



With financial support from the DAPHNE Programme of the European Union.

Posibilidad de estudios europeos de evaluación multisitio/multipaís de programas para agresores de violencia de género

Documento de trabajo 3 del proyecto de Daphne III "IMPACT: Evaluación de programas europeos para agresores"

This publication has been produced with the financial support of the DAPHNE Programme of the European Union. The contents of this website are the sole responsibility of Dissens - Institute for Education and Research e.V. and the project partners and can in no way be taken to reflect the views of the European Commission.

Christian Scambor, Katarzyna Wojnicka & Elli Scambor



Lista de abreviaturas frecuentes

TCC ... Terapia cognitivo-conductual

CCR ... Respuesta coordinada de la comunidad (Co-ordinated community response)

DVPP ... Programa para agresores de violencia de género (Domestic violence perpetrator programme)

ECA ... Ensayo controlado aleatorio

T0 ... Momento de la observación: en la derivación; fase de acogida

T1 ... Momento de la observación: "Pre", al principio de la intervención (p. ej., trabajo en grupo)

T2 ... Momento de la observación: "Durante", durante la intervención (p. ej., trabajo en grupo)

T3 ... Momento de la observación: "Post", al final de la intervención (p. ej., trabajo en grupo)

T4 ... Momento de la observación: "Seguimiento", un período de tiempo tras finalizar la intervención (p. ej., 6 meses tras finalizar el trabajo en grupo)

IMPACT ... Proyecto de Daphne III "Evaluación de programas europeos para agresores"

WAVE ... Women Against Violence in Europe (organización internacional)

WWP ... Trabajo con agresores (Work with perpetrators)

WWP – EN ... Trabajo con agresores – Red Europea (Work With Perpetrators – European Network) (organización internacional)

1. Introducción

Según los expertos en la materia, la evaluación de los programas para agresores es muy importante para los profesionales y el personal de los programas a fin de monitorizar y mejorar la calidad de su trabajo, así como para ayudar a los dirigentes políticos a tomar decisiones estratégicas (Geldschläger et al., 2014; Hester, Lilley, O'Prey, Budde, 2014). Por desgracia, a día de hoy, solo se han publicado unos pocos proyectos de evaluación sobre programas europeos en revistas científicas, aunque en su revisión Hester et al. (2014) han encontrado como mínimo 65 estudios nacionales (publicados en revistas o como "literatura gris", informes u otros documentos no publicados como libros o en revistas). No obstante, los estudios parecen parciales. La mayoría de proyectos se llevaron a cabo en un único sitio y solo unos pocos se podrían describir como proyectos de evaluación multisitio. En general, los informes están compuestos por estudios nacionales, con diferentes enfoques, diferentes teorías de cambio, métodos, instrumentos, etc. Los resultados de Hester et al. (2014) dan la impresión de que la evaluación de los programas europeos para agresores es un terreno disperso y complejo, y resulta complicado extraer conclusiones metaanalíticas sobre el trabajo con agresores. Esta situación provoca que la especificidad europea realmente no sea tomada en consideración en los debates internacionales sobre investigación de evaluación de programas para agresores; sin embargo, "...las evidencias existentes (p. ej., de estudios norteamericanos) no se pueden extrapolar fácilmente a un contexto europeo" Hester et al., 2014, pág. 3).

Por consiguiente, los objetivos principales del proyecto "IMPACT: Evaluación de programas europeos para agresores", financiado por la Comisión Europea (programa Daphne III), se han definido como sigue:

- a) Suplir la falta de conocimientos en relación con las evaluaciones de programas para agresores en Europa, y
- b) Ofrecer soluciones para una armonización de la medición de resultados (como un primer paso para la unificación de la metodología de evaluación)

por medio de:

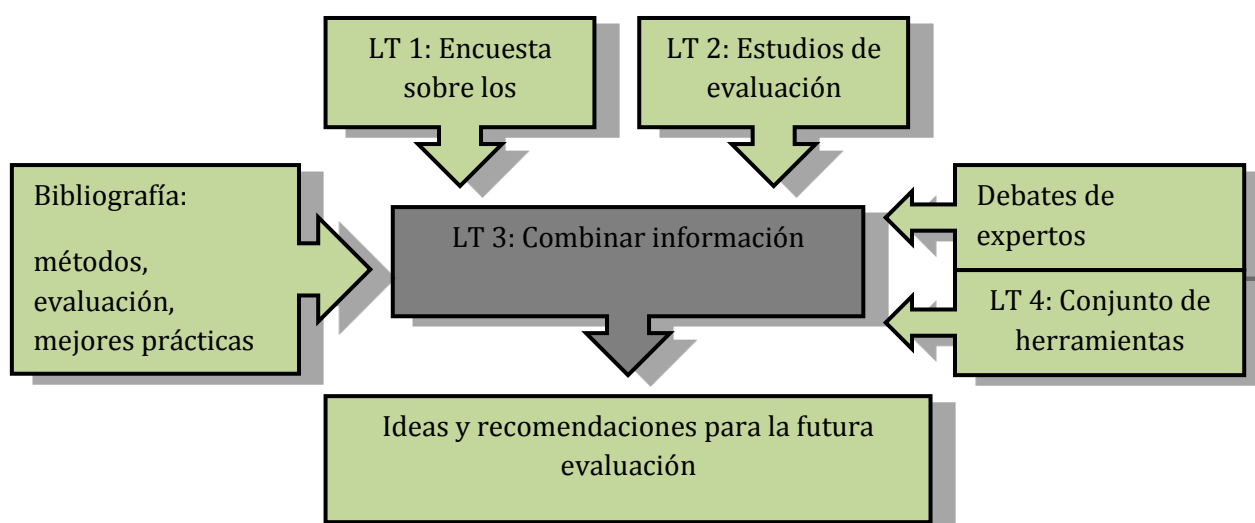
- Proporcionar una visión general y un análisis de la práctica actual de monitorización de resultados en programas europeos para agresores (Línea de trabajo 1)
- Proporcionar una visión general y un análisis de los estudios de investigación que evalúan los programas para agresores (Línea de trabajo 2)
- Identificar posibilidades y obstáculos para estudios europeos de investigación de resultados multipaís (Línea de trabajo 3)
- Desarrollar un conjunto de herramientas y guías de buenas prácticas para la medición de los resultados (Línea de trabajo 4)

Este documento de trabajo presenta la metodología y los resultados de las actividades llevadas a cabo en la Línea de trabajo 3. La identificación de las posibilidades y los obstáculos para realizar estudios europeos de investigación multisitio/multipaís se basó en:

- Los resultados (provisionales y definitivos) de la Línea de trabajo 1: visión general de la práctica de la medición de resultados (encuesta)
- Los resultados (provisionales y definitivos) de la Línea de trabajo 2: visión general de los estudios europeos de evaluación
- Revisión de la bibliografía relevante en el ámbito de la metodología de investigación y evaluación
- Entrevistas y debates realizados con personas expertas y profesionales del área de trabajo con agresores y servicios que cooperan con ellos.

Sobre la base de los resultados de la encuesta de la Línea de trabajo 1, las preguntas que se han formulado son: ¿hasta qué punto ya se están usando métodos e instrumentos comparables para medir los resultados o evaluar en diferentes organizaciones?, y ¿qué requisitos nacionales e institucionales se pueden identificar que promuevan u obstruyan proyectos multinacionales de evaluación? A partir del conocimiento generado por la revisión de la investigación en la Línea de trabajo 2, se han recopilado metodologías para proyectos de evaluación que muestran una elevada probabilidad de transferibilidad a diversos contextos nacionales. En este sentido, la Línea de trabajo 3 se centra en la viabilidad de los proyectos de evaluación que incluyen más de un sitio y más de un país. Las actividades de la Línea de trabajo 3 se han vinculado con la Línea de trabajo 4 con el fin de informar sobre el desarrollo del producto “caja de herramientas”, de tal modo que este producto contribuya a una armonización de las prácticas relacionadas con la medición de los resultados en los programas y, por consiguiente, ayude a poner los cimientos para proyectos de evaluación multipaís.

Figura 1. Posición y funcionamiento de la Línea de trabajo 3 dentro del proyecto IMPACT.



Nota. LT ... Línea de trabajo. DVPP ... Programas para agresores de violencia de género (Domestic violence perpetrator programmes)

En este documento de trabajo de la Línea de trabajo 3, se resumen con énfasis las aportaciones específicas de los expertos y expertas, así como nuestras conclusiones para futuros proyectos de evaluación en el ámbito de trabajo con agresores. Se proporciona un análisis más detallado de las cuestiones debatidas en este documento de trabajo en dos artículos científicos independientes.

Los autores desean agradecer a los y las participantes del proyecto IMPACT, en particular a Heinrich Geldschläger y Marianne Hester y a sus equipos, por haber proporcionado importantes resultados provisionales que ayudaron a avanzar en la Línea de trabajo 3. También nos gustaría dar las gracias a todos los expertos y expertas que participaron en las entrevistas y los debates.

2. Metodología

El principal objetivo de la Línea de trabajo 3 era identificar posibilidades y obstáculos para proyectos europeos de evaluación multisitio/multipaís de programas para perpetradores de violencia contra mujeres, niños y niñas (DVPP – Programas para agresores de violencia de género). El enfoque empleado para identificar modelos de evaluación con una posible aplicabilidad en diferentes contextos y países europeos consistió en:

- (1) El análisis de la bibliografía existente sobre evaluación de programas para agresores (tanto en contextos europeos como no europeos), así como otras publicaciones relevantes.
- (2) La utilización de los resultados de la Línea de trabajo 1 y la Línea de trabajo 2, junto con fases de análisis adicional de los datos recopilados en ambas Líneas de trabajo.
- (3) La realización de entrevistas y debates con personas expertas y profesionales internacionales que trabajan con agresores de violencia de género.

Cabe subrayar que la Línea de trabajo 3 se centraba en (a) los *resultados* (no en el *proceso* u otros aspectos de la evaluación del trabajo de programas para agresores) y (b) en la *evaluación* y, por consiguiente, en los proyectos de *investigación*, lo que va más allá de la medición de los resultados en la práctica diaria de programas para agresores (no obstante, estos aspectos están relacionados y, a menudo, tienen que considerarse conjuntamente: los resultados dependen de los procesos subyacentes, e, idealmente, las prácticas para la medición de los resultados son compatibles o incluso forman parte de la metodología que se emplea en los proyectos de evaluación.)

2.1 Análisis de la bibliografía

Las conclusiones de las publicaciones permitieron una visión general de los modelos, instrumentos, estrategias y métodos de evaluación más frecuentes, y apoyaron la formulación de las preguntas para las entrevistas con grupos focales y debates con personas expertas. Entre otras, han sido fuentes importantes de información: *Batterer Intervention Systems: Issues, Outcomes, and Recommendations* (Gondolf, 2002); *Domestic Violence*

Perpetrator Programs in Europe , Part I: A Survey of Current Practice; Part II: A Systematic Review of the State of Evidence (Hamilton, Koehler & Lösel, 2013; Akoensi, Koehler, Lösel & Humphreys, 2013); *Changing Violent Men* (Dobash, Dobash, Cavanagh & Lewis, 2000); *Violent Men and Violent Context* (Dobash & Dobash, 1998); *Evaluating Criminal Justice Interventions for Domestic Violence* (Dobash & Dobash, 2000); *Rethinking Survey Research on Violence Against Women* (Johnson, 1998); *Description of a Voluntary Attendance Program for Abusive Men* (Payarola, 2011); *Interviewing Violent Men: Challenge or Compromise?* (Cavanagh & Lewis, 1996); *Men Researching Violent Men: Epistemologies, Ethics and Emotions in Qualitative Research* (Cowburn, 2013); *Encountering Violent Men: Strange and Familiar* (Gottzen, 2013); *Batterer Intervention Programs: A Report From the Field* (Price & Rosenbaum, 2009); *Counting Men: Qualitative Approaches to the Study of Men and Masculinities* (Patulny & Pini, 2013); *World Report on Violence and Health* (WHO, 2002); así como dos documentos de trabajo del proyecto IMPACT: *Outcome Measurement in European Perpetrator Programmes: A Survey* (Geldschläger, Ginés, Nax & Ponce, 2014); *Overview and Analysis of Research Studies Evaluating European Perpetrator Programmes* (Hester, Lilley, O'Prey & Budde, 2014).

Otra fuente de información importante para la Línea de trabajo 3 fue la bibliografía metodológica sobre evaluación (p. ej., Bamberger, Rugh & Mabry, 2006; Rossi, Lipsey & Freeman, 2004; Shaddish, Cook & Campbell, 2002; Stufflebeam & Shinkfield, 2007). Se proporcionaron glosarios y terminología específica sobre evaluación al grupo del proyecto con el fin de emplear una terminología común.

