitaba un «sistema de indicadores de resultados» necesario para «hacer un seguimiento de los avances y realizar la evaluación de impacto». Dicho requisito era una de las condiciones *ex ante* para el período de programación 2014-2020.

seminarios internacionales, también podrían favorecer el intercambio de experiencias.

### 4.3. Superar las barreras jurídicas

El acceso a los datos es uno de los problemas más sustanciales y de mayor envergadura a los que se enfrentan los investigadores que realizan EIC en los Estados miembros. En particular, se topan a menudo con barreras jurídicas destinadas a proteger la confidencialidad de las personas representadas en los conjuntos de datos. Como se ha indicado anteriormente en relación con las características del RGPD, la solución a estos problemas reside en acometer reformas más amplias y celebrar acuerdos que permitan poner a disposición de los evaluadores los datos pertinentes de forma controlada y permanente.

*Eliminar las barreras jurídicas para acceder a los datos*

A partir de los datos administrativos que obran en poder de las autoridades de los Estados miembros, se podrían generar periódicamente versiones analíticas de los conjuntos de datos administrativos, las cuales se documentarían y depositarían en un archivo con acceso controlado. Los contratistas autorizados podrían extraer datos de dichas carteras con autorización. Si no fuera posible obtener un consentimiento específico, los datos se anonimizarían por completo con identificadores personales cifrados para respetar las normas del RGPD. En varios países se han creado carteras de datos de este tipo. No obstante, cuando el acceso siga suponiendo un obstáculo debido a las diferentes interpretaciones de las normas de privacidad, una iniciativa nacional a nivel gubernamental debe promover acuerdos y sistemas capaces de suministrar datos con fines de investigación en un período de tiempo relativamente corto.

*Crear conjuntos de datos analíticos*

Si persistiese la preocupación por la confidencialidad de los datos personales, debe valorarse la posibilidad de establecer laboratorios de datos. En este caso, los evaluadores que trabajen en conjuntos de datos administrativos solo tendrían acceso a los registros en lugares seguros, en los que se supervisaría y controlaría estrictamente el acceso a los datos. Los datos tendrían que tratarse y analizarse *in situ*, y solo los resultados de eventuales análisis podrían abandonar las instalaciones.

*Crear laboratorios de datos*

### 4.4. Avanzar hacia enfoques más prospectivos

Una característica común del reducido número de EIC de intervenciones financiadas por el FSE que se han realizado hasta la fecha es que han sido retrospectivas, en lugar de prospectivas. Esto significa que se ha encargado a los evaluadores expertos que lleven a cabo evaluaciones de intervenciones que ya han tenido lugar sin tener en cuenta la evaluación, y a veces en circunstancias en las que la planificación de una evaluación del impacto ha sido escasa o nula. De este modo, los evaluadores han tenido que construir fuentes de datos mediante un proceso largo y costoso y en condiciones que distan mucho de ser óptimas, aprovechando los datos disponibles, en lugar de contar con fuentes de datos creadas pensando en la evaluación del impacto.

Por el contrario, un enfoque prospectivo implicaría a los evaluadores en la planificación de una EIC lo antes posible y les permitiría influir en las intervenciones (nuevas o existentes), a menudo de forma bastante sutil, a fin de que resultase más fácil someterlas a una EIC. Planificar con antelación

*Los enfoques prospectivos cuentan con los*

una EIC puede suponer la diferencia entre poder llevar a cabo una evaluación rigurosa o no poder realizar una evaluación. La participación de personal interno debidamente formado o la contratación de contratistas expertos externos en una fase temprana de una intervención o cuando se tomen decisiones de financiación implica que:

*evaluadores desde el principio*

- se puede integrar un registro en la ejecución de los programas y las intervenciones;
- se puede determinar cuáles son las fuentes de datos necesarias en una fase temprana y se pueden resolver a su debido tiempo los problemas relacionados con el acceso y la protección de datos;
- se puede especificar la recopilación de datos de referencia y, en su caso, realizar encuestas;
- se pueden abordar las cuestiones prácticas relacionadas con la forma en que se capta a los participantes en las intervenciones de manera que los procesos de contratación sean más coherentes con una evaluación rigurosa.

La participación de evaluadores formados en métodos de EIC (ya sean evaluadores internos de las AG o expertos externos) en el proceso de desarrollo de nuevas intervenciones del FSE+ o en las decisiones relativas a las intervenciones existentes permitiría iniciar la planificación de la evaluación del impacto al principio del período de programación, así como obtener beneficios significativos en la toma de decisiones políticas con base empírica.

## 4.5. Ampliar el alcance de las EIC

Una última cuestión muy importante se refiere a la cobertura política de las EIC y abarca diferentes dimensiones: las políticas cubiertas, los efectos directos analizados y la integridad de la interpretación de las conclusiones.

En el período 2014-2020, la difusión de las EIC se centró casi en exclusiva en las políticas activas del mercado de trabajo y en los efectos relacionados con la situación laboral. Esto puede explicarse probablemente por el hecho de que, en general, suele ser más fácil acceder a datos administrativos sobre el empleo, y estos datos proporcionan la variable de resultado clave para el análisis de muchas políticas activas del mercado de trabajo.

*Mayor atención a las EIC en las políticas de inclusión social y educación*

Las políticas de educación e inclusión social no han sido sometidas a una EIC, o solo en contadas ocasiones, a pesar de su importancia en la estrategia y la asignación de fondos del FSE. En el recuadro 20 se presentan algunos de los pocos ejemplos disponibles de España, Polonia y Portugal.

Para contrarrestar esta escasez de EIC y mejorar los conocimientos sobre estas importantes políticas, es preciso combinar una serie de iniciativas:

*Implicar a agentes adicionales y desarrollar fuentes de datos adecuadas*

- Planificar y preparar EIC de medidas de inclusión o educación con antelación, con vistas a evaluar a tiempo la disponibilidad de datos; activar las colaboraciones necesarias entre los distintos agentes para aumentar esa disponibilidad, y especificar los grupos de tratamiento y de control en una fase temprana.
- Promover la participación de los propietarios de datos en el diseño de las EIC. Una mayor participación en las EIC de las administraciones responsables de las políticas educativas y sociales, cuando no coincidan

con la AG, podría dar lugar a un compromiso más firme y generalizado con la evaluación y el suministro de datos.

- Estimular los acuerdos institucionales y las herramientas informáticas para hacer que los datos administrativos sobre las condiciones sociales y la educación puedan utilizarse en EIC. A tal fin, puede ser importante que intervenga la oficina nacional de estadística, como institución «puente» y facilitadora, ya que estas oficinas cuentan con las capacidades necesarias y ya elaboran muchos conjuntos de datos administrativos para generar estadísticas nacionales.
- Apoyar la recopilación de los datos pertinentes de los participantes en los grupos de tratamiento y de control al inicio de la intervención con el fin de disponer de un conjunto de datos coherente sobre el antes y el después. Cuando no se disponga de datos administrativos o estos no sean pertinentes, se pueden obtener los datos necesarios mediante una encuesta específica. No obstante, esta debe comenzar al inicio de la intervención e implicar también al grupo de control para garantizar las comparaciones necesarias.

### Recuadro 20 Ejemplos de evaluaciones en el ámbito educativo

#### Intervenciones del FSE contra el abandono escolar prematuro llevadas a cabo en Asturias<sup>98</sup>

En el marco de una evaluación general del PO del FSE llevado a cabo en el Principado de Asturias en 2014-2020, se efectuó un análisis del impacto específico de los Programas de Diversificación Curricular y de Mejora del Aprendizaje y del Rendimiento (PMAR). Los PMAR son una medida del FSE ejecutada en el marco de la PI 10.i que tiene por objeto luchar contra el abandono escolar prematuro. En el marco de los PMAR, el alumnado se divide en grupos específicos (con entre ocho y quince estudiantes, pero podrían ser de otros tamaños si concurren circunstancias especiales), para cursar materias en los ámbitos de la lingüística, las ciencias sociales, las ciencias y las matemáticas, así como de las lenguas extranjeras (las asignaturas restantes se imparten en la clase principal). La medida se puso en marcha en el año académico 2016-2017, en el segundo y el tercer curso de la enseñanza secundaria obligatoria (ESO). Eran elegibles para participar los alumnos que habían repetido al menos un curso en cualquier etapa educativa y los que hubieran completado primero de la ESO, pero no estuvieran en condiciones de pasar de curso al año siguiente. La selección final se realizaba mediante una evaluación (académica y psicopedagógica) individual de los alumnos realizada por los equipos docentes.

Participaron un total de 1 053 estudiantes, 512 de segundo de la ESO y 541 de tercero. Se comparó a estos estudiantes con otros con dificultades de aprendizaje<sup>99</sup> para seleccionar alumnos de control comparables a los tratados en términos de criterios de elegibilidad (3 852 personas en total).

Se aplicó un enfoque de emparejamiento utilizando los datos administrativos de la base de datos del Sistema Informático para la Administración Unificada de Centros Educativos (SAUCE), que el Ministerio de Educación aportó de forma anonimizada. Estos datos se usaron para calcular la puntuación de propensión<sup>100</sup> y para medir la variable de resultado, que era pasar al curso siguiente. La evaluación puso de manifiesto que las intervenciones tenían un efecto positivo y estadísticamente significativo: los alumnos de segundo curso tratados presentaban una tasa de éxito un 18 % superior a la del grupo de control, mientras que en el caso de los estudiantes de tercer curso la diferencia era de aproximadamente el 16 %. En ambos grupos, el impacto fue más positivo para las alumnas que para los alumnos.

<sup>98</sup> Véase [Díaz, J.M. et al., 2019](#).

<sup>99</sup> Eran estudiantes que, en el año académico 2015-2016, se encontraban en primero o segundo de la ESO, habían terminado el curso con una o varias materias suspensas, y habían repetido al menos un curso en su vida escolar.