Paralelamente al proyecto IMPACT, se ha llevado a cabo un importante proyecto de evaluación en el Reino Unido, denominado MIRABAL¹. Este proyecto pretende responder algunas preguntas que también tienen relevancia para el proyecto IMPACT. El socio del proyecto, Respect (Reino Unido), desempeña un papel destacado en el proyecto MIRABAL, por lo que ha sido posible la transferencia de importantes desarrollos entre los dos proyectos. Los contactos y los debates entre los dos proyectos han sido una valiosa fuente de información (p. ej., en el taller realizado en Bristol, Reino Unido, en 2013, donde Nicole Westmarland presentó desarrollos recientes y la metodología del proyecto MIRABAL). Las publicaciones hasta la fecha relacionadas con el proyecto MIRABAL han constituido una valiosa fuente de información (Alderson, Kelly & Westmarland, 2013; Alderson, Westmarland & Kelly, 2012; Downes, Kelly & Westmarland, 2014; Phillips, Kelly & Westmarland, 2013; Westmarland & Kelly, 2012; Westmarland, Kelly & Chalder-Mills, 2010). Los resultados del proyecto MIRABAL se presentarán a principios de 2015 y, sin duda, dará nuevos impulsos para proyectos de evaluación multisitio en el futuro. Es muy significativo que el experto norteamericano Edward Gondolf (2012) haya tomado el proyecto MIRABAL como ejemplo de una nueva tendencia en la investigación sobre evaluación (pág. 81), que denomina “evaluaciones conducidas por profesionales” y que pretenden superar algunas de las deficiencias de los enfoques académicos –a veces inadecuados– respecto de la evaluación de programas para agresores.

¹ <http://respect.uk.net/research/our-research-partnerships/mirabal-multi-site-evaluation-project/>
<https://www.dur.ac.uk/criva/projectmirabal/>

2.2. Aplicación de los resultados de la Línea de trabajo 1

La Línea de trabajo 3 se ha centrado en los modelos de evaluación, es decir, diseños, instrumentos y procedimientos de evaluación. En la Línea de trabajo 1 se llevó a cabo una encuesta para recabar información sobre programas europeos en el ámbito de trabajo con agresores de violencia de género, así como sus prácticas en cuanto a la medición de los resultados y la la evaluación de su trabajo. En consecuencia, esperábamos obtener prácticas prometedoras que ya se hubieran implementado o usado en proyectos de evaluación, y que mostraran un cierto grado de transferibilidad. Aunque los autores afirman que "... el énfasis no recae en la investigación de los resultados científicos, sino en el seguimiento día a día de los resultados que los programas para agresores realizan regularmente como parte de sus servicios" (Geldschläger et al., 2014, pág. 4), esperábamos obtener sugerencias de la Línea de trabajo 1 para la planificación de futuros estudios de evaluación. Repasando los resultados de la Línea de trabajo 1, queríamos explorar si existía un denominador común de elementos que ya estuvieran implementados y enfoques que se pudieran adoptar para futuros estudios.

2.2.1. Aspectos relevantes de la Línea de trabajo 1

En la encuesta de la Línea de trabajo 1 se recabaron datos de 134 programas de 22 países, y los resultados de la encuesta se han publicado en un documento de trabajo (Geldschläger et al., 2014). Los resultados y las consecuencias más importantes para la Línea de trabajo 3 se resumen a continuación. Para obtener información más detallada, deberá consultarse Geldschläger et al. (2014).

Derivaciones

La mayoría de los programas para agresores trabajan con hombres que acuden por propia iniciativa, así como hombres derivados por otros servicios. Una parte de los programas excluye algún tipo de derivaciones o se concentra en un determinado tipo de derivaciones (un 22 % de los programas trabajan únicamente con hombres derivados por orden judicial; un 19 % no trabajan nunca con hombres por orden judicial; Geldschläger et al., 2014, pág. 12). Muchos programas se centran en hombres derivados por orden judicial o derivados por otros servicios, pero la mayoría de programas trabajan con varias formas de entrada en el programa. Los futuros estudios de evaluación multisitio deberían tener en cuenta esta diversidad.

Cooperación

Prácticamente todos los programas (98 %) de la encuesta colaboran con algún servicio, entre los más relevantes: servicios sociales, policía, unidades de protección de las víctimas, tribunales penales, servicios de tratamiento de alcoholismo y abuso de sustancias. Un 75 % de los programas afirman formar parte de una alianza interinstitucional contra la violencia de género, y una parte de los programas ofrecen apoyo a las víctimas dentro de su organización. Por consiguiente, podemos concluir que los programas de la encuesta están siguiendo mayoritariamente un "enfoque de sistema" y en una variedad de países

diferentes. La clase y la calidad de las redes puede diferir según los países y es algo que se debe tener en cuenta en los diseños de evaluaciones multisitio. No obstante, esta conclusión también significa que es posible concentrarse en esos programas que forman parte de una red coordinada de instituciones y servicios en primera instancia. Los programas que difieren de ese enfoque también son de interés aunque podrían formar parte de un programa de evaluación y se podrían estudiar en una segunda fase, puesto que requieren comparaciones y diseños diferentes.

Acogida

Prácticamente todos los programas (95 %) cuentan con algún tipo de procedimiento de acogida: desde un contacto inicial hasta fases más prolongadas de asesoramiento o evaluación individual durante la fase de acogida. Básicamente, se trata de una condición útil para la realización de proyectos de evaluación. Para muchos programas, la implementación de algún tipo de recogida inicial de datos constituye un procedimiento “normal”. Dos tercios de los programas emplean algún tipo de instrumento durante la acogida, pero no existen instrumentos que predominen. Se mencionan varios cuestionarios, manuales y otros instrumentos. En este aspecto, se requiere cierta armonización, pero también es importante que la mayoría de proyectos están acostumbrados a trabajar con algún instrumento. Los estudios de evaluación en el futuro pueden basarse en esta competencia.

La mayoría de programas pueden aplicar algún tipo de criterio de inclusión/exclusión, basándose en el procedimiento de acogida (p. ej., tres cuartas partes mencionan “alcoholismo/abuso de otras sustancias” y “trastornos mentales” como condiciones excluyentes). Esto significa que la mayoría de programas pueden realizar afirmaciones acerca de sus clientes en relación con determinadas características conductuales y psicológicas. Nuevamente, estamos ante una conclusión importante para los estudios de evaluación multisitio: parece factible implementar procedimientos de recogida de datos antes del inicio del programa con el fin de estudiar las diferencias entre determinados tipos de participantes y abandonos, más adelante.

Marco

Aproximadamente dos tercios de los programas utilizan el trabajo en grupo (eventualmente mezclado con otros tipos de intervención). La atención individual también es frecuente, pero el trabajo en grupo es la forma predominante de intervención. Para los estudios de evaluación multisitio, ambos enfoques (trabajo en grupo, atención individual) parecen interesantes. Otros enfoques son menos frecuentes (p. ej., terapia conyugal, mediación) y dichos enfoques podrían incluirse en determinados diseños de evaluación, pero requerirían modelos específicos de investigación.

En cuanto a la “dosis” de tratamiento, existe mucha variación entre los programas, situándose la mayoría de ellos en un rango de 14-52 sesiones, y 14-52 semanas de duración. Esta variabilidad es importante para los estudios de evaluación en relación con un enfoque de tipo dosis-resultado, lo que significa que se podrían comparar las intervenciones más largas y más cortas.

Enfoque de trabajo

En cuanto al enfoque de trabajo de los programas para su trabajo concreto con agresores, la mayoría de programas se sitúan en el ámbito de la terapia cognitivo-conductual (TCC) y un enfoque psicoeducativo. Junto con combinaciones de TCC, entre dos tercios y tres cuartas partes de los programas tienen un enfoque psicoeducativo o un modelo tipo Duluth. Evidentemente, estos enfoques no son idénticos, pero tienen algunos elementos en común, especialmente a nivel teórico (teoría del aprendizaje). En el caso de EE. UU., Gondolf (2012) habla de dos grandes categorías de enfoques en el ámbito de trabajo con agresores: enfoque cognitivo-conductual y psicodinámico, cada uno de ellos muy diversificado en sí mismo. En el caso europeo, apreciamos una variedad de enfoques, en ocasiones combinando elementos de diferentes teorías.

Medición de los resultados

En cuanto a las mediciones de los resultados, un 81 % de programas afirman que miden los resultados de algún modo (pág. 33). Las fuentes más importantes de información son los propios hombres (un 94 % afirmó haber medido sus resultados así), los facilitadores (63 %) y las (ex) parejas (60 %), pero también se incluyeron otros servicios (54 %), informes policiales/judiciales (43 %) y servicios de atención a las víctimas (32 %). Dos tercios de los programas afirmaron realizar seguimientos.

Este resultado muestra que integrar otras fuentes de información además de la de los hombres que participan en los programas es una práctica frecuente y común en los proyectos, algo que los futuros estudios de evaluación pueden aprovechar. La *perspectiva de la pareja* es considerada como la principal fuente de información en los proyectos de *evaluación* y en la mayoría de países se pueden encontrar programas que incluyen información de las (ex) parejas. De acuerdo con la información de los programas en la encuesta existen algunos países, como Reino Unido, donde esto parece ser común, mientras que en otros no lo es (p. ej., Francia y Suiza). Asimismo, se debe tener en cuenta el vínculo con los *programas dentro del Sistema de Justicia Penal* en algunos países, a saber que “principalmente los programas 'por orden judicial' tienden a no usar la información de las parejas y servicios de atención a las víctimas como cabría esperar, porque muchos de estos programas indicaron que no se ponen en contacto con las parejas... y no miden los resultados con ellas...” (Geldschläger et al. 2014, pág. 46).

No obstante, hemos encontrado programas en 17 países, repartidos por todo el continente, que incluyen la información de las (ex) parejas en la medición de los resultados. Por consiguiente, concluimos que no debería existir ningún problema grave con un estudio multipaís que incluyera a las parejas a la hora de evaluar los resultados de los programas. Cuando esto no fuera posible, los programas respectivos desempeñarían un papel diferente en una futura estrategia de evaluación.

En cuanto a los *hombres que participan en los programas*, los resultados más frecuentes que se midieron fueron los siguientes: ausencia de violencia/disminución de la violencia, actitudes y creencias, habilidades de comunicación de los clientes, así como disminución del riesgo de violencia. También se midieron aspectos psicológicos, la paternidad y la calidad de vida en entre la mitad y dos tercios de los programas.

Las dimensiones más importantes que se preguntaron a las *(ex) parejas* fueron: violencia, sensación de seguridad, disminución del riesgo de violencia, seguridad de los niños y niñas,

calidad de vida, actitudes del hombre, habilidades de comunicación del hombre, y habilidades parentales del hombre.

En conjunto, existe una elevada variabilidad de métodos de medición de resultados de los programas para agresores. Por este motivo, una de las principales conclusiones de esta encuesta es que los programas europeos para agresores están lejos de alcanzar un consenso sobre los métodos estándares de medición de los resultados y que la necesidad de cierta armonización parece bastante evidente (Geldschläger et al., 2014, pág. 56). En la Línea de trabajo 4 del proyecto IMPACT damos un paso en esa dirección.

2.2.2. Conclusión

A partir de los resultados de la encuesta realizada en la Línea de trabajo 1 llegamos a la conclusión de que no existen problemas graves para medir los resultados en muchos países y para muchos programas de una forma amplia, incluyendo las perspectivas de los hombres que participan en los programas, de las parejas y los niños, los registros oficiales y los facilitadores. No obstante, se requieren recursos adicionales para que la mayoría de programas participen en un proyecto de *evaluación*.² Los profesionales parecen abiertos a la evaluación, y se están usando varios procedimientos de recogida de datos. Sin embargo, los estudios de evaluación van mucho más allá de la medición y monitorización de los resultados. En consecuencia, los estudios de evaluación multisitio deberían llevarse a cabo como proyectos independientes con financiación adicional. Basándonos en las conclusiones de la Línea de trabajo 1, no consideramos que sea factible realizar estudios de evaluación multisitio/multipaís que se basen en las prácticas actuales de los programas, en relación con la medición de los resultados o la recogida de información en la fase de acogida. Las prácticas existentes sencillamente son demasiado heterogéneas para su aplicación en un enfoque multisitio. Los procedimientos adicionales en los estudios de evaluación requerirán recursos añadidos para prácticamente todos los programas implicados.

2.3. Aplicación de los resultados de la Línea de trabajo 2

En la Línea de trabajo 2 el énfasis se trasladó de la práctica a la investigación. El objetivo era adquirir conocimientos acerca de los enfoques empleados en los estudios de investigación de evaluación sobre programas para agresores en Europa. El equipo de investigación identificó estudios publicados en revistas científicas, así como “literatura gris”, tanto en inglés como otros idiomas europeos. Se analizaron los métodos, los diseños, las mediciones de resultados y otros aspectos de los estudios, de tal modo que los resultados de la Línea de

² Esto también es aplicable a la medición de los resultados de programas para agresores: en lo que se refiere a los obstáculos para la medición de los resultados: los principales problemas para muchos programas son limitaciones de tiempo y recursos. Una pequeña parte de los programas (aprox. un tercio) aducen “falta de metodología” y el “contexto legal/institucional” como obstáculos para medir los resultados (en lo relativo a los obstáculos legales, véase más arriba: programas en el Sistema de Justicia Penal en algunos países). No obstante, prácticamente todos los programas de la encuesta están interesados en mejorar las mediciones de sus resultados, en la evaluación y en el intercambio internacional. Por lo tanto, muchos profesionales desean más recursos humanos, recursos económicos, métodos, recursos para el análisis de datos, así como guías/recomendaciones y formación/asesoramiento.

trabajo 2 fueron de gran relevancia para el objetivo de la Línea de trabajo 3, es decir, dar recomendaciones para futuros estudios de evaluación

2.3.1. Aspectos relevantes de la Línea de trabajo 2

Frecuencias por país

Hester et al. (2014) describen detalladamente su enfoque encaminado a la identificación de estudios existentes. Se identificaron cuatro metaanálisis y síntesis basadas en un total de quince estudios. Curiosamente, la mayoría de esos estudios provienen del Reino Unido (7) y España (5), y estos dos países también estaban entre los países con más programas identificados en la Línea de trabajo 1.