<sup>100</sup> Para emparejar a ambos grupos se utilizaron las siguientes variables: rendimiento académico en el curso anterior, tipo de centro educativo (público o privado), matriculación en otro curso durante el año académico anterior (sí o no), sexo, ingresos medios del municipio de residencia, lugar de residencia, país de nacimiento (España o resto del mundo).

**Intervenciones del FSE para mejorar la formación profesional en Podlaskie<sup>101</sup>**

El estudio examinó los efectos del proyecto Buena Profesión, Vida Fantástica, destinado a promover la educación y formación profesionales en la región polaca de Podlaskie. El proyecto se dirigía al personal de los centros de primer ciclo de enseñanza secundaria, estudiantes y padres, y en su marco se llevó a cabo una serie de actividades tanto para las instituciones como para las personas (estrategias de comercialización y comunicación a medida para los centros educativos, formación para profesores, cooperación con empresas y líderes industriales locales, tutorías para estudiantes y padres, asesoramiento profesional y educativo para estudiantes). El proyecto tuvo lugar entre 2017 y 2019.

La población tratada se compuso de los alumnos que participaron en el proyecto evaluado. De los 9 403 alumnos, 1 500 fueron seleccionados mediante muestreo aleatorio estratificado, y 200 participaron en la encuesta del estudio. La información sobre los alumnos que participaron en el proyecto se obtuvo a partir de los datos de seguimiento. El grupo de control se formó con alumnos con características similares a las de los integrantes del grupo tratado, pero que no participaron en el proyecto. Se planificó un grupo de control con un tamaño que prácticamente duplicaba el del grupo de tratamiento y cuyos integrantes se seleccionaron aleatoriamente del censo nacional de población (PESEL, un registro que asigna a cada ciudadano polaco un número individual). Sin embargo, debido a las dificultades causadas por la COVID-19, la metodología prevista no fue viable. Finalmente, se utilizó la técnica de bola de nieve (referencia en cadena) como método de muestreo primario, con el que se realizaron 401 entrevistas, de las cuales 384 se utilizaron en el estudio.

Tanto para el grupo de tratamiento como para el grupo de control, los datos sobre las variables de resultado<sup>102</sup> se recogieron mediante encuestas por cuestionario realizadas en entrevistas personales, telefónicas o por internet. El estudio utilizó la ponderación de las puntuaciones de propensión como método contrafactual para evaluar el impacto (solo a nivel general, no para subgrupos específicos). Los análisis contrafactuales efectuados en el estudio no arrojaron conclusiones concluyentes. La significación estadística de los resultados no fue satisfactoria.

**Evaluación del sistema de becas de enseñanza superior para estudiantes menos favorecidos en Portugal<sup>103</sup>**

Una segunda evaluación, que todavía no se ha publicado oficialmente, examina un programa de becas dirigido a estudiantes de hogares con bajos ingresos cuyo objetivo es favorecer el acceso a la enseñanza superior y aumentar el nivel de asistencia. La beca cuenta con el apoyo del FSE (desde el período 2007-2013) en las regiones Norte, Centro y Alentejo, y con el apoyo del Estado en el resto de regiones. Son elegibles los estudiantes cuyos recursos se encuentren por debajo de cierto umbral (entre 6 800 y 7 900 EUR de renta per cápita) y el curso anterior hayan aprobado un número mínimo de créditos, en el caso de aquellos que no estén en el primer año académico. Se aplican las mismas normas de elegibilidad, independientemente de que se matriculen en universidades públicas o privadas. El importe de la beca es proporcional a los ingresos del hogar.

Desde 2011, el programa ha contado con la participación de unos 70 000 estudiantes cada año, pero la evaluación abarca el período a partir de 2012, puesto que se dispone de datos de ese año en adelante. Además, la evaluación se centra en los estudiantes que solicitaron la beca por primera vez y que se matricularon en el primer año de un curso de grado (licenciatura o máster); la elegibilidad se determina únicamente por el criterio de los ingresos<sup>104</sup>.

Se utilizaron dos conjuntos de datos administrativos: datos sobre los estudiantes que solicitaron la beca entre 2012 y 2018 (facilitados por la Dirección General de Educación Superior de Portugal o DGES, por sus siglas en portugués), que se fusionaron por medio de un identificador único con otro conjunto de datos

<sup>101</sup> Véase [Zub M. et al., 2020](#).

<sup>102</sup> Las variables de resultado (basadas en las respuestas de las encuestas) fueron las siguientes: la asistencia a una escuela técnica en el momento del estudio, la asistencia a una escuela de comercio de primer grado en el momento del estudio, la asistencia a una escuela técnica o una escuela de comercio de primer grado en el momento del estudio, la solicitud de matriculación en un instituto de secundaria, la solicitud de matriculación en una escuela técnica, la solicitud de matriculación en una escuela de comercio de primer grado, la solicitud de matriculación en un centro donde se impartiese formación profesional (ya fuese una escuela técnica o una escuela de comercio de primer grado), el hecho de que el centro educativo cumpliera en gran medida las expectativas de los alumnos, y si la profesión estaba totalmente en consonancia con los intereses del alumno.

<sup>103</sup> Véase Guthmuller, S. y Meroni, E.C., pendiente de publicación.

<sup>104</sup> Está previsto hacer un análisis de toda la muestra de estudiantes (no solo de los del primer año), pero todavía no se ha completado.

que contenía información sobre la progresión y las carreras académicas (facilitada por la Dirección General de Estadística y Ciencia o DGEEC). La muestra final analizada está compuesta por 156 002 estudiantes, de los cuales 130 602 fueron tratados y 25 400 no. El umbral de ingresos (la variable de asignación) permitió al evaluador aplicar un diseño de regresión discontinua. Además, se midieron varias variables de resultado: a corto plazo, el porcentaje de estudiantes que seguían matriculados al final del primer año y los créditos aprobados al final del primer año; a largo plazo, la probabilidad de graduarse, la nota final obtenida y el número de años necesarios para graduarse.

En el período inicial, la beca demuestra tener efectos positivos en las tasas de matriculación, pero no en lo que respecta a los créditos aprobados. A más largo plazo, los estudiantes tratados se gradúan más, en menos tiempo y con mejor nota que el grupo de control. Los efectos fueron mayores para los varones, los estudiantes que residían en territorios cubiertos por el FSE, y los estudiantes matriculados en máster y en universidades públicas.

Ya se ha recalcado anteriormente la utilidad y la capacidad explicativa de las meta evaluaciones. El aumento general de las EIC y, por consiguiente, el posible incremento del número de meta evaluaciones reforzaría el debate relativo a las mediciones específicas, en el que participarían de forma simultánea múltiples AG y Estados miembros. Como se ha mencionado antes, la complejidad metodológica de estas evaluaciones debe superarse planificando mejor las EIC a nivel nacional e internacional.

*Aumentar el número de meta evaluaciones*

En las intervenciones del FSE, las EIC solían medir los efectos con variables cuantitativas o binarias (sí/no), como los ingresos o la situación laboral. Sin embargo, los efectos directos difíciles de cuantificar, que están relacionados con la autopercepción y la capacidad de los participantes, son elementos importantes para medir el éxito de las intervenciones. Por ejemplo, la «empleabilidad», la autoestima o las competencias adquiridas son condiciones previas para encontrar trabajo o para participar activamente en el mercado laboral. En muchos contextos, los análisis de efectos directos difíciles de cuantificar serían incluso más útiles que las cifras de salida del empleo y arrojarían luz sobre la correspondencia entre las intervenciones y las necesidades individuales.

*Valorar los efectos directos difíciles de cuantificar*

Las EIC pueden aplicarse a estos efectos directos difíciles de cuantificar, siempre y cuando la información se recopile antes y después de la intervención tanto en el grupo de tratamiento como en el grupo de control, de conformidad con la teoría del cambio de la intervención. Estas encuestas deben planificarse en una fase temprana e implicar a todos los agentes pertinentes, y en particular a los beneficiarios que estén en contacto directo y constante con los participantes. En el recuadro 21 se presenta un ejemplo de una EIC que evaluó los impactos sobre efectos directos difíciles de cuantificar.

### **Recuadro 21 Ejemplo de evaluación de los efectos sobre los efectos directos difíciles de cuantificar en Alemania**

**«Do job creation schemes improve the social integration and well-being of the long-term unemployed?» («¿Mejoran los planes de creación de empleo la integración social y el bienestar de los parados de larga duración?»)<sup>105</sup>**

El estudio examina los efectos de un plan de creación de empleo dirigido a un grupo vulnerable de personas, parados de larga duración que han sido solicitantes de asistencia social durante al menos cuatro años y con problemas de salud o hijos, o ambos. El programa estuvo operativo de 2015 a 2018 y contó

<sup>105</sup> Véase [Ivanovm, B. et al., 2020](#).

con la participación de unas 20 000 personas. La medida subvencionó hasta treinta y seis meses de contratos de trabajo ordinarios de treinta horas semanales (principalmente con empleadores públicos u organizaciones benéficas).

Dado que la medida de empleo subvencionado, que duraba hasta treinta y seis meses, tenía como objetivo explícito promover la integración social de los grupos destinatarios, los autores trataron de evaluar sus impactos en este sentido.

Para el análisis se utilizó un abundante conjunto de datos que integraba datos administrativos (las biografías laborales integradas alemanas) con una encuesta panel que respondieron los participantes en el programa y las personas de control. La encuesta panel permitió recabar información sobre mediciones subjetivas de la calidad de vida: la satisfacción vital, la salud mental, la pertenencia social y el estatus social. Para recoger información sobre estas mediciones se utilizaron escalas de Likert, en línea con otras encuestas (como la encuesta nacional PASS), para extraer conclusiones comparables. Estas fueron las etapas principales de la encuesta: en los datos administrativos se identificó a las personas tratadas y no tratadas, que fueron emparejadas; se entrevistó a los grupos de tratamiento y control resultantes (tres rondas); se creó el conjunto de datos final, cribando los datos y eliminando los casos en los que faltaban datos. La muestra se compuso de 2 531 parejas en la fase 1, de 1 191 en la fase 2 y de 450 en la fase 3 (de una muestra total de 62 000 personas, de las cuales 12 400 recibieron el tratamiento y 49 600 sirvieron de control).

Sobre la base del emparejamiento por puntuaciones de propensión y la medición de los efectos siete, dieciocho y veintinueve meses después de la entrada en el programa, los autores constataron que las intervenciones tenían efectos positivos en las mediciones de bienestar, pero en distintos grados. Por ejemplo, la satisfacción vital aumentó de forma sustancial, mientras que el estatus social solo mejoró moderadamente. Sin embargo, los efectos tienden a disminuir a medida que avanzan los programas. Una posible explicación puede ser el incremento de la cantidad de participantes que abandonan el programa y del número de personas de control que encuentran trabajo. En el caso de las personas más vulnerables, los efectos positivos fueron mayores.

Como se ha subrayado al principio de la presente Guía, la EIC es un valioso instrumento analítico para evaluar «qué» ha generado una intervención del FSE+. Para entender «cómo» o «por qué» se han producido los efectos medidos, es necesario recurrir a otros instrumentos.

*Integrar la EIC con una evaluación basada en la teoría*

En algunos casos, en los que las intervenciones son de sobra conocidas y ya se han evaluado en profundidad, un análisis relativamente limitado de su ejecución y algunas entrevistas con los beneficiarios bastan para determinar los principales elementos que dieron lugar a los efectos medidos. En intervenciones más complejas, menos conocidas o innovadoras, con frecuencia es necesario combinar la EIC con otras evaluaciones.

Una evaluación de la ejecución, o del proceso, que se complete antes de la EIC podría poner de relieve cómo se llevó a cabo la intervención, los problemas que surgieron en la ejecución y el modo en que los distintos agentes colaboraron para que tuviera éxito. Las conclusiones de esta evaluación pueden nutrir la EIC, pues indican qué efectos deben investigarse, cómo debe componerse el grupo de control en relación con las características de los participantes, y cuál es el calendario adecuado para una EIC.

En paralelo a la EIC, se podría llevar a cabo una evaluación basada en la teoría. Ambas evaluaciones podrían alimentarse mutuamente: la evaluación basada en la teoría aporta explicaciones relativas a los mecanismos contextuales, sociales e individuales que influyeron en los efectos y los hicieron posibles, mientras que la EIC calcula los efectos netos y aclara hasta qué punto la intervención tuvo efectos directos significativos. En una evaluación del impacto, se recomienda integrar la EIC y la evaluación basada en la teoría para obtener una visión global de los efectos y sus causas.

# Glosarios

## Acrónimos

ACB	Análisis coste-beneficio
AG	Autoridad de gestión
BLI	Biografías laborales integradas
CE	Comisión Europea
DED	Diferencia en diferencias
DG EMPL	Dirección General de Empleo, Asuntos Sociales e Inclusión
DG REGIO	Dirección General de Política Regional y Urbana
DRD	Diseño de regresión discontinua
EAC	Ensayo aleatorio controlado
EIC	Evaluación de impacto contrafactual
EPA	Encuesta de población activa
EPP	Emparejamiento por puntuaciones de propensión
FEDER	Fondo Europeo de Desarrollo Regional
FSE	Fondo Social Europeo
IEJ	Iniciativa de Empleo Juvenil
JLD	Conjunto de datos longitudinales de los demandantes de empleo
OI	Organismo intermediario
ONG	Organización no gubernamental
PAMT	Política activa de mercado de trabajo
PE	Plan de evaluación
PO	Programa operativo
Pyme	Pequeña y mediana empresa
RDC	Reglamento sobre Disposiciones Comunes
RGPD	Reglamento General de Protección de Datos [Reglamento (UE) 2016/679]
SPE	Servicio Público de Empleo
UE	Unión Europea
VI	Variable instrumental

## Definiciones

<b>Término</b>	<b>Definición</b>
Aleatorización	Es un método en el que los miembros de un grupo destinatario se asignan aleatoriamente a una serie de tratamientos o al grupo de control. La aleatorización garantiza que los grupos sean estadísticamente equivalentes en todos los aspectos en el momento en que son aleatorizados.
Análisis contrafactual	Una comparación entre lo que realmente ha sucedido y lo que habría ocurrido sin la intervención, en términos de resultados. Puesto que la diferencia entre los resultados reales y los contrafactuales define el efecto causal de la intervención, el análisis contrafactual abarca todos los enfoques destinados a evaluar la proporción del cambio observado que puede atribuirse a la intervención evaluada.
Beneficiario	Según el artículo 2 del Reglamento (UE) 2021/1060 del Parlamento Europeo y del Consejo (RDC), un beneficiario es «un organismo público o privado, una entidad con o sin personalidad jurídica, o una persona física, responsable de iniciar operaciones, o de iniciarlas y ejecutarlas». En el contexto de los instrumentos financieros (que en el marco del FSE+ en ocasiones se utilizan, por ejemplo, para conceder microcréditos para trabajar por cuenta propia), un beneficiario es el organismo que ejecuta el fondo de cartera o, si no existe una estructura de fondos de cartera, el organismo que ejecuta el fondo específico de que se trate o, si la autoridad de gestión gestiona el instrumento financiero, la autoridad de gestión.
Datos de referencia	Datos sobre las variables medidas antes de que una unidad (persona o empresa) sea expuesta a una intervención. En muchos casos, se recogerán mediciones de los resultados de una intervención realizadas antes del tratamiento tanto para el grupo de tratamiento como para el grupo de control.
Diferencia en diferencias (DED)	En su forma más sencilla, la diferencia en un resultado antes y después del tratamiento en un grupo de control se deduce de la misma diferencia observada en un grupo tratado, a fin de obtener una estimación del impacto de una intervención. Los impactos calculados a partir de la diferencia en diferencias se obtienen normalmente dentro de un marco de regresión.
Diseño de regresión discontinua (DRD)	Este método puede adoptarse cuando el acceso a una intervención venga determinado por un punto de corte a lo largo de una calificación, escala o medida continua. El enfoque se basa en el hecho de que las personas que se encuentren inmediatamente alrededor del punto de corte serán muy similares entre sí, salvo porque las que se encuentren en un lado de dicho punto de corte participarán en la intervención y las que se encuentren al otro, no. Los resultados de quienes estén por encima y por debajo del punto de corte pueden compararse para conocer el impacto de una intervención.
Diseño no aleatorizado o cuasi experimental	Enfoques de la evaluación de impacto contrafactual en los que los grupos de control se crean mediante métodos distintos de la aleatorización.
Efecto directo	El conjunto de efectos a corto y medio plazo probables o logrados por los productos de una intervención <sup>106</sup> . El término «efecto directo» es similar a «resultado», pero se utiliza con mayor frecuencia en las evaluaciones del impacto.
Eficacia	Se refiere a la «consecución de los objetivos» y se evalúa comparando lo que se ha conseguido con lo que se tenía previsto conseguir (o con una situación de referencia), o bien comparando lo observado después de que la acción haya tenido lugar con lo que habría ocurrido sin la acción (situación contrafactual).
Eficiencia	La eficiencia se define como la obtención de un determinado producto al mínimo coste o, de manera equivalente, como la maximización del producto para un nivel determinado de recursos. Puede determinarse mediante un análisis coste-beneficio o un análisis de coste/eficacia.
Emparejamiento	Se trata de un método en el que las muestras de intervención y de control se emparejan entre sí atendiendo a sus características observadas.
Emparejamiento por	El EPP implica estimar un modelo estadístico para toda la muestra (unidades de tratamiento y posibles unidades de control), lo cual da lugar a una propensión estimada

<sup>106</sup> Véase [OCDE, 2010](#).

Término	Definición
puntuaciones de propensión (EPP)	a participar para cada persona o empresa, independientemente de si realmente participaron o no. A continuación, se empareja a las personas o empresas tratadas con una o varias personas o empresas no tratadas, atendiendo a la puntuación de propensión.
Enfoque de las variables instrumentales (VI)	Cuando se aplica este método, la selección para acceder al tratamiento debe estar determinada, al menos en parte, por un factor (o instrumento) exógeno que no guarde relación con resultados distintos de los obtenidos con el tratamiento. Por lo tanto, el factor exógeno influye en la participación, pero no directamente en los resultados.
Evaluación de impacto contrafactual	Tipo de evaluación del impacto que intenta determinar los efectos causales de las intervenciones mediante la estimación de los resultados contrafactuales medios y su sustracción de los resultados medios observados en las unidades tratadas. Las estimaciones de los resultados contrafactuales se obtienen normalmente de grupos de control cuidadosamente seleccionados para ser lo más similares posible al grupo tratado.
Evaluación del proceso	La evaluación del proceso se centra en la ejecución del programa, lo que comprende, entre otras cosas, cómo se prestan los servicios, las diferencias entre la población a la que se quiere apoyar y la población a la que se apoya realmente, el acceso al programa y las prácticas de gestión.
Evaluación externa	Evaluación realizada por un evaluador externo e independiente sobre la base de un procedimiento de licitación.
Evaluación interna	Evaluación realizada internamente, es decir, encargada directamente por una institución o unidad pública independiente (la AG o el OI) sin un proceso de licitación o como parte de un proceso de seguimiento y análisis ampliado.
Grupo de control	Un grupo de personas, empresas u otras unidades que es lo más similar posible al grupo de tratamiento, pero que no recibe tratamiento alguno, y del que se obtienen estimaciones contrafactuales de los resultados. En sentido estricto, el término «grupo de control» se refiere a contextos experimentales (véase la definición de «aleatorización»), y el término «grupo de comparación» se refiere a contextos cuasi experimentales, pero en la práctica ambos se utilizan indistintamente.
Grupo de tratamiento	Un grupo de personas, empresas u otras unidades que se benefician de una intervención o están expuestas a ella (engloba tanto a aquellas a las que se ofrece el tratamiento como las que realmente lo reciben).
Impacto	En el contexto de las EIC, los impactos se refieren a los efectos netos, definidos como la diferencia entre los resultados del tratamiento medio y los resultados contrafactuales. A efectos de la presente Guía, el término «impactos» se utiliza de forma intercambiable con «efectos netos».
Intervenciones	Se refiere generalmente a las operaciones de los programas operativos del FSE o a los proyectos cofinanciados por el FSE.
Participantes	Un participante es una persona física que se beneficia directamente de una operación, sin ser responsable de ponerla en marcha ni de ejecutarla. En el contexto del FSE, se refiere a las personas apoyadas por las intervenciones del FSE.
Pertinencia	La pertinencia se refiere a la idoneidad de los objetivos explícitos de una intervención en relación con los problemas socioeconómicos que se pretende resolver <sup>107</sup> .
Plan de evaluación	De conformidad con el artículo 44 del Reglamento (UE) 2021/1060 del Parlamento Europeo y del Consejo (RDC), el Estado miembro o la autoridad de gestión elaborará un plan de evaluación (PE) que podrá aplicarse a más de un programa y deberá presentarse al comité de seguimiento como máximo un año después de la decisión por la que se aprueba el programa.
Producto	Se considera que un producto es aquello que se produce o se suministra directamente mediante la ejecución de una operación del FSE, medido en unidades físicas o monetarias. Los productos se miden principalmente en términos del número de personas a las que se ha apoyado, las entidades apoyadas, los bienes y servicios