Junto con otras medidas para identificar estudios de evaluación tanto publicados como sin publicar (o publicados como “literatura gris”), se pudieron usar un total de 65 estudios para su ulterior análisis en la Línea de trabajo 2. Nuevamente, muchos de esos estudios provienen de España y del Reino Unido, así como de países de Europa central³. Podemos observar una superposición de frecuencias con respecto a los *programas* (Línea de trabajo 1) y los *estudios de evaluación* (Línea de trabajo 2), que tiene sentido: los países con más programas para agresores tienden a realizar y publicar más estudios de evaluación.

Problemas relacionados con los estudios

Los estudios de evaluación analizados abordaban varios problemas. A continuación se enumeran algunos de los más importantes:

- *Diseño*: en cuanto al diseño de la investigación, prácticamente no pudieron hallarse ensayos controlados aleatorios (ECA). De un total de n=58 estudios de las cinco regiones europeas que abordaban los *resultados* (se centraban únicamente en los resultados o en los resultados y el proceso), solo dos tenían un diseño ECA, y eran estudios realizados en una prisión y una clínica psiquiátrica.
- *Grupos de control*: muchos estudios carecen de un grupo de control, o el grupo de control es pequeño. Este hecho constituye una limitación importante de los estudios disponibles, puesto que los grupos de control (o elementos equivalentes de control que permitan una inferencia causal) son cruciales para la validez interna de cualquier estudio.
- *Terminología*: en los estudios se han podido observar varias definiciones de términos clave (p. ej., “inicio”, “finalización”). No existe una terminología coherente.
- *Tiempo*: se han constatado diferencias en los enfoques y las metodologías en cuanto a los momentos de recogida de los datos. Hester et al. (2014) han introducido un esquema para su análisis que se debería utilizar y fomentar para un entendimiento común respecto de los momentos de observación en los estudios de evaluación. Consta de cinco momentos en los que se debe recabar información durante una

³ Para su análisis, Hester et al. (2014) emplearon una clasificación de países en cinco regiones donde se identificaron estudios de evaluación: Europa Occidental (Reino Unido, Irlanda), del Sur (España, Portugal), del Norte (Suecia, Dinamarca, Finlandia, Islandia), Oriental (Croacia) y Central (Alemania, Austria, Suiza, Países Bajos).

evaluación (T0: preinicio; T1: inicio; T2: durante la intervención; T3: fin; T4: seguimiento). Este esquema se describe más adelante.

- *Abandono*: las tasas de abandono no estaban claras, o no estaba claro en qué fases se produjo el abandono. En la mayoría de casos faltaba información sobre el tamaño de las muestras y el abandono en *cada* fase del proceso de intervención. La información aportada varía mucho en relación con las muestras y el tamaño de las mismas, lo que se debe a una falta de estándares y terminología establecida para estudios de evaluación de programas para agresores.
- *Derivación*: es fundamental contar con información sobre las vías de derivación o de entrada en los programas (voluntaria o asistencia por orden judicial, así como otras vías de derivación). También se requieren los motivos de abandono, información sociodemográfica y de otra índole de todas las submuestras (usuarios derivados antes de que empiece la intervención; al principio de la intervención; abandonos; clientes que completan el programa). En muchos casos falta esta información.

Medición de los resultados e impacto

En relación con los resultados, los datos reportados por los agresores, los datos oficiales del sistema de justicia penal y los datos de las víctimas/supervivientes se emplean en todas las regiones, a excepción del Sur, donde los datos de las parejas no aparecen en los estudios disponibles. No obstante, existen programas en la región Sur que incluyen la perspectiva de las víctimas, tal como sabemos por el análisis realizado en la Línea de trabajo 1. Sin embargo, parece existir una diferencia en las prácticas y posibilidades legales de los programas dentro de las prisiones y programas comunitarios. Los programas dentro del sistema de justicia penal no integran las perspectivas de las parejas; no obstante, se evalúan más a menudo que en los programas comunitarios. Asimismo, utilizan muchos instrumentos psicométricos validados/estandarizados, centrándose en aspectos psicológicos y cambios internos de los agresores. Esto refleja un marcado enfoque psicológico en la evaluación de programas para agresores en el sistema de justicia penal en la región Sur, con la mayoría de estudios procedentes de España.

Las mediciones más amplias de los resultados se han hallado en la región Occidental, "... incluyendo el cambio en las actitudes respecto de las mujeres y la violencia, la reincidencia oficial y autorreportada, la revictimización, la calidad de vida, la seguridad y el bienestar de las mujeres/víctimas y sus hijos e hijas, los niveles de estrés parental y el consumo de alcohol y drogas por parte del agresor" (Hester et al., 2014, pág. 26). Se pueden encontrar instrumentos de evaluación del riesgo en estudios de las regiones Occidental y Sur.

En general, las dimensiones que se estudian y los instrumentos que se emplean varían mucho de una región a otra (y de un estudio a otro). Esta ausencia de una metodología coherente dificulta o incluso imposibilita comparar los estudios.

Por lo tanto, cuando se trata de dar una opinión sobre la efectividad de los programas para agresores sobre la base de los estudios de evaluación disponibles, solo se puede aportar una imagen muy general: existen algunas conclusiones relevantes desde un punto de vista estadístico que apuntan en la dirección de cambios positivos o existen conclusiones prometedoras que no son estadísticamente significativas. Este patrón se pudo constatar para una variedad de dimensiones (cambio conductual; cambio de actitud; cambio

psicológico; cambio en los factores de riesgo; motivación y adherencia). No obstante, existen limitaciones en las conclusiones en relación con el impacto, a saber, tamaños pequeños de muestra en muchos casos, elevados niveles de abandono (o falta de información sobre el abandono), así como una excesiva dependencia de los datos reportados por los propios hombres en la región Sur.

2.3.2. Conclusión

En relación con la Línea de trabajo 3, a pesar de todos los problemas hallados en los estudios, los resultados apuntan que es posible obtener y combinar varios datos en todas las regiones. Existen diferentes enfoques respecto de la evaluación, pero parece factible aplicar determinados instrumentos e integrar determinadas fuentes de información en todos los países (con algunas excepciones, p. ej., programas en las prisiones, que deben desempeñar un papel diferente en la futura estrategia de investigación de evaluación).

Hoy en día, el principal problema parece ser la metodología incoherente respecto de la evaluación de programas para agresores, lo que dificulta comparar los estudios en diferentes aspectos (diseños y grupos de control, terminología, instrumentos, fuentes de datos). Se requiere una armonización de la metodología de evaluación a fin de superar la fragmentación en la investigación de evaluación sobre programas para agresores. Esto parece posible, puesto que los autores del documento de trabajo sobre la Línea de trabajo 2 concluyen:

“Existen algunas diferencias y similitudes interesantes entre las regiones y dentro de las mismas, y en general sería posible tomar elementos de diferentes enfoques para empezar a desarrollar una robusta metodología de evaluación” (Hester et al., 2014, pág. 39).

2.4. Entrevistas con grupos focales de expertos y expertas

En lo que se refiere a los estudios de evaluación de programas para agresores, existe una gran variedad de partes implicadas, todas ellas con sus intereses y perspectivas, que pueden diferir entre sí pero también se pueden solapar:

- *Víctimas/Supervivientes de violencia de género*: las parejas y los hijos e hijas de hombres violentos querrán saber si un programa es eficaz para detener la violencia de su pareja/padre y aumentar su seguridad.
- *Profesionales de apoyo a las víctimas*: las y los profesionales de las organizaciones de apoyo a las víctimas quieren saber si los programas son eficaces y si son lo suficientemente efectivos para justificar el uso de recursos financieros adicionales que, de lo contrario, se destinarían a incrementar el apoyo a las víctimas.
- *Profesionales del sistema de intervención*: la policía, fiscales, los tribunales, agentes de libertad condicional y otras partes implicadas también tienen interés en ver si los DVPP tienen un efecto específico (aparte de las intervenciones existentes).
- *Profesionales en los DVPP*: la mayoría de profesionales querrán conocer los resultados de su trabajo y querrán saber específicamente si es o no efectivo, y si la

violencia disminuye y, en tal caso, cómo pueden hacerlo para obtener aún mejores resultados.

- *Dirección de programas DVPP*: tendrán interés en los efectos de los programas, así como en mejorar la calidad de los mismos, es decir, su interés es demostrar la eficacia de los programas y mejorar los resultados.
- *La población en general*: puede estar interesada en reducir la violencia de género en la sociedad e incrementar la seguridad. El debate público sobre este tema a menudo tiene lugar en los medios de comunicación, que tienen su propia lógica y sus propios intereses. Los medios de comunicación de calidad pueden plantear cuestiones similares a los demás grupos interesados, mientras otros segmentos de los medios de comunicación optan por contenidos impactantes, llamativos y extraordinarios. A menudo, esta parte de los medios de comunicación hace un llamamiento a aplicar penas más duras para los agresores y leyes más estrictas, sin embargo, existe cierto interés en los efectos de alternativas como los DVPP.
- *Financiadores, partidos políticos*: las partes interesadas pertenecientes a esta categoría pueden querer demostrar a la población sus actividades contra la violencia y, al mismo tiempo, asegurarse de que los fondos que aportan a estos programas tienen efectos y permiten el cumplimiento de los objetivos predefinidos. Simultáneamente, es posible que deseen optimizar el uso de los recursos, es decir, están interesados en contestar preguntas acerca de la rentabilidad.

En las *entrevistas con grupos focales* y en los *debates con personas expertas* (véase más adelante), la cuestión de los diferentes intereses así como el solapamiento de intereses entre los diferentes grupos de partes interesadas tuvo relevancia.

Las **entrevistas con grupos focales de personas expertas** (un método de investigación social que combina la entrevista con expertos y expertas, y la entrevista con grupos focales; véase Babbie, 2003; Silverman, 2013) se llevaron a cabo en octubre de 2013 durante el *Taller A* del proyecto, organizado en la Universidad de Bristol (Reino Unido). Se realizaron cuatro entrevistas de 45 minutos con grupos focales de 18 expertos y expertas (personal investigador y profesionales que trabajan con agresores) de 11 países europeos (Alemania, Austria, Dinamarca, España, Finlandia, Irlanda, Noruega, Polonia, Portugal, Reino Unido, República Checa). En cada grupo un miembro del equipo del proyecto formuló tres preguntas principales:

- ¿Qué condiciones ve que favorezcan y obstaculicen la viabilidad de un diseño multisitio en diferentes países, a escala nacional o local?
- ¿Qué condiciones ve que favorezcan y obstaculicen la viabilidad de un diseño multisitio en diferentes países, a escala internacional?
- ¿Qué posibilidades y obstáculos ve en relación con la siguiente idea: “Proporcionar una herramienta a los profesionales (que trabajan en programas para agresores) para recabar datos comparables en diferentes países, agrupar estos datos y analizarlos de forma centralizada”?

Tras las entrevistas, se prepararon informes internos, que sirvieron de base para la presentación durante el Taller B, así como para su ulterior análisis (véase más adelante).

Los **debates con personas expertas** tuvieron lugar durante el Taller B organizado en abril de 2014 en Copenhague (Dinamarca). 41 profesionales, personal investigador y responsables políticos de Alemania, Austria, Bulgaria, Croacia, Dinamarca, España, Estonia, Finlandia, Irlanda, Italia, Lituania, Países Bajos, Reino Unido, República Checa, Rusia, Suecia y Suiza debatieron cuestiones y preguntas planteadas en la presentación *Posibilidad de evaluación en un enfoque multisitio/multipaís*:

- ¿Qué posibilidades y obstáculos veo en mi organización?
- ¿Qué preguntas deberían abordarse en los estudios de evaluación?:
 - ¿Funciona?
 - ¿En qué medida / con qué grado de eficacia?
 - ¿Qué funciona?
 - Preguntas específicas (quién vuelve a agredir, quién necesita qué, etc.)

Los resultados de los debates llevados a cabo en cuatro grupos se recopilaron y sirvieron como base para su ulterior análisis (véase más adelante). Se recabaron más opiniones acerca del concepto de estudios de evaluación multisitio/multipaís sobre DVPP en la conferencia final del proyecto celebrada en Barcelona, España (octubre de 2014).

3. Resultados: perspectiva de los expertos y expertas

Los resultados de las entrevistas con grupos focales de expertos así como los debates se resumieron en una recopilación de temas, que han sido una fuente de información sobre cómo avanzar en el desarrollo de una metodología para evaluar los programas para agresores. Estos temas se han clasificado en cinco grandes grupos: preguntas, posibilidades, ventajas, retos y obstáculos.

3.1. Preguntas

Una de las preguntas más comunes e importantes que surgió durante las entrevistas y los debates hacía referencia a la comprensión fundamental de los “programas para agresores”. Según las personas expertas de varios países, la definición de tales programas puede variar. Además de los típicos programas (a saber, las reuniones regulares de grupo, dirigidas por profesionales, estructuradas, coordinadas, evaluadas, etc.) las siguientes intervenciones fueron mencionadas como elementos en el trabajo con agresores de violencia de género en relaciones íntimas: líneas de ayuda telefónica; intervenciones individuales; reuniones informales y voluntarias en la comunidad y/o la iglesia; reuniones de Alcohólicos Anónimos, etc. Por lo tanto, uno de los primeros pasos al pensar en investigación de evaluación europea multipaís sobre el trabajo con agresores sería aportar una **definición de qué es un programa para agresores** que fuera común pero también relativamente amplia.

Otra cuestión importante abordó el tema de la **similitud de los programas**. La mayoría de expertos compartían la opinión de que solo deberían compararse los resultados de programas similares (mismo enfoque; métodos de trabajo similares; carácter, tiempo,

alcance similar, etc.). No obstante, esta suposición despertó dudas acerca del número de programas europeos lo suficientemente similares como para ser comparados.

Esta cuestión de la similitud estuvo seguida por el tema del **tipo de programa** a estudiar. Algunos expertos y expertas subrayaron que el estudio futuro debería centrarse en un único tipo: bien los programas *obligatorios* o bien los *voluntarios*, mientras que otra parte propuso estudiar ambos enfoques o concentrarse en programas “mixtos” (con ambos tipos de vías de derivación en el mismo programa).

Finalmente, las personas expertas abordaron la cuestión del **procedimiento de investigación** de un futuro estudio. Prevalcieron las cuestiones concretas, tales como quiénes formarían parte del personal investigador, quién debería recoger los datos, etc. Se debatieron los pros y los contras de diferentes posibilidades, por ejemplo, contar con personal investigador externo independiente o contar con profesionales y personas que trabajen directamente con agresores para recoger datos y, quizás, realizar los pasos analíticos. La motivación y el interés genuino en la evaluación del trabajo con agresores pueden ser una ventaja en el último caso, pero los conflictos de intereses pueden constituir una desventaja. Según los expertos y expertas, ambas aproximaciones tienen sus pros y sus contras (diferente acceso a los datos y muestras; costes; distinto nivel de habilidades analíticas; la cuestión de la “objetividad”; el tiempo, etc.), y también debería considerarse la decisión sobre qué procedimiento elegir.

3.2. Posibilidades

Muchos expertos y expertas expresaron su interés en la realización y el apoyo a algún tipo de estudio de resultados multipaís a escala europea. En consecuencia, la cuestión sobre las posibilidades de tales estudios fue la más debatida. En los debates relacionados, los **estándares para programas** así como las **mediciones de los resultados** a emplear en los estudios de evaluación fueron de particular interés para muchas de las personas expertas. Un punto de partida para la definición de estándares mínimos para programas para agresores, así como para las mediciones de sus resultados, puede encontrarse en el *Convenio sobre prevención y lucha contra la violencia contra la mujer y la violencia doméstica* (“Convenio de Estambul”, Consejo de Europa, 2011), en la nota informativa sobre criterios de resultados elaborada por Westmarland, Kelly y Chalder-Mills (2010), o en las *Directrices para el desarrollo de estándares para los programas que trabajan con agresores de violencia de género* (Trabajo con agresores [WWP] – Red Europea, 2008). También hay que destacar que, en opinión del grupo de expertos y expertas, los estándares deben basarse en políticas de igualdad de género.

Asimismo, se mencionó el término “mínimo” en relación con los estudios de investigación y evaluación, en el sentido de que se requiere un **mínimo de estándares en cada estudio específico de país** o en cada parte del estudio, a fin de poder “combinar” las diferentes partes. Se mencionaron varios elementos que ya se han empleado en varias ocasiones, por ejemplo, el modelo ecológico; el modelo de observación por etapas (pre, post, seguimiento); determinados instrumentos (por ejemplo, inventario de violencia y malos tratos). El personal investigador ya ha empezado a describir y analizar varias cuestiones relevantes, por ejemplo,

el contexto europeo y nacional (legislación, vía de derivación, etc.), tipos de agresores, frecuencias y tipos de violencia, áreas deseables de cambio (conducta violenta de los agresores; comunicación; opiniones e identidades; relaciones familiares, etc.). El debate llevó a algunos expertos y expertas a plantear las primeras preguntas relativas a la investigación (sin aportar respuestas en ese punto del debate) sobre cómo abordar estas cuestiones:

- ¿Qué contribuye en mayor medida a poner fin a la violencia y aumentar la seguridad de las víctimas?
- ¿Los hombres cambian? ¿Por qué? ¿Cómo?
- ¿Qué funciona para quién?
- ¿Qué (otros) factores influyen en la vida de los agresores (clase, religión, origen inmigrante/no inmigrante, etc.)?

Finalmente, las personas expertas abordaron las condiciones que pueden ayudar a incrementar la **calidad de futuros estudios de evaluación multisitio/multipaís**, por ejemplo: triangulación de métodos de investigación (cualitativos, cuantitativos, observaciones); cruce de datos y teorías; tener en cuenta las perspectivas de agresores, víctimas y profesionales. Asimismo, en opinión del grupo de expertas y expertos, la implicación de organizaciones paraguas y ONG locales/especializadas (grupos de inmigrantes, grupos de LGBTQ, grupos religiosos) en el proceso de investigación también puede resultar extremadamente útil.

3.3. Ventajas

En todas las entrevistas y debates, los expertos y expertas subrayaron las ventajas de estudios de evaluación multisitio. Una manera **correcta y útil** de medir la violencia y otras **mediciones de resultados** deberían usarse o desarrollarse en tales estudios, lo que no solo sería beneficioso para profesionales de todos los países europeos, sino que también contribuiría a llenar las lagunas en el conocimiento y la investigación sobre la violencia en general. Según algunos expertos y expertas, la metodología existente es insuficiente. En particular, los y las profesionales expresaron la esperanza de que la evaluación pudiese apoyar la **continuidad de los programas existentes**, puesto que algunos órganos de decisión vinculan la aportación de más financiación a la obtención de resultados positivos en la evaluación de los programas.

Asimismo, tal como señalaron las personas expertas, la idea de integrar la perspectiva de las víctimas y las organizaciones de víctimas permitiría **estudios de evaluación más exhaustivos**, incluido el **sistema más amplio** implicado en la intervención. Los estudios europeos de evaluación multipaís también podrían contribuir a la **promoción** de programas nacionales para agresores que aún no son bien conocidos entre las diferentes partes interesadas, y los resultados relativos a los programas servirían de **base de conocimiento** para profesionales que desean mejorar la calidad de su trabajo, evitar errores comunes, etc. Además, como efecto adicional, la atención proporcionada a un estudio internacional sobre el trabajo con agresores aumentará la **concienciación social** sobre la existencia y el papel de los programas para agresores en determinados países y, según algunos expertos y expertas, esto ayudará a realizar una promoción más efectiva tanto a escala nacional como internacional. Si un

estudio europeo de evaluación multipaís se realizara como proyecto europeo, esta podría ser la única oportunidad para evaluar programas en aquellos países donde los programas aún no se han evaluado por falta de recursos y metodología.

3.4. Retos

Los expertos y expertas coincidieron en que el **proceso de reclutamiento** es uno de los mayores retos en relación con los estudios de evaluación multipaís, especialmente en el caso de que no exista financiación adicional. Subrayaron que un estudio de este tipo solo puede tener éxito si el personal investigador puede convencer a profesionales, personal de los programas y a otros actores clave para que sacrifiquen **tiempo y recursos** a fin de participar en la investigación. En su opinión, el argumento de llenar la laguna de conocimiento en el área de los estudios de evaluación de programas para agresores puede no ser suficiente para motivar a los actores clave a participar. Por este motivo, también se debería poner énfasis en otros **beneficios** derivados de la participación en estudios de investigación de evaluación, tales como la formación externa para el personal de los programas (no solo sobre métodos de evaluación), visitas de estudios, certificación (por ejemplo, por parte de WWP - Red Europea, o servicios nacionales como Respect, Reino Unido, o BAG Täterarbeit HG, Alemania). Según el personal investigador entre el grupo de expertos y expertas, uno de los beneficios más importantes de un estudio de evaluación será el feedback obtenido de los usuarios, lo que puede suponer un incremento de la calidad y la eficacia de los programas. Además, los programas que participen en el estudio podrían tener mejores oportunidades de recibir financiación en el futuro si se ha llevado a cabo un proceso de evaluación y los resultados demuestran que el programa tiene cierto impacto. La dimensión europea de un estudio de evaluación realizado externamente aumentará la credibilidad de los resultados a escala nacional.

Otro reto consiste en llegar a un **acuerdo** acerca del uso de los mismos **modelos de evaluación** (criterios, diseños, métodos, instrumentos). Puesto que los programas y métodos prácticos son diferentes en muchos países, en este momento se emplean distintos criterios e instrumentos para medir los resultados. Por lo tanto, el reto no consiste solo en seleccionar y emplear las mismas herramientas, sino en seleccionar las más universales y transferibles.

3.5. Obstáculos

Según los expertos y expertas, llevar a cabo una investigación europea multipaís de resultados puede encontrarse con una serie de obstáculos. Uno de los problemas más graves es la **financiación**, puesto que la realización de un estudio tan complejo generará una gran cantidad de costes. Por consiguiente, encontrar la ayuda financiera para un proyecto de este tipo es fundamental, pero también constituye un gran reto. La financiación no solo es necesaria para cubrir el trabajo de un equipo de investigación externo y otros gastos del proyecto (gastos de viaje, equipamiento, etc.), sino también para cubrir el trabajo adicional del personal de los programas, especialmente en aquellas situaciones donde los y las profesionales estén mal pagados o incluso trabajen de forma voluntaria (lo que suele ser el

caso en los programas ubicados en países postsocialistas), y/o donde la evaluación de los programas no esté incluida en su financiación regional o nacional.

La **diversidad de contextos** en diferentes países europeos constituye otro obstáculo que influye tanto en el carácter de los programas como en el futuro proceso de investigación. Los distintos contextos jurídicos pueden influir en la elección de métodos específicos y presuposiciones básicas de un proyecto de investigación. Las regulaciones legales en relación con el sistema de intervención, la protección de los datos personales, etc. pueden impedir que el personal investigador pueda **recuperar determinados datos** en determinados contextos. Uno de los ejemplos comunes de esta situación fue dado por un experto de España, donde el personal de los programas para agresores en el sistema de justicia no tiene acceso a las víctimas y, por consiguiente, la inclusión de la perspectiva de la pareja puede resultar difícil o incluso imposible de obtener.

Otra cuestión es el **contexto cultural** de diferentes países europeos, lo que también puede influir en el proceso de investigación. En algunos países, la **interpretación de términos fundamentales** tales como violencia de género, derechos de las víctimas, comportamiento inadecuado, roles de género, igualdad de género, etc. puede diferir y, por lo tanto, diseñar herramientas de investigación multipaís puede constituir un gran reto. Asimismo, los países varían en cuanto a su situación **económica**, el papel de la(s) **religión(es)** (el grado en el que la política se ve influida por ideas religiosas, países más o menos laicos), la variedad de trasfondos culturales en relación con la **inmigración**, etc., pero también en cuanto al grado de **implementación de políticas de género** y la perspectiva de género, el grado de implementación y desarrollo de **programas para agresores**, así como el nivel de **aceptación de la importancia** de este tipo de sistema de intervención.

En cuanto a los programas, los expertos aprecian diferencias importantes entre países en términos de **teorías de cambio**, **criterios de éxito**, **tipos de hombres** que pueden participar en programas para agresores, y diferentes enfoques sobre el **sexo de los usuarios** en los grupos: existen programas que trabajan únicamente con hombres, otros presentan un enfoque mixto, con diferentes grupos/vías para agresores hombres y mujeres o sin diferenciación alguna. Para resumir, en opinión de los expertos y expertas, el obstáculo (metodológico) más serio para la investigación europea de evaluación multipaís es el hecho de que los **programas para agresores**, por diversos motivos, sencillamente son muy **diferentes**.

Finalmente, algunos expertos y expertas dudaron de que sea posible llevar a cabo un estudio europeo complejo puesto que aún no existen suficientes programas para agresores (en algunos países, como Hungría, los programas ni tan siquiera existen; Kutrovátz, 2013), lo que podría suponer un grave problema de muestreo.