<sup>107</sup> Véase [Comisión Europea, 2013](#).

<b>Término</b>	<b>Definición</b>
Resultado	<p>prestados, y los proyectos ejecutados.</p> <p>Los efectos sobre los participantes o entidades generados por una operación; por ejemplo, sobre su situación laboral, sus ingresos, sus puntuaciones en pruebas educativas normalizadas, sus beneficios, etc. Los efectos pueden medirse a corto o largo plazo. En el contexto del FSE+, al medir los indicadores de resultados, se considera que el corto plazo se refiere al período inmediatamente posterior a la participación (cuatro semanas), mientras se considera que el largo plazo es transcurridos seis meses<sup>108</sup>. Sin embargo, en el contexto del análisis del impacto, se suele entender que el largo plazo empieza después, y concretamente veinticuatro meses o más después de la intervención.</p>

---

<sup>108</sup> Véase [Comisión Europea, 2021a](#), y [Comisión Europea, 2018](#).

## Bibliografía

- Ashenfelter, O. (1978): «Estimating the effect of training programmes on earnings», *Review of Economics and Statistics*, número 6, pp. 47-57.  
<https://www.jstor.org/stable/1924332?origin=crossref>
- Axdorph, E.; Egebark, J.; Lundström, T., y Özcan, G. (2019): *Effekter av förstärkta förmedlingsinsatser för unga arbetssökande – Resultat från utvärderingen av Ung framtid*, Arbetsförmedlingen.
- Baran, J. et al. (2016): *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. I Raport Wskaźnikowy*.  
<http://files.evaluationhelpdesk.eu/Evaluations/PLE49.pdf>
- Baran, J. et al. (2017): *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój II Raport wskaźnikowy*.  
<http://files.evaluationhelpdesk.eu/Evaluations/PLE196.pdf>
- Baran, J. et al. (2018): *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój III Raport wskaźnikowy*.  
<http://files.evaluationhelpdesk.eu/Evaluations/PLE197.pdf>
- Baran, J. et al. (2018a): *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. II Raport Tematyczny*.  
<http://files.evaluationhelpdesk.eu/Evaluations/PLE221.pdf>
- Bazzoli, M.; De Poli, S.; Rettore, E., y Schizzerotto, A. (2018): «Are Vocational Training Programmes Worth Their Cost? Evidence from a Cost-Benefit Analysis», *Politica Economica/Journal of Economic Policy*, vol. XXXIV(3), pp. 215-240.  
<https://www.rivisteweb.it/doi/10.1429/92119>
- Bloom, H. S. (1995): «Minimum detectable effects: A simple way to report the statistical power of experimental designs», *Evaluation Review*, 8(2), pp. 225-246. <https://cpb-us-e2.wpmucdn.com/sites.uci.edu/dist/1/1159/files/2021/03/Bloom-MDES-Eval-Rev-1995-Bloom.pdf>
- Bloom, H. S. (2009): «Modern regression discontinuity analysis», *MDRC Working Papers on Research Methodology*, MDRC, Nueva York. <https://www.mdrc.org/publication/modern-regression-discontinuity-analysis>
- Boockmann, B. et al. (2017): *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziale*.  
<http://files.evaluationhelpdesk.eu/Evaluations/DEE189.pdf>
- Boockmann, B. et al. (2018): *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziale*.  
<http://files.evaluationhelpdesk.eu/Evaluations/DEE52.pdf>
- Boockmann, B. et al. (2019): *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziale*.  
<http://files.evaluationhelpdesk.eu/Evaluations/DEE94.pdf>
- Boockmann, B. et al. (2021): *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziale*.  
<http://files.evaluationhelpdesk.eu/Evaluations/DEE202.pdf>

- Borik, V. et al. (2015): *The net effects of graduate work experience and the promotion of self-employment. Technical Report.* <http://files.evaluationhelpdesk.eu/Evaluations/SKE8.pdf>
- Bratti, M. et al. (2018): «Vocational Training for Unemployed Youth in Latvia: Evidence from a Regression Discontinuity Design», *IZA DP*, n.º 11870. <http://ftp.iza.org/dp11870.pdf>
- Bratu, C. et al. (2014): *Knowledge gaps in evaluating labour market and social inclusion policies: use of counterfactual impact evaluation*, doi: 10.2788/083390. <https://op.europa.eu/en/publication-detail/-/publication/d0e73217-2a40-47af-a5c3-a700dcd47da3/language-en>
- Bredgaard, T. (2015): «Evaluating what Works for whom in active Labour market policies», *European Journal of Social Security*, 17(4), pp. 436-452. <https://doi.org/10.1177/138826271501700403>
- Bryson, A.; Dorsett, R., y Purdon, S. (2002): *The use of propensity score matching in the evaluation of active labour market policies*, documento de trabajo n.º 4, Departamento de Trabajo y Pensiones. [http://eprints.lse.ac.uk/4993/1/The\\_use\\_of\\_propensity\\_score\\_matching\\_in\\_the\\_evaluation\\_of\\_active\\_labour\\_market\\_policies.pdf](http://eprints.lse.ac.uk/4993/1/The_use_of_propensity_score_matching_in_the_evaluation_of_active_labour_market_policies.pdf)
- Caliendo, M. y Kopeinig, S. (2008): «Some practical Guidance for the implementation of propensity score matching», *Journal of Economic Surveys*, vol. 22, 1.ª ed., pp. 31-72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Caliendo, M.; Mahlstedt, R., y Mitnik, O. A. (2017): «Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labour market policies», *Labour Economics*, 46, pp. 14-25. <https://doi.org/10.1016/j.labeco.2017.02.001>
- Card, D.; Ibarra, P., y Villa, J. M. (2011): *Building in an evaluation component for active labour market programs: a practitioner's guide*, documento de reflexión n.º 6 085, IZA, Bonn, Alemania. <http://ftp.iza.org/dp6085.pdf>
- Card, D.; Kluve, J., y Weber, A. (2017): «What works? A meta-analysis of recent active labor market program evaluations», *Journal of the European Economic Association*, vol. 16, 3.ª edición, junio de 2018, pp. 894-931. <https://doi.org/10.1093/jeea/jvx028>
- Centre for Disease Control and Prevention (CDC) (2013): *Good Evaluation Questions: Checklist to Help Focus Your Evaluation*, Departamento de Salud y Servicios Humanos de Estados Unidos. [http://www.cdc.gov/asthma/program\\_eval/AssessingEvaluationQuestionChecklist.pdf](http://www.cdc.gov/asthma/program_eval/AssessingEvaluationQuestionChecklist.pdf)
- Danmarks Statistik (2017): *Effektmåling og monitorering 2016 - strukturfondsindsatsen i Syddanmark.* <http://files.evaluationhelpdesk.eu/Evaluations/DKE7.pdf>
- Danmarks Statistik (2018): *Effektmåling af den Virksomhedsrettede Strukturfondsindsats 2007-2013.* <http://files.evaluationhelpdesk.eu/Evaluations/DKE10.pdf>
- Díaz, J. M. et al. (2019): *Evaluación del PO-FSE 2014/2020 del Principado de Asturias para el informe anual a presentar en 2019.* <http://files.evaluationhelpdesk.eu/Evaluations/ESE101.pdf>
- Ecorys e Ismeri Europa (2020): *Study supporting the evaluation of ESF support to education and training (Thematic Objective 10) for the Commission, Directorate-General for Employment, Social Affairs and Inclusion.* <https://op.europa.eu/en/publication-detail/-/publication/d0c1a558-077d-11eb-a511-01aa75ed71a1/language-en/format-PDF/source-173162502>
- Comisión Europea (2007): *Indicative Guidelines on evaluation methods: evaluation during the programming period*, documento de trabajo n.º 5, DG Política Regional y Urbana. [https://ec.europa.eu/regional\\_policy/sources/docoffic/2007/working/wd5\\_ongoing\\_en.pdf](https://ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd5_ongoing_en.pdf)
- Comisión Europea (2013): *EVALSED: The Resource for the Evaluation of Socio-Economic Development*, versión actualizada.

[https://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/guide/guide\\_evalsed.pdf](https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf).