3.6. Conclusiones de los grupos focales y el debate con expertos y expertas

Las contribuciones del grupo de expertos y expertas a los debates y sus respuestas a diferentes preguntas se resumieron en las siguientes recomendaciones:

- En primer lugar, se debe aportar una **definición** de “programas para agresores” que se tenga en cuenta en los estudios de evaluación multipaís. Los programas estándares (cooperación con los servicios de atención a las víctimas y otros servicios; basados en TCC o un enfoque psicoeducativo, con un enfoque de género) parecen los más adecuados en este punto, también por motivos metodológicos: necesitamos programas similares en diferentes contextos, y no diferentes programas en diferentes contextos. Los programas también deben ser capaces de recoger los datos adecuados de los resultados, lo que depende de la cooperación, es decir, un enfoque coordinado. Asimismo, el estudio debe centrarse en los programas que sean **compatibles** con los **estándares** y las **políticas** existentes (por ejemplo, el Convenio de Estambul).
- **Proyecto de investigación:** trabajar juntos en un nuevo proyecto financiado fue considerado como la manera más factible de realizar un proyecto de evaluación multisitio/multipaís. Se cuestionó –por lo menos en este punto– la idea inicial de utilizar un conjunto de herramientas de evaluación mediante la recogida y el análisis centralizado de los datos. No obstante, el conjunto de herramientas debería implementarse ampliamente, la investigación debería basarse en ese conjunto de herramientas en la medida de lo posible y, en un fase posterior, quizás sea posible volver a la idea inicial. Los expertos hicieron hincapié en la diferencia del conjunto de herramientas (como producto resultante de la Línea de trabajo 4, para medir los resultados) y de los instrumentos como parte de una metodología de evaluación. Incluir proyectos de evaluación en el trabajo diario parece difícil.
- Los estudios futuros deben diseñarse de una **manera más bien amplia**, incluyendo el **sistema** implicado en la intervención. El impacto social de los programas para agresores puede ampliar el diseño básico. La diferencia de **contextos** en diferentes países debe tenerse en cuenta desde el principio, tanto en lo relativo a las regulaciones legales, como aspectos del sistema de intervención y cuestiones culturales.
- Los criterios de evaluación, diseños, métodos, instrumentos, etc. deben hacer referencia a las **buenas prácticas** existentes **en la evaluación** y tener en cuenta los puntos de vista del grupo de expertos y expertas.
- Las múltiples **preguntas** que deben responder los estudios de evaluación se deben **sistematizar**, preferiblemente asignándolas a determinados grupos y **partes implicadas**.
- Los modelos de evaluación deben mostrar un alto grado de **transferibilidad**, a fin de poder incluir múltiples proyectos en varios países a largo plazo.
- En relación con el tipo de programa: las **vías de derivación** (autoderivación o derivación por otros servicios, mixto) deben utilizarse como una variable, no como un criterio de selección.
- En cualquier caso, en el proceso de reclutamiento debe ponerse énfasis en los **beneficios** derivados de la participación en estudios de evaluación.

4. Cómo avanzar en la evaluación de DVPP

4.1. Punto de partida: Dónde estamos actualmente

Como se ha puesto de manifiesto en las fuentes de estudio disponibles sobre evaluación de DVPP, así como en las perspectivas de los expertos y expertas, la idea inicial de la Línea de trabajo 3 debe modificarse. No ha resultado ser un plan factible diseñar y realizar grandes estudios de evaluación que respondan a las diversas cuestiones planteadas por las diferentes partes interesadas de una manera amplia, por lo menos no en este momento. Por el contrario, hemos desarrollado un concepto de esfuerzos coordinados de evaluación, que deben ser vistos como una serie de proyectos de evaluación más pequeños con una metodología similar y la posibilidad de coordinación de estos proyectos.

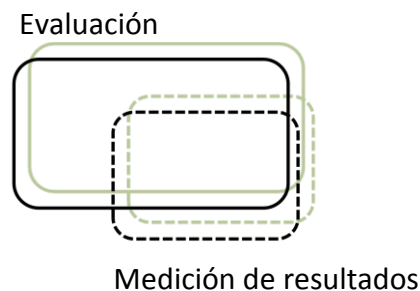
A modo de resumen, los requisitos previos para proyectos de evaluación multipaís/multisitio son los siguientes:

La **principal motivación** de un proyecto de esta índole es hacer uso de la elevada variabilidad de condiciones nacionales y “sistemas” (prácticas de organismos, atención a las víctimas, DVPP, etc.) en toda Europa. Esa variabilidad puede ser un recurso para los estudios de evaluación, para el estudio de DVPP en diferentes contextos. Además, como muchos programas aún no se han evaluado, los estudios de evaluación multipaís/multisitio pueden brindar una oportunidad para ampliar la base empírica en lo relativo a DVPP. Como se ha indicado anteriormente, existe una gran diversidad de trabajo con agresores en los distintos países, pero también se pueden hallar programas similares (por ejemplo, DVPP basados en TCC o modelos tipo Duluth), y las respuestas coordinadas por la comunidad (CCR) son frecuentes. Existen procedimientos para medir los resultados en varios países, pero la variabilidad metodológica es elevada, lo que dificulta la comparación de los estudios existentes. Hay diferencias y problemas en relación con la terminología y los conceptos (por ejemplo, “finalización”, “éxito”, etc.), así como los diseños (por ejemplo, respecto de los momentos de medición y la fuente de información); los grupos de control resultan difíciles de establecer y existen preocupaciones éticas en relación con los grupos de control; el abandono es un gran problema en la práctica, tanto respecto de los hombres que participan en los programas como de las víctimas/supervivientes. No obstante, Hester et al. (2014) concluyen en su recopilación de estudios que “... en general sería posible tomar elementos de diferentes enfoques para empezar a desarrollar una robusta metodología de evaluación” (Hester et al., 2014, pág. 39).

Al grupo de expertos y expertas les pareció que vale la pena apostar por la idea de un enfoque multipaís, y los diferentes grupos implicados tienen interés en obtener respuestas a la cuestión de la eficacia de los programas DVPP. Tal como han argumentado las diferentes partes interesadas, debería producirse un énfasis en programas que sean compatibles con los estándares y las políticas existentes (por ejemplo, el Convenio de Estambul). No obstante, se hizo evidente que los estudios de evaluación que son capaces de responder a las preguntas relevantes requieren muchos recursos y esfuerzo, y la dirección de los programas y profesionales se han mostrado escépticos acerca de la integración de ese tipo

de trabajo en sus rutinas diarias. En otras palabras, no parece factible ir hacia estudios de evaluación internacionales multipaís/multisitio sin recursos adicionales. En la práctica, se puede convencer a las partes interesadas para que cambien los instrumentos que están empleando para evaluar sus resultados a instrumentos alternativos, pero persisten una serie de problemas respecto de la evaluación del trabajo con agresores, y dichos problemas no se pueden solucionar dentro de las prácticas diarias de los programas. Es importante aceptar que los estudios de evaluación sólidos son más amplios en su concepto y que necesitan más recursos que la medición de los resultados. El trabajo necesario y los esfuerzos invertidos en proyectos de evaluación van más allá de las rutinas diarias dentro de la medición de los resultados, aunque la medición de los resultados puede ser un buen punto de partida y constituye la base para que los DVPP estén “preparados para la investigación” (véase Gondolf, 2012; véase la Figura 2). En consecuencia, los y las profesionales hicieron hincapié en la necesidad de financiación adicional si debían participar en estudios de evaluación.

Figura 2. Aspectos solapados y específicos de (a) estudios de evaluación y (b) prácticas para medición de resultados.



En conjunto, podemos concluir que existen unos cimientos para proyectos de evaluación multipaís, pero definitivamente se requiere una armonización de la metodología. Los proyectos de evaluación necesitan a los DVPP y los servicios de apoyo a las víctimas. Por consiguiente, para mejorar las posibilidades de éxito, los futuros estudios de evaluación se deberían planificar como proyectos separados con suficiente financiación para cubrir los costes adicionales de las organizaciones implicadas.

4.2. Hacia dónde debemos ir

Lo ideal sería que, dentro de unos años, se haya superado la situación fragmentada en la evaluación de DVPP mediante una armonización de la metodología de evaluación, de tal modo que los resultados de diversos estudios se puedan comparar y se complementen entre sí. La idea es contar con una institución de coordinación que promueva, apoye y monitorice los futuros proyectos de evaluación de DVPP en términos de diseños, métodos, instrumentos, así como supervisar qué cuestiones específicas deben abordarse y contestarse. No será un *programa de investigación de evaluación de DVPP* completo, puesto que no es realista que una serie de estudios durante varios años se puedan anticipar y planificar de manera integral, principalmente debido a restricciones en la financiación. No obstante, es muy probable que existan varios estudios de evaluación, algunos a escala nacional, otros a escala internacional (por ejemplo, proyectos de la UE) y dichos estudios

deben ser comparables, y debe ser posible vincularlos como si de un “rompecabezas” se tratase.

Nos proponemos utilizar la *Red europea para el trabajo con agresores* (WWP-EN) como institución de coordinación, con el soporte de la *Red WAVE*, por varios motivos: WWP-EN ha liderado la vinculación de DVPP en la UE y más allá de sus fronteras. Como organización paraguas de DVPP, junto con los servicios de apoyo a las víctimas que cooperan, ya se han dado los primeros pasos hacia unos estándares para el trabajo concreto con agresores, y es posible que se adopten en el futuro. La evaluación de este trabajo es tan solo otro paso lógico, como lo es ocuparse de los **estándares** sobre cómo realizar este trabajo de **evaluación** (véase más adelante).

Esto se corresponde con lo que Gondolf ha denominado “evaluaciones conducidas por profesionales”, esto es, proyectos de evaluación que se han iniciado o codiseñado o han recibido la influencia del ámbito práctico. En muchos casos, como Gondolf ha señalado, la cooperación entre personal investigador y profesionales ha demostrado ser difícil, donde cada parte solo comprende parcialmente a la otra, donde hay margen para la mejora.

4.3. Estándares para proyectos de evaluación

Como hemos aprendido en los la Línea de trabajo 1 y 2, así como en la bibliografía específica, los estándares metodológicos y una terminología común parecen fundamentales para un desarrollo significativo de los proyectos de evaluación en el futuro. Evaluar el trabajo con los agresores en general y, en particular, los DVPP es una tarea muy específica, y los procedimientos de otras áreas (por ejemplo, la evaluación de la psicoterapia) no se pueden aplicar con un mero “copiar y pegar”. Algunos de los aspectos específicos de la evaluación de DVPP son:

- *Conceptos y terminología:* hasta la fecha se han producido amplios debates sobre hacia dónde deben dirigirse los DVPP, con consecuencias para los términos y los conceptos como “éxito”. Hay varios grupos de partes interesadas que han expresado sus posiciones e intereses en relación con los resultados de los DVPP. Dichas posiciones se deben tener en cuenta al planificar estudios de evaluación.
- *Fuente de información:* la información facilitada por los hombres es considerada como insuficiente por los expertos y expertas en la materia (véase, por ejemplo, Gondolf, 2012). La información facilitada por la pareja, la policía o los registros judiciales se requiere para complementar la información necesaria. La información facilitada por la pareja es considerada como la más fiable y significativa y, por lo tanto, debería constituir un indicador central. No obstante, a su vez, esto plantea cuestiones éticas que se deben tomar en consideración (véase Downes, Kelly & Westmarland, 2014).
- *Tiempos de observación:* a fin de hacer frente a los niveles relativamente elevados de abandono (tanto en cuanto a los agresores como a las víctimas/supervivientes), por un lado, y evaluar la sostenibilidad de los cambios logrados por los programas, por el otro, un simple diseño pre/post no es suficiente para evaluar los DVPP. Hester et al. (2014) han propuesto un esquema de momentos de observación a emplear en este

contexto. Han definido cuatro momentos: T0, cuando un hombre es derivado al programa; T1, cuando se inicia el programa; T2: durante el programa; T3: al final del programa; T4: seguimiento, por ejemplo, 6 meses tras finalizar el programa. Este esquema, junto con los requisitos relativos a la información que se debe recoger en cada momento (a quién hay que preguntar, qué instrumentos se deben usar), puede constituir un estándar central para la evaluación de los DVPP. Junto con una descripción suficiente sobre quiénes son los hombres en el momento T0 (por ejemplo, atendiendo a características sociodemográficas o psicológicas) y cuál ha sido la vía de derivación al programa, el problema del abandono se puede estudiar mucho mejor que en la actualidad.