Comisión Europea (2017): *Better Regulation Toolbox, Tool #47 on Evaluation criteria and questions*. [https://ec.europa.eu/info/sites/default/files/file\\_import/better-regulation-toolbox-47\\_en\\_0.pdf](https://ec.europa.eu/info/sites/default/files/file_import/better-regulation-toolbox-47_en_0.pdf)

Comisión Europea: (2018): *Monitoring and Evaluation of European Cohesion Policy European Social Fund*, documento de orientación.

<https://ec.europa.eu/sfc/en/system/files/ged/ESF%20monitoring%20and%20evaluation%20guidance.pdf>

Comisión Europea (2019): *Advanced counterfactual evaluation methods. Guidance document*, doi:10.2767/464242.

<https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8254&furtherPubs=yes>

Comisión Europea (2020): *Counterfactual impact evaluation of European Social Fund interventions in practice. Guidance document for Managing Authorities*, doi: 10.2767/55495.

<https://op.europa.eu/en/publication-detail/-/publication/f82c5fb8-bb40-11ea-811c-01aa75ed71a1/language-en>

Comisión Europea (2020): *How to use administrative data for European Social Funds counterfactual impact evaluations. A step-by-step guide for managing authorities*.

<https://op.europa.eu/en/publication-detail/-/publication/d96feed3-f30c-11ea-991b-01aa75ed71a1>

Comisión Europea (2021): Reglamento (UE) 2021/1060 del Parlamento Europeo y del Consejo, de 24 de junio de 2021, por el que se establecen las disposiciones comunes relativas al Fondo Europeo de Desarrollo Regional, al Fondo Social Europeo Plus, al Fondo de Cohesión, al Fondo de Transición Justa y al Fondo Europeo Marítimo, de Pesca y de Acuicultura, así como las normas financieras para dichos Fondos y para el Fondo de Asilo, Migración e Integración, el Fondo de Seguridad Interior y el Instrumento de Apoyo Financiero a la Gestión de Fronteras y la Política de Visados.

Comisión Europea (2021a): *Common Indicators Toolbox Working document*, junio de 2021.

[https://ec.europa.eu/sfc/en/system/files/2021/ged/toolbox\\_june\\_2021\\_final.pdf](https://ec.europa.eu/sfc/en/system/files/2021/ged/toolbox_june_2021_final.pdf)

Fondazione G. Brodolini, Metis GmbH, Applica, y Ockham IPS (2020): *Study for the Evaluation of ESF Support to Employment and Labour Mobility (Thematic Objective 8), for the European Commission, Directorate- General for Employment, Social Affairs and Inclusion*.

<https://ec.europa.eu/social/BlobServlet?docId=22899&langId=en>

Frolich, M. (2004): «Programme evaluation with multiple treatments», *Journal of Economic Surveys*, 18(2), pp. 181-224.

<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.620.9209&rep=rep1&type=pdf>

Guthmuller, S. y Meroni, E. C.: *Evaluation of the higher education grant system for less privileged students in Portugal*. Informe técnico del JRC. Pendiente de publicación.

Hagglund, P. (2006): *A description of three randomised experiments in Swedish labour market policy*, *Institute for Labour Market Policy Evaluation*, informe 2006:4.

<https://www.ifau.se/globalassets/pdf/se/2006/r06-04.pdf>

Heckman, J. J.; Ichimura, H.; Smith, J., y Todd, P. (1998): «Characterizing selection bias using experimental data», *Econometrica*, vol. 66, n.º 5, pp. 1017-1098.

[http://jenni.uchicago.edu/papers/Heckman\\_Ichimura\\_etal\\_1998\\_Econometrica\\_v66\\_n5\\_r.pdf](http://jenni.uchicago.edu/papers/Heckman_Ichimura_etal_1998_Econometrica_v66_n5_r.pdf)

Hunger, K. y Sattler, K. (2017): *Evaluationsbericht zum spezifischen Ziel A 1.1. im Rahmen der Evaluierung des Operationellen Programms des Europäischen Sozialfonds in Baden-Württemberg 2014–2020*. <http://files.evaluationhelpdesk.eu/Evaluations/DEE26.pdf>

Holland, P. W. (1986): «Statistics and Causal Inference», *Journal of the American Statistical Association*, 81 (396), pp. 945-960, doi: 10.1080/01621459.1986.10478354.

<http://people.umass.edu/~stane/pdffiles/causal-holland.pdf>

IAW Institut für Angewandte Wirtschaftsforschung e ISG Institut für Sozialforschung und Gesellschaftspolitik, GmbH (2015): *Evaluation der Modellprojekte „Bürgerarbeit“*. Endbericht. <http://files.evaluationhelpdesk.eu/Evaluations/DEE7.pdf>

ICF, Cambridge Econometrics y Eurocentre (2020): *Study supporting the 2020 evaluation of promoting social inclusion, combating poverty and any discrimination by the European Social Fund (Thematic Objective 9) for the European Commission*, Dirección General de Empleo, Asuntos Sociales e Inclusión. <https://op.europa.eu/en/publication-detail/-/publication/8788ec85-2308-11eb-b57e-01aa75ed71a1>

Indecon (2016): *Indecon Evaluation of JobBridge Activation Programme*. <http://files.evaluationhelpdesk.eu/Evaluations/IEE2.pdf>

Instytut Badań Strukturalnych, Imapp e IQS (2015): *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój*. I Raport Tematyczny. <http://files.evaluationhelpdesk.eu/Evaluations/PLE34.pdf>

Ires Piemonte. (2019): *Misure di sostegno all'imprenditorialità*. <http://files.evaluationhelpdesk.eu/Evaluations/ITE61.pdf>

Isfol (2016): *Primo rapporto di valutazione del piano italiano Garanzia giovani*. <http://files.evaluationhelpdesk.eu/Evaluations/ITE13.pdf>

Ismeri Europa (2018): *I Rapporto tematico di valutazione – I risultati di PIPOL*. <http://files.evaluationhelpdesk.eu/Evaluations/ITE36.pdf>

Ismeri Europa, Ecorys e Instituto de Estudios de Empleo (2019): *Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations*, doi: 10.2767/39339. <https://op.europa.eu/es/publication-detail/-/publication/84cc9eb9-b33d-11e9-9d01-01aa75ed71a1>

Ivanov, B. et al. (2020): «Do Job creation schemes improve the social integration and well-being of the long-term unemployed?», *Labour Economics*. <https://doi.org/10.1016/j.labeco.2020.101836>.

Kalinowski, H. (2020): *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój V Raport wskaźnikowy*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE296.pdf>

Kalinowski, H. et al. (2020): *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój Raport końcowy*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE339.pdf>

Kluve, J.; Lehmann, H., y Schmidt, C. M. (2008): «Disentangling treatment effects of Active Labor Market Policies: The role of labor force status sequences», *Labour Economics*, 15(6), pp. 1270-1295. <https://doi.org/10.1016/j.labeco.2007.12.002>

Krug, G. y Stephan, G. (2011): «Is contracting-out intensified placement services more effective than in-house production? Evidence from a randomized field experiment», *LASER Discussion Papers*, artículo n.º 5. <http://doku.iab.de/externe/2011/k110912303.pdf>

Lammers, M. y Kok, L. (2021): «Are active labor market policies (cost-)effective in the long run? Evidence from the Netherlands», *Empirical Economics*, 60, pp. 1719–1746. <https://doi.org/10.1007/s00181-019-01812-3>

Lang, J. (2021): «Employment effects of language training for unemployed immigrants», *Journal of Population Economics*. <https://doi.org/10.1007/s00148-021-00832-7>  
<http://files.evaluationhelpdesk.eu/Evaluations/PLE359.pdf>

Martini, A. (2009): *Counterfactual impact evaluation: what it can (and cannot) do for cohesion*

*policy, prepared for the 6th European Conference on Evaluation of Cohesion Policy*, Varsovia, 30 de noviembre.

[https://ec.europa.eu/regional\\_policy/archive/conferences/evaluation2009/abstracts/martini.doc](https://ec.europa.eu/regional_policy/archive/conferences/evaluation2009/abstracts/martini.doc)

Morgan, S. L. y Winship, C. (2014): *Counterfactual and causal inference: Methods and principles for social research*, 2.ª ed., Cambridge University Press, Cambridge y Nueva York.

OCDE (2010): *Glosario de los principales términos sobre evaluación y gestión basada en resultados*, París. <https://www.oecd.org/dac/evaluation/2754804.pdf>

Openfield (2019): *Analiza skuteczności i efektywności dotacji na założenie działalności gospodarczej udzielonych w ramach 8 osi priorytetowej RPO WM - komponent 3*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE312.pdf>

Palczyńska, M. et al. (2019): *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój IV Raport wskaźnikowy*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE274.pdf>

Pomatto, G. (2017): *L'attuazione del Buono per Servizi al Lavoro nella Regione Piemonte*, Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE55.pdf>

Pomatto, G. (2019): *Buoni per servizi al lavoro nella Regione Piemonte: qualità percepita dai destinatari e meccanismi dell'attuazione*, Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE77.pdf>

Pompili, M. y Giorgetti, I. (2020): *Rapporto di Placement. Servizio di attività di valutazione del POR FSE 2014/2020 – Regione Marche*. <http://files.evaluationhelpdesk.eu/Evaluations/ITE115.pdf>

Pompili, M. y Giorgetti, I. (2020a): *Rapporto tematico «Disoccupazione di lunga durata». Servizio di attività di valutazione del POR FSE 2014/2020 – Regione Marche*. <http://files.evaluationhelpdesk.eu/Evaluations/ITE214.pdf>

Poy, S. (2019): *Gli effetti occupazionali del buono per servizi al lavoro nella Regione Piemonte: prime evidenze. Misura per disoccupati da almeno 6 mesi, anno 2017*, Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE123.pdf>

Poy, S. (2020): *Nuove evidenze sull'effetto occupazionale del buono per servizi al lavoro. Target persone disoccupate da almeno 6 mesi*, Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE245.pdf>

Re-source Pracownia Badań i Doradztwa (2015): *Ocena działań na rzecz wydłużenia aktywności zawodowej osób starszych w kontekście niekorzystnej sytuacji demograficznej w województwie lubelskim*. <http://files.evaluationhelpdesk.eu/evaluations/PLE126.pdf>

Riccio, J.; Friedlander, D., y Freedman, S. (1994): *GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program*, MDRC, Nueva York. <https://www.mdrc.org/publication/gain-benefits-costs-and-three-year-impacts-welfare-work-program>

Scheller, F. y Seidel, K. (2020): *Zweiter Evaluationsbericht zum spezifischen Ziel A 1.1: Teilnehmer\*innenperspektive und Wirkungsanalyse. im Rahmen der Evaluierung des Operationellen Programms des Europäischen Sozialfonds in Baden-Württemberg 2014–2020*. <http://files.evaluationhelpdesk.eu/Evaluations/DEE171.pdf>

Shadish, W. R.; Cook, T. D., y Campbell, D. T. (2002): *Experimental and quasi-experimental designs for generalised causal inference*, Houghton Mifflin Company Boston, Estados Unidos.

Stern, E. et al. (2012): *Broadening the range of designs and methods for impact evaluations*, Departamento de Desarrollo Internacional del Reino Unido, documento de trabajo 38. <https://assets.publishing.service.gov.uk/media/57a08a6740f0b6497400059e/DFIDWorkingPaper>

38.pdf

WK Kellogg Foundation (2004): *Logic Model Development Guide*. <https://www.wkcf.org/resource-directory/resources/2004/01/logic-model-development-guide>

Zub M. *et al.* (2020): *Ocena wpływu wsparcia RPOWP na popularyzację szkolnictwa zawodowego w województwie podlaskim. Raport końcowy.*

# Anexos

## Anexo 1. Bibliografía adicional

A continuación se proponen lecturas adicionales para el personal de la autoridad de gestión interesado en profundizar en las cuestiones tratadas en la presente Guía. Existe una amplia bibliografía en torno al tema de la evaluación. El objetivo de esta lista es presentar los principales análisis de confianza que proporcionan información de utilidad inmediata para planificar una EIC. Tras la mayoría de las referencias se proporciona una breve descripción de su contenido.

### Evaluación general

- Rossi, P. H.; Lipsey, M. W., y Henry G. T (2018): *Evaluation: A Systematic Approach*. 8.ª ed., SAGE Publications, Thousand Oaks, California.

El clásico manual sobre prácticas y métodos de evaluación. Incluye métodos y ejemplos.

- Gertler, P. J.; Martínez, S. Premand, P.; Rawlings, L. B., y Vermeersch, C. M. J. (2016): *La evaluación de impacto en la práctica*, 2.ª ed., Banco Interamericano de Desarrollo y Banco Mundial, Washington D. C. © Banco Mundial.  
<https://openknowledge.worldbank.org/handle/10986/25030> (Disponible en inglés, portugués y español).

Al igual que la presente Guía, este manual comienza con la evaluación clásica y, a continuación, examina las alternativas. Aunque se escribió pensando en los responsables de programas en países de renta baja, el análisis es pertinente y fácilmente aplicable en el contexto de los Estados miembros de la UE.

- White, H. y Raitzer, D. A. (2017): *Impact evaluation of development interventions: A practical guide*, Banco Asiático de Desarrollo.  
<https://www.adb.org/sites/default/files/publication/392376/impact-evaluation-development-interventions-guide.pdf>

Al igual que Gertler *et al.* (2016), este manual, que se ha mencionado anteriormente al hablar de la evaluación del impacto de intervenciones en países de renta media-baja, contiene un análisis exhaustivo de los métodos de EIC generales. Aunque es más técnico que Gertler *et al.* (2016), contiene numerosas explicaciones técnicas que podrían ser de interés para los responsables de programas del FSE.

- Comisión Europea (2019): *Advanced counterfactual evaluation methods. Guidance document*, doi:10.2767/464242. <https://op.europa.eu/en/publication-detail/-/publication/11968bbb-fac9-11e9-8c1f-01aa75ed71a1/language-en>

El documento presenta los métodos de evaluación de impacto contrafactual recientes y más avanzados, como el análisis de secuencia, el emparejamiento dinámico y los controles sintéticos.

- Csillag, M.; Kreko, J., y Scharle, A. (2020). *Evaluación contrafactual de las políticas de empleo para jóvenes*, elaborado en el marco del proyecto Youth Employment PartnerSHIP. [https://www.fedea.net/ship/docs/YEP\\_guia\\_metodologica.pdf](https://www.fedea.net/ship/docs/YEP_guia_metodologica.pdf) [Disponible en inglés, español, húngaro (traducción en curso), italiano y polaco].

Se trata de una introducción «paso a paso» a la evaluación contrafactual de las políticas del mercado laboral dirigidas a los jóvenes, centrada en el uso de datos administrativos. Se presentan distintas cuestiones sobre la base de los problemas prácticos encontrados en cuatro EIC de subvenciones para la contratación de jóvenes en España, Hungría, Italia y Polonia.

- HM Treasury (2020): *The Magenta Book: Guidance for evaluation*, The Agency, Londres, Reino Unido. <https://www.gov.uk/government/publications/the-magenta-book>

El «Libro magenta» examina en profundidad la metodología de evaluación. Estos documentos son interesantes como ejemplos de la perspectiva de evaluación interna del Gobierno.

### Ensayos aleatorios controlados

- White, H. (2013): «An introduction to the use of randomised control trials to evaluate development interventions», *Journal of Development Effectiveness*, 5:1, pp. 30-49, doi: 10.1080/19439342.2013.764652. <https://www.tandfonline.com/doi/pdf/10.1080/19439342.2013.764652>
- White, H.; Sabarwal, S., y De Hoop, T. (2014): *Randomized Controlled Trials (RCTs), Methodological Briefs: Impact Evaluation 7*, Oficina de Investigación del UNICEF, Florencia. [https://www.unicef-irc.org/publications/pdf/brief\\_7\\_randomized\\_controlled\\_trials\\_eng.pdf](https://www.unicef-irc.org/publications/pdf/brief_7_randomized_controlled_trials_eng.pdf)

Dos introducciones no técnicas a la lógica del EAC, que además analizan varios diseños y críticas de este tipo de ensayos.

- Glennerster, R. y Takavarasha, K. (2013): *Running Randomized Evaluations: A Practical Guide*, Princeton University Press, Princeton, Nueva Jersey.

El libro es una guía paso a paso sobre cómo diseñar y llevar a cabo EAC en el ámbito de los programas sociales. Se basa en los EAC concretos efectuados por Abdul Latif Jameel Poverty Action Lab.

### Diferencia en diferencias

- Card, D.; Ibararán, P. y Villa, J. M. (2011): *Building in an Evaluation Component for Active Labor Market Programs: A Practitioner's Guide*. documento de reflexión del IZA n.º 6085, Forschungsinstitut zur Zukunft der Arbeit, Bonn. <http://ftp.iza.org/dp6085.pdf>

Compara la diferencia en diferencias con el EAC.

- Lechner, M. (2011): «The Estimation of Causal Effects by Difference-in-Difference Methods», *Foundations and Trends in Econometrics*, vol. 4, n.º 3, 2010, pp. 165-224, doi: 10.1561/08000000014. [https://michael-lechner.eu/ml\\_pdf/journals/2011\\_Lechner\\_DiD\\_2011\\_ECO%200403%20Lechner\\_darf%20aufs%20Netz.pdf](https://michael-lechner.eu/ml_pdf/journals/2011_Lechner_DiD_2011_ECO%200403%20Lechner_darf%20aufs%20Netz.pdf)

Este artículo analiza en profundidad el enfoque de la DED y algunas de las principales dificultades para ponerlo en práctica. También presenta aplicaciones de la DED, como las aplicaciones no lineales y el emparejamiento de puntuaciones de propensión con DED.

- Fredriksson, A. y De Oliveira, G. M. (2019): «Impact evaluation using Difference-in-Differences», *RAUSP Management Journal*, vol. 54, n.º 4, pp. 519-532, doi: 10.1108/RAUSP-05-2019-0112. <https://www.emerald.com/insight/content/doi/10.1108/RAUSP-05-2019-0112/full/pdf?title=impact-evaluation-using-difference-in-differences>

Un resumen de los métodos de DED, con recomendaciones prácticas

- Card, D. y Krueger, A. B. (1994): «Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania», *American Economic Review*, 84 (4), pp. 774-775. <https://davidcard.berkeley.edu/papers/njmin-aer.pdf>

Ejemplo clásico de aplicación de una técnica de diferencia en diferencias.

### Variables instrumentales

- Morgan, S. L. y Winship, C. (2014): *Counterfactuals and Causal Inference: Methods and*

*Principles for Social Research*, 2.<sup>a</sup> ed., Cambridge University Press, Cambridge y Nueva York.

Una especie de revisión técnica de los métodos de EIC utilizando terminología sociológica. El capítulo 9, «Instrumental Variable Estimators of Causal Effects» «Estimadores con variables instrumentales de los efectos causales» (pp. 291-324) ofrece una visión general de la lógica y los procedimientos de la estimación de VI.

- Kuhn, A.; Wuellrich, J. P., y Zweimüller, J. (2010): *Fatal Attraction? Access to Early Retirement and Mortality*, documento de reflexión del IZA n.º 5160, Forschungsinstitut zur Zukunft der Arbeit, Bonn. <http://ftp.iza.org/dp5160.pdf>

Se sirve de la variación regional del cambio en la edad de jubilación en Austria como variable instrumental en un estudio del efecto de la jubilación anticipada en la salud de los trabajadores.

- Galiani, S.; Rossi, M. A., y Schargrodsy, E. (2011): «Conscription and crime: Evidence from the Argentine draft lottery», *American Economic Journal: Applied Economics*. <https://www.aeaweb.org/articles?id=10.1257/app.3.2.119>

Uso innovador e ilustrativo de un sorteo aleatorizado (para reclutar para el servicio militar) como variable instrumental. Estudio muy fácil de leer, prototipo de una línea de investigación que utiliza los procedimientos de sorteo como instrumentos.