4.4. Promoción de estándares

Una vez desarrollados esos estándares metodológicos, se deben promover y usar en estudios de evaluación. Dentro de la Línea de trabajo 3 se debatieron varias maneras de promover tales estándares. Se han identificado las siguientes posibilidades:

- Un grupo de trabajo dentro de WWP-EN, responsable de cuestiones de evaluación, podría desempeñar un papel clave en la armonización de la metodología de evaluación, recogiendo las mejores prácticas en la investigación de evaluación de DVPP y “encauzando” series de estudios de evaluación. Los futuros proyectos de evaluación necesitarán cooperar con los programas sobre el terreno, tal como se ha realizado en el pasado. Por lo tanto, es probable que los proyectos de evaluación se basen en organizaciones miembros de WWP-EN, que tendrán sensibilidad en cuanto a los estándares de evaluación de WWP-EN.
- Del mismo modo, las redes internacionales de organizaciones de apoyo a las víctimas, especialmente la *Red WAVE*, desempeñarán un papel clave. La evaluación de los DVPP requiere la perspectiva de las víctimas, por lo tanto, tales proyectos estarán en contacto y cooperarán no solo con organizaciones miembros de WWP-EN, sino también con organizaciones miembros de WAVE. Puesto que WAVE y WWP-EN cooperan, estas dos organizaciones paraguas europeas pueden tener un impacto significativo en la armonización de la metodología de evaluación a través de sus organizaciones miembros. Asimismo, existe la posibilidad de iniciar estudios de evaluación (nacionales e internacionales) dentro de WWP-EN. Especialmente en relación con proyectos multipaís de la UE, WWP-EN es la organización europea ideal para coordinar propuestas en el ámbito de la investigación de evaluación.
- Promoción de los estándares metodológicos a través de un proyecto de evaluación financiado por la UE como siguiente hito. La metodología concreta se puede desarrollar dentro de este proyecto futuro sobre la base del estudio de viabilidad de la Línea de trabajo 3, con un conjunto de programas aparentemente adecuados en varios países, a identificar a partir de los datos de la Línea de trabajo 1 y WWP-EN. Una vez que un proyecto se haya propuesto realizar un primer paso multipaís en la evaluación de DVPP, pueden seguirle otros proyectos basándose en las experiencias del proyecto piloto.
- Promoción de las prácticas similares a la investigación mediante el conjunto de herramientas de la Línea de trabajo 4. Aunque el conjunto de herramientas trata la

medición de los resultados en la práctica diaria de los programas WWP, es un paso hacia la armonización en la recogida de datos, en los que puedan basarse los estudios futuros en varios países. Si más programas que hoy en día utilizan el conjunto de herramientas, integraran rutinas similares a la investigación en su trabajo diario prepararían el terreno para un proyecto de evaluación. La medición de los resultados hará que los equipos de los programas estén “preparados para la investigación”.

- Los programas de la UE que armonizan la investigación y la práctica en toda la UE también constituyen oportunidades para armonizar los métodos en la investigación de evaluación. Existen programas de la UE que se esfuerzan por desarrollar un Espacio Europeo de Investigación, por ejemplo, COST. En el plano de la práctica, las redes WWP-EN y WAVE desempeñarían un papel de liderazgo en la promoción de estándares metodológicos en la evaluación, ayudando a elegir programas y alentando a sus miembros a participar en estudios de evaluación multipaís.

4.5. Un enfoque global de la evaluación de DVPP y el papel de los estudios multipaís

Para empezar con la metáfora del “rompecabezas” mencionada más arriba, diferentes tipos de estudios de evaluación pueden responder a diferentes preguntas, lo importante es asegurarse de que esos estudios se puedan comparar y vincular entre sí. Esto se puede garantizar mediante el uso de la misma metodología (o, como mínimo, un conjunto de procedimientos e instrumentos comparables). Nos proponemos analizar los diferentes tipos de estudios de evaluación desde la perspectiva de la categorización descrita a continuación. (Se proporcionará más información detallada sobre este y otros aspectos en artículos separados que se escribirán y publicarán como parte del proyecto IMPACT.)

4.5.1. Estudios básicos de contexto único

Se trata de estudios que evalúan el impacto de un DVPP como parte de una CCR. Tales estudios se llevan a cabo sobre el terreno, no en condiciones de laboratorio, y por consiguiente, es más probable que apliquen diseños cuasi experimentales, porque la asignación aleatoria de unidades de investigación (por ejemplo, hombres) a condiciones de tratamiento a menudo no es posible por motivos metodológicos y éticos. Los abandonos han sido utilizados como grupos de control, u otros grupos de hombres que no han completado las intervenciones (por ejemplo, hombres excluidos de los programas o no admitidos en los programas) se comparan con aquellos que han completado los programas.

En este punto nos gustaría proponer tres elementos como parte de una futura metodología de evaluación, los cuales deberán ser perfeccionados, por ejemplo, por el *grupo de trabajo de evaluación de WWP-EN*: (a) observaciones / tiempo; (b) indicadores de éxito; (c) diseño

(a) Observaciones / Tiempo

Hester et al. (2014) han proporcionado un esquema en su análisis de los estudios de evaluación europeos para describir “la forma y el tamaño de la muestra en las diferentes fases del proceso de evaluación” (pág. 17). Dicho esquema se debería utilizar en los estudios de evaluación sobre DVPP para definir diferentes momentos de observación. El esquema

está compuesto por cinco momentos, junto con propuestas de fuentes de información en los respectivos momentos:

- T0: “Preinicio”. Es importante saber qué tipos de hombres se derivan a los programas (datos sociodemográficos, características psicológicas, abuso de sustancias, vía de derivación, etc.), así como si entran o no en los programas. La comparación de las muestras en T0 y T1, junto con los motivos de abandono, pueden proporcionar información útil para el posterior desarrollo de intervenciones para aquellos hombres que no entran en el programa. Gran parte de la información se puede obtener de las víctimas/supervivientes, parte de los hombres y parte de los profesionales. (Sobre los aspectos éticos de la evaluación de programas por las víctimas/supervivientes, véase Downes, Kelly & Westmarland, 2014.) En relación con las víctimas/supervivientes, los indicadores de intervenciones exitosas (véase más adelante) se deben recopilar en T0. Dichos indicadores constituirán las mediciones centrales de los resultados en T3 y T4.
- T1: “Inicio”. La comparación de las muestras en T0 y T1 aportará información acerca de la primera ola de abandonos, mientras que la comparación de T1 y T2, así como en T3, mostrará qué hombres permanecen en el programa y cuáles lo abandonan. Los cambios en la situación vital y la información sobre las variables de contexto serán vitales de T0 a T3, a fin de interpretar los diferentes abandonos. En cuanto a los **hombres**, el énfasis en este momento se situará en variables que deben ser cambiadas por el programa (mediciones de resultados para el tratamiento), observadas durante T1 (por ejemplo, conducta, actitudes, motivación, factores de riesgo, etc.). Una vez más, se puede preguntar a las **víctimas/supervivientes** acerca de los indicadores para intervenciones exitosas.
- T2: “Durante”. Se recomiendan medidas repetidas durante la intervención. El desarrollo de los indicadores a lo largo del tiempo puede aportar información importante, tanto con respecto a los hombres como a las víctimas/supervivientes. Asimismo, se pueden abordar las preguntas relacionadas con el proceso y se pueden vincular con los resultados.
- T3: “Fin”. La comparación entre el grupo de tratamiento y el grupo de control, durante T1 (pretratamiento) y T3 (post-tratamiento), constituye el núcleo de los estudios para evaluar el impacto de las intervenciones. De nuevo, el análisis debe incluir como mínimo dos perspectivas: la perspectiva de las víctimas/supervivientes, en relación con los indicadores de éxito, y la perspectiva de los hombres en los programas, así como el grupo de control, en relación con las mediciones de los resultados para el propio tratamiento. La información aportada por los y las profesionales y, adicionalmente, los registros oficiales que se pueden recopilar durante T3.
- T4: “Seguimiento”. La estabilidad de los cambios se evalúa mediante un seguimiento, idealmente seis meses después de que los hombres finalicen el programa. De nuevo, se debe preguntar a las víctimas/supervivientes y los hombres que han participado en programas, mientras que los datos de los y las profesionales y los registros oficiales pueden actuar como fuentes adicionales de información.

Se puede realizar un análisis de abandono para cada momento, tanto para los hombres como para las víctimas/supervivientes, siempre que se haya recabado la información relevante (por ejemplo, variables sociodemográficas, riesgos, necesidades, etc.) durante T0.

(b) Diseños

Los ensayos controlados aleatorios (ECA) son considerados como el estándar de oro en la investigación clínica y psicológica. Los ECA se basan en la asignación aleatoria de unidades de investigación (por ejemplo, hombres) a condiciones de tratamiento (por ejemplo, intervención vs. no intervención). Existen múltiples problemas asociados a este procedimiento (por ejemplo, problemas éticos), y allí donde se han realizado ECA, se enfrentan a críticas específicas (véase Gondolf, 2002, 2012). Incluso en el sistema de justicia penal, a menudo resulta difícil realizar ECA (véase por ejemplo Dobash et al., 2000), aunque probablemente las mejores condiciones para realizar ECA se den en las prisiones (véase más adelante).

Figura 3. Diseño de un ensayo controlado aleatorio (ECA)

Asignación		T1 (pre)	(T2, durante TM)	T3 (post)	(T4, seguimiento)
aleatoria	Grupo de tratamiento	0	X 0	0	0
de individuos					
a grupos	Grupo de control	0	0	0	0

Como consecuencia, en varios estudios sobre abandonos, los que no asisten o los hombres que han sido excluidos de los programas han sido definidos como grupos de control, mientras que quienes completan los DVPP han sido definidos como el grupo de tratamiento. En consecuencia, el grupo de tratamiento y el grupo de control no son equivalentes, lo que constituye una amenaza para la validez interna (Shadish, Cook & Campbell, 2002) de tales estudios. En su conocido estudio de evaluación, Gondolf (2002) halló pequeños efectos para quienes completan los programas en comparación con los abandonos. Con el fin de abordar el problema del grupo de control no equivalente, Gondolf utilizó el análisis de puntuación de propensión y el análisis de variables instrumentales para ajustar los resultados. Dichos procedimientos supusieron una mayor eficacia. El principio de tales procedimientos de ajuste es el siguiente: La diferencia entre los resultados del grupo de tratamiento y el grupo de control no equivalente se corrige estadísticamente empleando las diferencias entre los grupos (que se deben conocer al comienzo del estudio) a modo de una especie de variables de control. Como resultado, se obtiene un tamaño de efecto que refleja qué se puede esperar como efecto, asumiendo que ambos grupos fueran lo más similares posibles. Este tamaño de efecto corregido es considerado más apropiado que el inicial sin corregir por los expertos y expertas en el ámbito estadístico (y el tamaño de efecto sin corregir se considera inapropiado).

Muchos estudios de evaluación anteriores se han centrado en los hombres que participan en los programas. En aquellos casos donde las víctimas/supervivientes se han incluido como fuente de información acerca del impacto del programa, a menudo los investigadores han estado primero en contacto con los agresores y, posteriormente, han tratado de ponerse en contacto con las correspondientes víctimas/supervivientes. Este enfoque también puede proseguir en el futuro, tal como ha demostrado Gondolf (2002) en su estudio, se podía establecer contacto con hasta el 60% de las víctimas/supervivientes iniciales para el seguimiento.

No obstante, también existen enfoques alternativos, como el siguiente, que se ha inspirado en el proyecto MIRABAL⁴. El énfasis de la investigación puede desplazarse de los hombres que participan en los programas a las víctimas/supervivientes en los centros de apoyo a las víctimas. Puesto que existen regiones geográficas con centros de apoyo a las víctimas –pero con y sin DVPP–, el efecto adicional de un DVPP se puede estudiar cuando los resultados de las víctimas/supervivientes en esa región o ciudad se comparan con una región/ciudad *sin* un DVPP (los centros de apoyo a las víctimas existen en ambos tipos de regiones). En este diseño, las víctimas/supervivientes reciben apoyo y son invitadas a seguimiento durante cierto tiempo (T1-T4), así como sus parejas –que son derivadas a un DVPP si hay uno disponible (grupo de tratamiento) o que no pueden ser derivadas a un DVPP porque no hay ninguno (grupo de control). Sobre la base de la información de las víctimas/supervivientes acerca de sus parejas, incluso es posible realizar una especie de asignación entre grupo de tratamiento y grupo de control durante T0. Aunque parece útil y significativo incluir la perspectiva de los hombres en la evaluación y proceder tal como se ha indicado arriba (“Observaciones / Tiempo”), incluso podríamos pensar en tener en cuenta únicamente la perspectiva de las víctimas/supervivientes, además de cierta información adicional facilitada por los y las profesionales acerca del abandono, la “dosis” o la finalización de los programas. En este caso, el DVPP se considera una “caja negra”, administrada a un grupo de agresores con una determinada dosis (dependiendo de la duración de su asistencia). El grupo de tratamiento (víctimas/supervivientes cuyas parejas están en un programa) se compara entonces con el grupo de control (víctimas/supervivientes cuyas parejas no están en un programa). Como condición previa, los contextos de ambas regiones deben ser muy similares, de tal modo que las diferencias entre el grupo de tratamiento y el grupo de control se puedan atribuir a la (no) existencia de un DVPP. Aunque los procedimientos detallados son más complejos⁵, la idea básica de estudiar el impacto de un DVPP por los resultados diferenciales de víctimas/supervivientes que están en programas de apoyo a las víctimas en diferentes regiones (con/sin DVPP) es prometedora.

De forma similar, también se pueden desarrollar otros diseños, por ejemplo, diseños de series de tiempo interrumpido, donde se realiza el seguimiento a las víctimas/supervivientes durante un lapso de tiempo que se corresponde con T0-T5 durante un largo período de

⁴ Los resultados del proyecto MIRABAL se presentarán a principios de 2015. El enfoque y la metodología del proyecto deben ser retomados por un futuro grupo de trabajo de evaluación de WWP-EN.
<https://www.dur.ac.uk/criva/projectmirabal/>

⁵ Por ejemplo, los grupos deben ser definidos post hoc: Abandonos tanto de las víctimas/supervivientes como de los agresores; hombres que completan los programas y (ex) parejas de hombres que completan los programas; lo mismo con hombres excluidos y no admitidos; etc.

tiempo, y en un determinado momento se instala un DVPP en la región. La hipótesis es que los resultados relacionados con la situación vital de las víctimas/supervivientes mejoran en cuanto sus parejas son derivadas a un DVPP de forma sistemática. Los diseños de series de tiempo también son adecuados para la investigación de caso único. En especial en marcos con muestras pequeñas no debería subestimarse esta posibilidad. La metodología de caso único, también a nivel estadístico, constituye un desarrollo prometedor y se puede usar en el contexto de evaluación de DVPP. Resulta tentador considerar cómo los efectos de los DVPP se pueden estudiar basándose en los datos de caso único. Este enfoque es especialmente útil cuando diferentes personas reciben diferentes tratamientos y diferentes intensidades de tratamiento (“dosis”), en función de sus necesidades individuales.

El proyecto MIRABAL ha anunciado que aportará otro enfoque cuantitativo progresivo a la evaluación de los programas para agresores, a saber, el *análisis de supervivencia*. Este tipo de análisis proviene originalmente de la investigación médica y el cálculo de la esperanza de vida, pero también se utiliza en otros ámbitos. El análisis de supervivencia parte de tablas de mortalidad a partir de las cuales se calculan las probabilidades acumuladas de supervivencia. Dichas probabilidades, que por lo general se muestran en forma de gráficos a lo largo de una cadena temporal, se pueden determinar para diferentes subgrupos, por ejemplo, personas que reciben tratamientos distintos. Asimismo, se pueden observar los efectos de las covariables. Recomendamos adoptar este enfoque innovador en un futuro grupo de trabajo de evaluación de WWP-EN.

Los resultados de los estudios de evaluación básicos tal como se han descrito aquí serán afirmaciones acerca del efecto adicional de un DVPP en una CCR (en comparación con una “CCR sin un DVPP”).

Independientemente del diseño que se adopte, los diferentes estudios solo se pueden comparar si se utiliza el esquema de T0-T4 de manera coherente (para saber qué tipos de hombres están participando en cada programa, quién está abandonando y cuándo), y si se utilizan las mismas mediciones de los resultados.

(c) Criterios de éxito y medición de los resultados

En un proyecto de investigación anterior para el proyecto MIRABAL, Westmarland, Kelly y Chalder-Mills (2010) propusieron los siguientes “criterios de éxito” de DVPP:

- Mejora de la relación entre los hombres que participan en los programas y las (ex) parejas; respeto y mejora de la comunicación.
- Aumento del margen de acción para las (ex) parejas, recuperando su voz, su capacidad de tomar decisiones, mejorar su bienestar.
- Seguridad y ausencia de violencia y maltrato a mujeres y niños.
- Paternidad segura, positiva y compartida.
- Mejora de la conciencia de uno mismo y de los demás por parte de los hombres, comprensión del impacto de la violencia en las parejas y los hijos.
- Infancias más seguras y saludables, los niños se sienten escuchados y atendidos

Estos criterios se emplearon para formular las preguntas y los elementos concretos para el estudio MIRABAL (véase Westmarland & Kelly, 2012). Las perspectivas de las

víctimas/supervivientes y los niños son la principal fuente de información en este enfoque para evaluar el impacto de un DVPP. Idealmente, esta información se complementa mediante la perspectiva de los hombres que participan en los programas, con las mediciones de los resultados detalladas anteriormente. Los profesionales y los registros oficiales pueden aportar información adicional que se puede utilizar para contestar preguntas que van más allá de una evaluación de los resultados de DVPP. Con el fin de anclar los futuros estudios de evaluación en un ejemplo de buenas prácticas, recomendamos que un futuro grupo de trabajo de evaluación de WWP-EN tenga en cuenta estos indicadores.

4.5.2. Estudios multisitio y multicontexto

En los *estudios básicos de contexto único*, el énfasis se situaba en determinar el impacto de los DVPP cuando las condiciones son bastante similares en ambas áreas. Ahora queremos centrarnos en otro caso, cuando se están realizando DVPP *similares en condiciones diferentes*. La Figura 4 aporta un marco sobre cómo aproximarse a esta constelación. El contexto de un DVPP está constituido por los componentes locales relevantes que pertenecen al sistema de intervención en el caso de violencia de género (policía, tribunales, protección de las víctimas, etc.), así como por las condiciones a un nivel más elevado, por ejemplo, la situación de la financiación a escala regional o leyes, campañas y medidas preventivas a escala nacional, etc. Las condiciones internacionales y globales también pueden influir en los resultados del sistema.

En los *estudios básicos de contexto único*, hemos asumido que el contexto es constante y existen dos áreas (con y sin DVPP) en un país, con un contexto local y regional igual o muy parecido. En los *estudios multisitio multicontexto*, asumimos que se realizan DVPP muy similares en condiciones diferentes, como es de esperar en diferentes países.

Los resultados de la Línea de trabajo 1 han arrojado que existen programas similares en diferentes países –probablemente debido al hecho de que muchos programas se han desarrollado a partir de las mismas raíces, tales como el modelo tipo Duluth o el programa CHANGE. Si en dos o más países se pueden identificar DVPP muy similares, a ser posible con la misma vía de derivación (por orden judicial, voluntario o ambos), se pueden comparar los resultados de las intervenciones de los sistemas. En este caso, podemos asumir que la forma de trabajo del DVPP es constante y que el contexto es diferente, lo que comporta resultados diferentes. En este tipo de diseño, el *papel del contexto y su impacto en los resultados de un DVPP* se comparan y estudian. Es exactamente en este punto donde tiene sentido adoptar un enfoque multipaís.

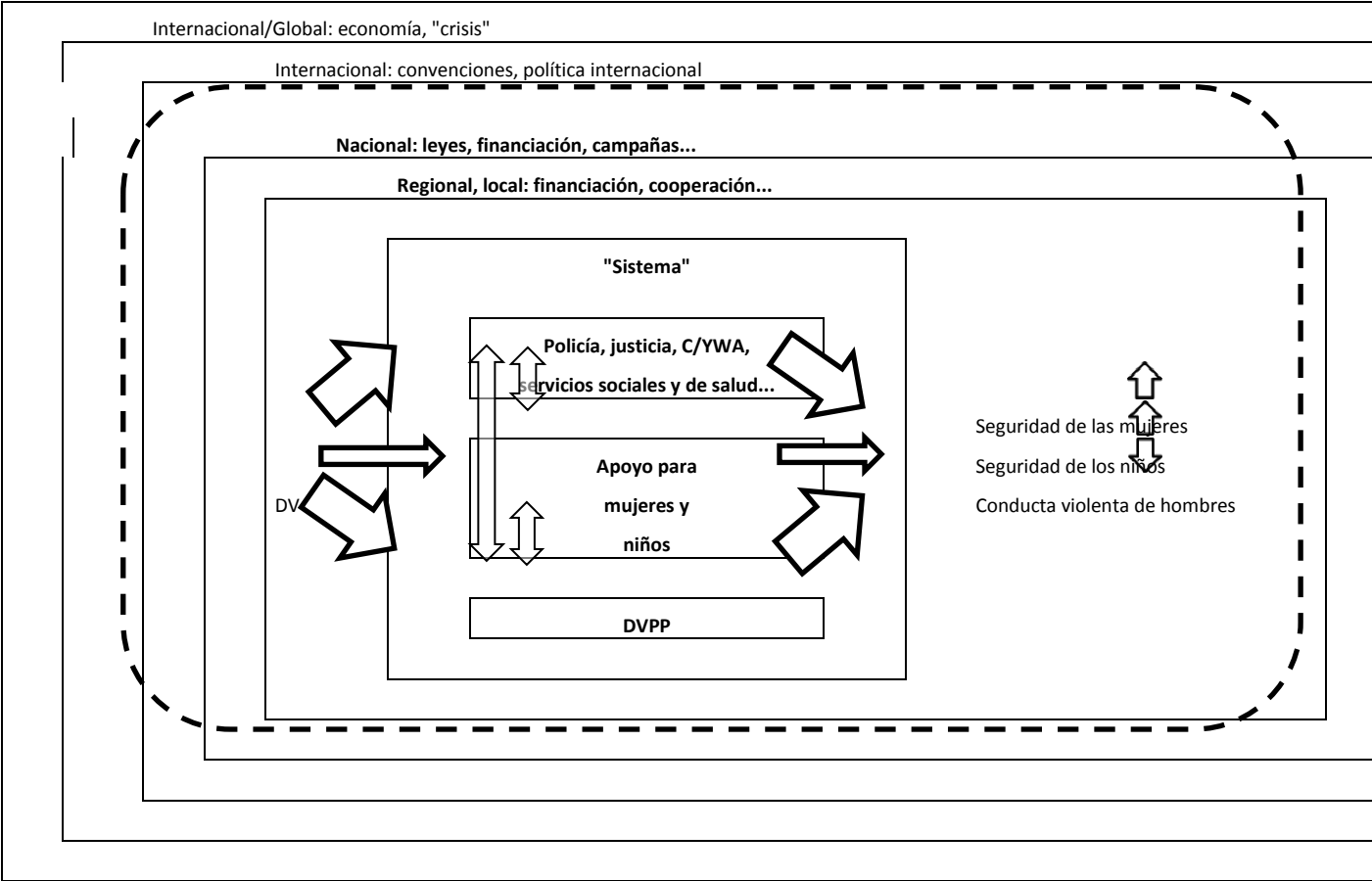
De nuevo, para poder vincular este estudio a otras evaluaciones, es necesario aplicar la misma metodología básica en todos los estudios. Como resultado de un estudio multisitio multicontexto, podemos esperar afirmaciones acerca de la influencia del contexto en los resultados de un determinado tipo de DVPP.

Si el contexto se debe analizar con más profundidad y no se debe tratar simplemente como una “caja negra” (con un impacto global conocido, pero con un impacto desconocido sobre los componentes individuales), el reto consiste en *operacionalizar el contexto*, a fin de determinar la influencia de los componentes en los resultados del DVPP. Empezando por la

Figura 4, la tarea consiste en transformar el contexto en variables para su análisis cuantitativo. Varios componentes se deben “traducir” de tal modo que sea posible su integración en procesos estadísticos cuantitativos:

- A escala local: apoyo a mujeres y niños, el trabajo de la policía, sistema de justicia penal, sistemas de bienestar infantil y juvenil, servicios sociales y de salud.
- A escala regional y nacional: medidas legales, medidas relativas a la prevención de la violencia, situación de la financiación, situación general de la igualdad de género.

Figura 4. Un marco para diferentes tipos de estudios de evaluación de DVPP.



Nota. DV... Casos de violencia de género; DVPP... Programa para agresores de violencia de género; C/YWA... Servicio de bienestar de la infancia/juventud. Los casos de violencia de género confluyen en un sistema de intervención que está compuesto por varios componentes que interactúan. Los resultados de las intervenciones del sistema dependen de cada componente y sus interacciones, así como de las condiciones regionales, nacionales e internacionales. Desde la perspectiva de un DVPP, los otros componentes del sistema y las condiciones a niveles superiores (nacional, internacional) constituyen el *contexto* del DVPP.

A día de hoy falta este tipo de operacionalización, y su desarrollo requerirá recursos significativos. Aquí solo queremos dar algunas pistas sobre cómo podría realizarse esa operacionalización, a modo de preparación para futuros *estudios multisitio multicontexto*.

A *nivel de sistema*, las reacciones de los tribunales se han identificado como influencias relevantes en los resultados de los DVPP, en el caso de agresores derivados por orden judicial (por ejemplo, Gondolf, 2002). Para operacionalizar esta influencia, se puede tomar como punto de partida la perspectiva de los hombres que participan en el programa, los profesionales, las víctimas/supervivientes o los expertos y expertas. Si utilizamos la perspectiva de los *hombres que participan en el programa por orden judicial, valorados por los y las profesionales*, se podrían emplear los indicadores siguientes:

A. Experiencias y expectativas de los clientes acerca de las reacciones de los tribunales

- “En caso de incumplimiento, se ha producido (o se espera que se produzca) una reacción rápida y consecuente por parte del tribunal.” (Valor=1)
- “En caso de incumplimiento, se ha producido (o se espera que se produzca) una reacción lenta y errática por parte del tribunal.” (0,5)⁶
- “En caso de incumplimiento, no se ha producido (o no se espera que se produzca) reacción alguna por parte del tribunal.” (0)

B. Significado subjetivo de las consecuencias anunciadas o impuestas

- “Las consecuencias dictadas por el tribunal son significativas/severas a ojos del usuario.” (1)
- “Las consecuencias dictadas por el tribunal tienen cierto significado negativo para el usuario.” (0,5)
- “Las consecuencias dictadas por el tribunal no tienen sentido a ojos del usuario.” (0)

Se precisa la información por caso acerca de A, B, A*B para todos los componentes relevantes de un sistema, y desde varias perspectivas (hombres, profesionales, víctimas/supervivientes). De este modo, se puede introducir mucha información relevante para el caso sobre el contexto en una matriz de datos que se utiliza para realizar, por ejemplo, un análisis multinivel para evaluar la influencia de las variables y los niveles en los resultados del DVPP. En un nivel superior, los expertos pueden evaluar las reacciones de los tribunales *en general*, lo que también se introduce como variable en la matriz de datos. Las influencias relevantes se determinarán empíricamente en un estudio de este tipo.