### Emparejamiento

- Heinrich, C.; Maffioli, A., y Vázquez, G. (2010): «A Primer for Applying Propensity Score Matching», *Impact-Evaluation Guidelines Technical Notes*, n.º IDB-TN-161, Banco Interamericano de Desarrollo, Washington. <https://publications.iadb.org/publications/english/document/A-Primer-for-Applying-Propensity-Score-Matching.pdf>

Al igual que la guía sobre la regresión discontinua que figura a continuación, se ha escrito pensando en los responsables de evaluaciones bien informados.

- Caliendo, M. y Kopeinig, S. (2008): «Some practical Guidance for the implementation of propensity score matching», *Journal of Economic Surveys*, vol. 22, 1.<sup>a</sup> ed., pp. 31-72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>

Un texto clásico sobre las decisiones que deben tomarse en la aplicación del EPP (en términos de estimación, algoritmos de emparejamiento, evaluación de la calidad del emparejamiento, sensibilidad de los efectos del tratamiento estimados).

- Harris, H. y Horst, S. J. (2016): «A Brief Guide to Decisions at Each Step of the Propensity Score Matching Process», *Practical Assessment, Research, and Evaluation*, vol. 21, artículo 4, doi: <https://doi.org/10.7275/yq7r-4820> Disponible en: <https://scholarworks.umass.edu/pare/vol21/iss1/4>.

Similar al artículo anterior.

### Diseño de regresión discontinua

- Jacob, R.; Zhu, P.; Somers, M. A., y Bloom, H. (2012): *A Practical Guide to Regression Discontinuity*, MDRC, Nueva York. [https://www.mdrc.org/sites/default/files/RDD%20Guide\\_Full%20rev%202016\\_0.pdf](https://www.mdrc.org/sites/default/files/RDD%20Guide_Full%20rev%202016_0.pdf)

Análisis excepcionalmente accesible y exhaustivo de la metodología de discontinuidad de la recesión que incluye una bibliografía cuidadosamente seleccionada

- Lee, D. S. y Lemieux, T. (2009): *Regression discontinuity designs in economics*, documento de trabajo de la NBER n.º 14723, National Bureau of Economic Research, Cambridge, Massachusetts, 2009. [http://www.nber.org/papers/w14723.pdf?new\\_window=1](http://www.nber.org/papers/w14723.pdf?new_window=1)

El documento es una especie de «guía del usuario» sobre la regresión discontinua (RD) que ofrece una descripción de la lógica del método y explica diferentes maneras de estimar los diseños de RD.

## Anexo 2. Propuesta de esquema para un curso sobre EIC

Un curso introductorio sobre la EIC dirigido a las AG y los funcionarios podría abarcar los siguientes aspectos:

- Introducción a los enfoques de evaluación: evaluaciones de procesos (por qué son importantes y qué relación tienen con las EIC) y evaluaciones del impacto (enfoque contrafactual y enfoque basado en la teoría).
- ¿Qué son las EIC? ¿Qué preguntas de evaluación pueden responder unas EIC sólidas? ¿Para qué criterios de evaluación proporcionan las pruebas principales las EIC? ¿Por qué son importantes las EIC?
- ¿Cómo funcionan? (Grupo tratados y grupo de control, comparaciones «antes y después», principios de inferencia).
- Resumen de las metodologías de EIC (características, puntos fuertes y débiles de cada método):
  - o Ensayo aleatorio controlado
  - o Emparejamiento
  - o Diferencia en diferencias
  - o Diseño de regresión discontinua
  - o Variable instrumental
  - o Indicación de otros métodos posibles (p. ej., técnicas de análisis espacial)
- Requisitos en materia de datos
  - o Requisitos del grupo tratado y el grupo de control
  - o Posibles fuentes (seguimiento del FSE, conjuntos de datos administrativos, encuestas)
  - o Normas del RGPD
- Resumen de los pasos para realizar una EIC:
  - o Planificar las EIC (plan de evaluación, viabilidad de una EIC)
  - o Encargar las EIC (términos de referencia, criterios de selección y métodos)
  - o Gestionar las EIC (verificación de los resultados e interacción con el evaluador)
  - o Difundir las conclusiones de las EIC (tipos de público y canales de comunicación)

Un curso estructurado de este modo debe incluir ejemplos prácticos que se ampliarían en talleres y tendría una duración de entre dos y tres días. Una estrategia para impartir un curso como este sería adoptar una metodología de aprendizaje basada en problemas y utilizar una medida política del programa como caso concreto a desarrollar.

Anexo 3. Evaluaciones de impacto contrafactuales: ejemplos mencionados en la Guía

Cuadro 6 Características de las EIC mencionadas en la Guía como ejemplos

País	Título y año	Política evaluada y período del tratamiento observado	Método	Grupo tratado y de control: definición y tamaño	Variables de resultado	período de observación	Datos
Suecia	Efectos de reforzar el apoyo a los jóvenes demandantes de empleo: resultados de la evaluación de Futuro Joven (2019)	Mayor apoyo del servicio público de empleo a los demandantes de empleo  De junio de 2017 a enero de 2018	Ensayo aleatorio controlado (EAC)	GT: jóvenes de entre dieciocho y veinticuatro años asignados aleatoriamente para recibir el apoyo reforzado  GC: jóvenes de entre dieciocho y veinticuatro años asignados aleatoriamente para recibir el apoyo habitual  GT: 2 792; GC: 1897	- Proporción de desempleados.  - Porcentaje de salidas para trabajar. - Promedio de días en situación de desempleo.	Doce meses después de acceder al tratamiento	- Registros públicos de empleo para las variables de resultado. - Encuesta a los funcionarios de empleo y a los intermediarios para medir la intensidad del apoyo prestado.
Italia (Las Marcas)*	Impactos de las intervenciones del FSE financiadas en 2014-2020 para parados de larga duración en la región de Las Marcas (2020)	Prácticas, experiencias laborales, becas de empleo, vales de formación para desempleados (entre ellos, parados de larga duración)  De 2017 a julio de 2019	Emparejamiento por puntuaciones de propensión (EPP)	GT: parados de larga duración tratados  GC: parados de larga duración en el período 2016-2018 matriculados en los SPE  GT: 526 para prácticas, 1 058 para becas de empleo, 236 para experiencias laborales en los municipios y 241 para vales de formación; GC: 77255	- Probabilidad de tener trabajo en un determinado momento después de las intervenciones. - Probabilidad de tener un empleo con un contrato indefinido. - Número de días trabajados un período después de las intervenciones.	Seis, nueve, doce, quince y dieciocho meses después de acceder a las intervenciones	- Datos de seguimiento del FSE para el grupo tratado. - Datos administrativos de los SPE sobre las personas registradas. - Datos administrativos sobre los contratos laborales activados en la región («Comunicazioni Obbligatorie» o «COB»).
Polonia (Lubelskie)*	Evaluación de las medidas dirigidas a los trabajadores de más edad en Lubelskie financiadas por el FSE en el período de programación 2007-2013 (2015)	Varios tipos de políticas activas del mercado de trabajo  2007-2013	EPP + diferencia en diferencias (DED)	GT: muestra de desempleados mayores de cincuenta años que recibieron apoyo  GC: desempleados mayores de cincuenta años que no participaron en ninguna medida y que estaban inscritos en una oficina regional de empleo en el momento en que los miembros del grupo tratado participaron en el programa evaluado  GT: 145; GC: 67102	- Probabilidad de dejar de estar inscritos como desempleados.	Seis, doce y dieciocho meses después de acceder a las intervenciones	- Datos administrativos de los SPE.

País	Título y año	Política evaluada y período del tratamiento observado	Método	Grupo tratado y de control: definición y tamaño	Variables de resultado	período de observación	Datos
Letonia	«Vocational Training for Unemployed Youth in Latvia: pruebas procedentes de un diseño de regresión discontinua (2018)	Formación profesional dirigida a niñas de entre quince y veintinueve años 2014-2015	Diseño de regresión discontinua (DRD)	GT: desempleados que participasen en la formación profesional en el plazo de un año a partir de la fecha de inscripción GC: desempleados inscritos entre junio de 2013 y diciembre de 2015 que no participan en ninguna medida GT: 898; GC: 10717	- Probabilidad de tener trabajo en un determinado momento después de las intervenciones. - Ingresos.	Entre doce y treinta y seis meses después de la inscripción	- Datos administrativos de la Agencia Estatal de Empleo de Letonia. - Información sobre los participantes y los no participantes inscritos como desempleados en fechas concretas. - Datos administrativos de la Administración Tributaria del Estado. - Información sobre la situación laboral en diferentes fechas y sobre los ingresos de las personas.
Alemania	Efectos en el empleo de la formación lingüística para inmigrantes desempleados (2021)	Cursos de formación lingüística (+ experiencia laboral para algunos de los tratados) para demandantes de empleo migrantes 2014	Variables instrumentales (VI)	GT: participantes migrantes que iniciaron la formación lingüística con fines profesionales en 2014 GC: muestra aleatoria de no participantes con al menos un período de ciudadanía no alemana que no participaron en ninguna medida de la política activa del mercado de trabajo o curso de integración GT: 8 968; GC: 26463	- Probabilidad de tener trabajo (empleo regular, empleo regular a tiempo completo, empleo regular durante más de seis meses, empleo con ingresos superiores al umbral de riesgo de pobreza, empleo cualificado). - Ingresos diarios y acumulados procedentes del empleo regular. - Días acumulados de empleo regular.	Veinticuatro meses después de acceder a las intervenciones	- Biografías laborales integradas (BLI) para la determinación de ambos grupos y para las variables de resultado. - Datos recopilados retrospectivamente durante reuniones con los demandantes de empleo y los asistentes sociales (Werdegangshistorie, WGH), con el fin de obtener más información sobre los episodios de trabajo por cuenta propia o permiso parental y sobre la educación y el empleo en el extranjero.
Italia (Trento)	«Are Vocational Training Programmes Worth Their Cost? Pruebas de un análisis coste-beneficio (2018)	Cursos de formación profesional para desempleados 2010-2011	EPP	GT: desempleados que participaron en cursos de formación profesional, financiados con recursos provinciales y por el FSE GC: personas inscritas como desempleadas en la fecha de inicio de cada curso formativo en los registros de los SPE locales GT: 818 (cursos provinciales), 114 (cursos del FSE); Grupo de control: 17 236 (relacionados con los cursos provinciales), 1 152 (relacionados con cursos del FSE)	- Probabilidad de tener trabajo. - Ingresos.	Tres, seis, doce, dieciocho, veinticuatro y treinta y seis meses después de acceder a las intervenciones	- Datos de seguimiento del FSE para el grupo tratado. - Datos administrativos de los SPE sobre las personas registradas. - Datos administrativos sobre los contratos laborales activados en la región («Comunicazioni Obbligatorie» o «COB») - Datos administrativos sobre las declaraciones fiscales del INPS. - Datos administrativos del Departamento de Educación de la provincia.

País	Título y año	Política evaluada y período del tratamiento observado	Método	Grupo tratado y de control: definición y tamaño	Variables de resultado	período de observación	Datos
Alemania	¿Mejoran los planes de creación de empleo la integración social y el bienestar de los parados de larga duración? (2020)	Plan de creación de empleo (empleos subvencionados con una duración de hasta treinta y seis meses) para personas vulnerables (parados de larga duración)  De 2015 a junio de 2017	EPP	GT: muestra de participantes que respondieron a todo el cuestionario  GC: no participantes inscritos en los SPE y con los mismos criterios de elegibilidad del grupo de control, que respondieron a todo el cuestionario  GT: 2 531 en la ronda 1, 1 191 en la ronda 2, 450 en la ronda 3; GC: Igual que arriba	- Probabilidad de tener trabajo. - Mediciones subjetivas de la satisfacción vital, la salud mental, la pertenencia social y el estatus social.	Siete, dieciocho y veintinueve meses después de la entrada en el programa	- Biografías laborales integradas (BLI) para la determinación de ambos grupos y para las variables de resultado. - Encuesta sobre variables de resultado subjetivos (efectos directos difíciles de cuantificar).
España (Asturias)*	Evaluación del PO del FSE de Asturias (2019)	Métodos de enseñanza para grupos de estudiantes en riesgo de abandono escolar prematuro (segundo y tercer curso de enseñanza secundaria)  2016-2017	EPP	GT: estudiantes que participaron en las intervenciones  GC: estudiantes comparables en lo que respecta a los criterios de elegibilidad  GT: 1 053; GC: 3852	- Probabilidad de pasar al siguiente curso.	Doce meses después de acceder al tratamiento	- Datos administrativos sobre educación: base de datos SAUCE (Ministerio de Educación).
Portugal	Evaluación del sistema de becas para la enseñanza superior dirigidas a los estudiantes menos favorecidos en Portugal (2020); pendiente de publicación	Becas para estudiantes de enseñanza superior procedentes de familias con bajos ingresos  2021-2018	DRD	GT: estudiantes que solicitaron matricularse en el primer curso de un máster o licenciatura  GC: estudiantes no elegibles porque sus ingresos superan el umbral establecido  GT: 130 602; GC: 25400	- Probabilidad de seguir matriculado al final del primer año. - Número de créditos aprobados al final del primer año. - Probabilidad de graduarse. - Número de años que se tarda en graduarse. - Nota final en la graduación.	Entre doce y cuarenta y ocho meses después de recibir la beca	- Datos administrativos sobre los estudiantes que solicitaron la beca (DGES). - Datos administrativos con información sobre la progresión y la carrera académica (DGEEC).
Polonia (Podlaskie)*	Evaluación del impacto en la popularización de la educación profesional del apoyo del PO llevado a cabo en Podlaskie en 2014-2020 (2020)	Proyecto del FSE Buena Profesión, Vida Fantástica, destinado a promover la educación y la formación profesionales y dirigido a los centros de enseñanza secundaria de primer ciclo  2017-2019	EPP	GT: muestra de estudiantes que participaron  GC: estudiantes con características similares a las del grupo tratado, pero que no participaron en el proyecto  GT: 200; GC: 384	- Asistencia a una escuela técnica en el momento del estudio. - Asistencia a una escuela de comercio de primer grado en el momento del estudio. - Asistencia a una escuela técnica o una escuela de comercio de primer grado en el momento del estudio. - Solicitud de matriculación en un centro de enseñanza secundaria, una escuela técnica, una escuela de comercio de primer grado, una escuela donde se impartiese formación profesional.	Entre doce y treinta y seis meses después de las intervenciones	- Datos de seguimiento. - Censo nacional de población (PESEL) para la selección del grupo de control. - Encuesta para medir las variables de resultado.

País	Título y año	Política evaluada y período del tratamiento observado	Método	Grupo tratado y de control: definición y tamaño	Variables de resultado	período de observación	Datos
Italia (Piamonte)*	Efectos en el empleo del régimen de vales para los servicios de empleo («Buono servizi lavoro» o «BSL», financiado por el PO del FSE en Piamonte en 2014-2020) (2019)	Vales para servicios de empleo (orientación + formación + etapa) para parados vulnerables con un período de desempleo de al menos seis meses 2018	EPP	GT: personas desempleadas tratadas  GC: desempleados inscritos en los SPE con características similares (treinta años y desempleados durante seis meses)  GT: 8 125; GC: 130000	- Probabilidad de tener trabajo. - Probabilidad de tener un empleo con un contrato indefinido.	Seis, doce y dieciséis meses después de las intervenciones	- Datos de seguimiento del FSE para el grupo tratado. - Datos administrativos de los SPE sobre las personas registradas. - Datos administrativos sobre los contratos laborales activados en la región («Comunicazioni Obbligatorie» o «COB»).
Alemania*	Evaluación del programa de integración de los parados de larga duración en Alemania en 2014-2020 (2019)	Medidas de integración para parados de larga duración durante veinticuatro meses o más (empleos subvencionados) 2015-2017	DED (se estimó la intención de tratamiento)	GT: parados de larga duración potencialmente elegibles durante el período del programa (a partir de 2015)  GC: personas que cumplían los criterios de elegibilidad antes de la ejecución del programa (2010-2012)  GT y GC: 134 515 personas en 237 874 episodios de desempleo (35 % para el GT y 65 % para el GC)	- Probabilidad de tener trabajo (empleo con cotizaciones a la Seguridad Social).	Veinticuatro meses después de las intervenciones	- Datos de seguimiento administrativo de la Oficina Federal de Administración (base de datos ZUWES). - Biografías laborales integradas (BLI) para las variables de resultado.
Alemania (Baden-Württemberg)*	Segundo informe de evaluación del objetivo específico A1.1 en el marco del PO del FSE llevado a cabo en Baden-Wurtemberg en 2014-2020 (2020)	Medidas para integrar a los desempleados, y en especial a los parados de larga duración y otras personas vulnerables De marzo de 2016 a diciembre de 2017	EPP	GT: participantes identificados en la base de datos de las BLI GC: personas que cumplen los criterios de elegibilidad en el mismo período  GT: 1 578 (de un total de 1 800); GC: N/A	- Probabilidad de tener trabajo (empleo con cotizaciones a la Seguridad Social). - Probabilidad de seguir percibiendo una prestación por desempleo.	Quince meses después de las intervenciones	- Datos de seguimiento. - Biografías laborales integradas (BLI) para las variables de resultado.

Nota: GT = grupo tratado, GC = grupo de control.

Nota: \* señala las evaluaciones especificadas en el Servicio de Asistencia.

## CÓMO PONERSE EN CONTACTO CON LA UE

### En persona

Hay cientos de centros de información Europe Direct repartidos por toda la Unión Europea. Puede encontrar la dirección del centro más cercano en:

[https://europa.eu/european-union/contact\\_es](https://europa.eu/european-union/contact_es)

### Por teléfono o por correo electrónico

Europe Direct es un servicio que responde a sus preguntas sobre la Unión Europea. Puede ponerse en contacto con este servicio de las siguientes formas:

- mediante el número de teléfono gratuito: 00 800 6 7 8 9 10 11 (algunos operadores pueden cobrar por las llamadas);
- mediante el número de teléfono normal: +32 22999696; o
- por correo electrónico: [https://europa.eu/european-union/contact\\_es](https://europa.eu/european-union/contact_es)

## BUSCAR INFORMACIÓN SOBRE LA UE

### En línea

Encontrará información sobre la Unión Europea en todas las lenguas oficiales de la UE en el sitio web Europa: [https://europa.eu/european-union/index\\_es](https://europa.eu/european-union/index_es)

### Publicaciones de la Unión Europea

Puede descargar o solicitar publicaciones gratuitas y de pago de la UE en: <https://op.europa.eu/es/publications>. Si desea obtener varios ejemplares de las publicaciones gratuitas, póngase en contacto con Europe Direct o con su centro de información local (véase [https://europa.eu/european-union/contact\\_es](https://europa.eu/european-union/contact_es)).

### Derecho de la Unión y documentos conexos

Para acceder a la información jurídica de la Unión Europea, incluido todo el Derecho de la Unión desde 1952, en todas las versiones lingüísticas oficiales, consulte el sitio web EUR-Lex: <https://eur-lex.europa.eu/homepage.html?locale=es>

### Datos abiertos de la Unión Europea

El portal de datos abiertos de la UE (<http://data.europa.eu/euodp/es>) permite acceder a conjuntos de datos de la UE. Los datos pueden descargarse y reutilizarse gratuitamente con fines comerciales o no comerciales.