Para determinar los componentes relevantes, el desarrollo de una herramienta para comparar contextos puede partir de la investigación comparativa sobre las medidas legales y de apoyo, como EIGE (2012). También existen proyectos anteriores que se han propuesto asignar y comparar los contextos sobre las medidas para víctimas/supervivientes en los diferentes países europeos (por ejemplo, EUCPN, 2013). Este trabajo muestra qué medidas legales y prácticas se han implantado en las diferentes naciones.

Otro ejemplo de un indicador potencialmente útil es el Índice de Igualdad de Género (IIG), desarrollado por EIGE (2012). Aunque la violencia no forma parte del IIG debido a la falta de datos comparables, el IIG se puede usar como índice global para determinar la posición de un país en relación con la igualdad de género, y puede formar parte de una ecuación como variable independiente o de control.

⁶ O un conjunto diferente de alternativas entre 0 y 1.

Otro enfoque interesante ha sido propuesto por Hester y Lilley (2014), señalando la necesidad de alterar varios factores de riesgo de violencia en varios niveles dentro de un marco ecológico (sociedad, organismos, familia/amigos, individuo). Esta propuesta se basa en un modelo interactivo de factores de riesgo que ha sido desarrollado por Hagemann-White y sus colegas (véase Comisión Europea, 2010). Su modelo se puede emplear para definir vías de prevención de la violencia en relaciones; los componentes preventivos relevantes se pueden operacionalizar y utilizar como variables en análisis multinivel.

4.5.3. Estudios multisitio de contexto único

El énfasis de este tipo de estudios es comparar diferentes tipos de DVPP dentro de un mismo contexto. Resulta difícil determinar qué conduce a resultados diferentes al estudiar *diferentes* programas en *diferentes* contextos, por lo que se recomienda mantener el contexto lo más constante posible, por ejemplo, comparando dos tipos de programas en una región.

Este enfoque no solo aborda diferentes formas de trabajar con agresores (por ejemplo, más del estilo TCC o más psicodinámico; o con/sin servicios adicionales, como tratamiento del alcoholismo o psicoterapia), sino que también se pueden comparar diferentes vías de derivación, por ejemplo, programas TCC en una región, donde un programa trabaja con hombres derivados por orden judicial, mientras otro trabaja con hombres que acuden voluntariamente, etc. El énfasis no debería ser sencillamente determinar qué programa funciona mejor que otro, sino qué enfoque funciona mejor para qué clientes, lo que significa que es necesario estudiar las características de los hombres en la fase de acogida. El debate sobre diferentes tratamientos para diferentes tipos de agresores puede ser abordado mediante estudios que comparen varios programas (pero dentro de un contexto), y también se pueden tratar cuestiones relacionadas con la intensidad del tratamiento. Los programas estándares pueden proporcionar un punto de referencia con respecto al cual se comparan programas con nuevos elementos.

De nuevo, es fundamental emplear la misma metodología en todos los estudios y, en este ejemplo, la información recabada durante T0 acerca de los hombres que participan en los programas, es de gran relevancia. Los estudios multisitio de contexto único son el enfoque correcto para aprender más acerca de la interacción de marcos de trabajo, métodos, tipos de agresores, etc. Como resultado, podemos esperar conclusiones sobre los efectos diferenciales de diversos tipos de DVPP para varios tipos de agresores.

4.5.4. Estudios similares a estudios de laboratorio

Finalmente, existe un tipo de estudios de evaluación donde no es significativo o no es posible (por ejemplo, por ley) incluir las perspectivas de las víctimas/supervivientes (por ejemplo, en el contexto de programas para agresores en las prisiones en España). Por otro lado, los únicos ECA se han encontrado en instituciones cerradas como prisiones u hospitales psiquiátricos, dentro de la búsqueda realizada en la Línea de trabajo 2 (Hester et al., 2014). Las prisiones proporcionan un marco donde no es posible observar los impactos de los

programas para agresores de forma inmediata, porque los presos no están conviviendo con sus parejas y, en general, tan solo tienen contacto limitado con ellas. No obstante, como parte de la rehabilitación de los delincuentes, los programas de las prisiones preparan a quienes participan en ellos para su vida tras el encarcelamiento y deben ser vistos como una intervención importante y significativa.

En términos de evaluación de DVPP, el papel de los programas en las prisiones se puede comparar con un entorno de laboratorio de experimentación psicológica. En las prisiones, en principio, es posible realizar ECA. En especial los diseños de lista de espera (es decir, las personas que no reciben tratamiento y que se definen como grupo de control reciben tratamiento posteriormente) son una forma factible de realizar ECA en prisiones sin mayores implicaciones éticas. Por lo tanto, los estudios en las prisiones pueden aspirar a mejorar los programas, centrándose en varios componentes del tratamiento y estudiándolos de manera sistemática. Los componentes efectivos se pueden implementar en programas sobre el terreno (y también se pueden verificar y validar mediante *estudios multisitio de contexto único*).

Como resultado de los estudios similares a los estudios de laboratorio, se pueden esperar afirmaciones diferenciadas acerca de los efectos de los tratamientos en el marco de una prisión (mayoritariamente basados únicamente en la perspectiva de los hombres), así como afirmaciones acerca de la eficacia de determinados componentes de los tratamientos. Se pueden realizar estudios de correlación con reincidencia; no obstante, la reincidencia depende de muchas condiciones e influencias, no de los resultados de aprendizaje de un hombre en un programa en la prisión, por lo que las expectativas con respecto a este tipo de validación no deberían ser demasiado elevadas.

5. Resumen y próximos pasos

Se ha tenido que modificar la idea inicial de un gran estudio de evaluación multipaís que aprovechara la diversidad europea en enfoques para el trabajo con agresores y que respondiera a una serie de preguntas de una vez. Nos centramos en el trabajo con agresores de violencia de género y en enfoques coordinados, por diversos motivos, pero también porque los expertos y expertas de las áreas de investigación y práctica nos aconsejaron hacerlo en los talleres realizados dentro de este proyecto. Hemos perfilado los fundamentos de una metodología para este ámbito, que aún se debe desarrollar en detalle, con el fin de avanzar hacia procedimientos más sistemáticos y armonizados en la evaluación de DVPP, tanto a escala nacional como internacional. La nueva organización internacional WWP-EN puede y debe desempeñar un papel importante en este sentido, porque puede desarrollar y promover estándares para la evaluación de DVPP entre sus miembros, y muchos futuros proyectos de evaluación tendrán que cooperar con estos. Un paso de gran relevancia que debe darse ahora es designar a un grupo investigador y de profesionales dentro de WWP-EN que asuman la responsabilidad de estas tareas. Cuando desarrollen métodos y estándares, pueden partir de los resultados del proyecto IMPACT en general y las recomendaciones de la Línea de trabajo 3 en particular.

Existen algunas cuestiones específicas en relación con la evaluación de los DVPP; una de las más importantes es la idea de que la información más fiable sobre cambios relevantes proviene de las parejas e hijos e hijas de los agresores. Como consecuencia, es necesaria la

cooperación con organizaciones de protección de las víctimas, tanto en el nivel concreto de centros individuales como a nivel de redes con organizaciones paraguas como WAVE. Ha resultado que un enfoque de sistema es el más apropiado, lo que significa que no debemos esforzarnos por evaluar programas aislados para agresores, sino abordar el sistema de intervención en su conjunto. Las maneras sobre cómo proceder a este respecto se han perfilado en este documento de trabajo, adoptando nuevos desarrollos y ejemplos de buenas prácticas en la evaluación de DVPP. Hemos propuesto un marco para estudios de evaluación que permite vincular y comparar estudios individuales, y aportar varios diseños a un sistema significativo, basándose en los desarrollos de otras Líneas de trabajo del proyecto IMPACT, estudios anteriores y algunos desarrollos prometedores en la evaluación de DVPP. Esperamos que esta propuesta contribuya a superar la situación dispersa y más bien problemática en relación con la comparabilidad de las evaluaciones de DVPP hoy en día. Paso a paso, deberíamos ser capaces de detectar y demostrar tamaños de efecto relevantes de intervenciones optimizadas (“Quién necesita qué”), tanto para agresores como para víctimas/supervivientes. Tener tamaños de efecto fiables para DVPP también constituye la base para análisis significativos coste-beneficio y coste-eficacia. En los sectores social y de salud, tales análisis normalmente prueban que los recursos asignados a intervenciones no son solo una cuestión de ética, sino que también están ayudando a ahorrar recursos públicos. Un resultado frecuente es que los costes de las intervenciones son bajos en comparación con los costes de las consecuencias de la inactividad. Más aspectos y consideraciones importantes se publicarán en documentos separados, basados en el trabajo realizado en la Línea de trabajo 3 del proyecto IMPACT.

6. Bibliografía

- Alderson, S., Kelly, L. and Westmarland, N. (2013). Domestic violence perpetrator programmes and children and young people. Londres y Durham: London Metropolitan University y Durham University.
- Alderson, S., Westmarland, N. & Kelly, L. (2012). The need for accountability to, and support for, children of men on domestic violence perpetrator programmes. *Child Abuse Review*, 3, 182-193.
- Akoensi, T., Koehler, J. A., Lösel, F. & Humphreys, D. K. (2013). Domestic violence perpetrator programs in Europe, Part II: A systematic review of the state of evidence. *International Journal of Offender Therapy and Comparative Criminology*, 57 (10), 1206-1225.
- Babbie, E. (2003). *Badania społeczne w praktyce*. Varsovia: Wydawnictwo Naukowe PWN.
- Bamberger, M., Rugh, J. & Mabry, L. (2006). *RealWorld evaluation. Working under budget, time, data, and political constraints*. Thousand Oaks: Sage
- Cavanagh, K. & Lewis, R. (1996). Interviewing violent men: challenge or compromise? In K. Cavanagh & V.E. Cree (eds.), *Working with men. Feminism and social work*. Londres: Routledge.
- Consejo de Europa (2011). Council of Europe Convention on preventing and combating violence against women and domestic violence. Obtenido en <http://www.coe.int/t/dghl/standardsetting/convention-violence/convention/Convention%20210%20English.pdf>

- Cowburn, M. (2013). Men researching violent men: Epistemologies, ethics and emotions in qualitative research. In B. Pini & B. Pease (eds.), *Men, masculinities and methodologies*. Nueva York: Palgrave Macmillan.
- Dobash, R. E. & Dobash, R. P. (2000). Evaluating criminal justice interventions for domestic violence. *Crime & Delinquency*, 46 (2), 252-271.
- Dobash, R. E., Dobash, P. R. (1998) Violent men and violent contexts. In R. E. Dobash & R. P. Dobash (eds.), *Rethinking violence against women*. Thousand Oaks, CA: Sage.
- Dobash R. E., Dobash, P. R., Cavanagh, K. & Lewis, R. (2000). *Changing violent men*. Thousand Oaks, CA: Sage.
- Downes, J., Kelly, L. and Westmarland, N. (2014) Ethics in violence and abuse research - a positive empowerment approach. *Sociological Research Online*, 19 (1) 2.
- EIGE (2013). Gender Equality Index. Report. Vilnius: European Institute for Gender Equality. Obtenido en <http://eige.europa.eu/content/document/gender-equality-index-report>
- EUCPN (2013). Tackling domestic violence in the EU – policies & practices. In EUCPN Secretariat (eds.), *EUCPN Toolbox Series*, no. 4, European Crime Prevention Network: Bruselas.
- Comisión Europea (2010). Feasibility study to assess the possibilities, opportunities and needs to standardize national legislation on violence against women, violence against children and sexual orientation violence. Luxemburg: Publications Office of the European Union.
- Geldschläger, H., Ginés, O., Nax, D. & Ponce, A. (2014). Outcome measurement in European perpetrator programmes: A survey. Obtenido en http://www.webseiten.cc/fileadmin/WWP_Network/redakteure/IMPACT/Daphne%20III%20Impact%20Working%20Paper%201%20rev.pdf
- Gondolf, E. W. (2002). *Batterer intervention systems: Issues, outcomes, and recommendations*. Thousand Oaks, CA: Sage.
- Gottzen, L. (2013). Encountering violent men: Strange and familiar. In B. Pini & B. Pease (eds.), *Men, masculinities and methodologies*. Nueva York: Palgrave Macmillan.
- Hamilton, L., Koehler, J. A. & Lösel, F. (2013). Domestic violence perpetrator programs in Europe, Part I: A survey of current Practice. *International Journal of Offender Therapy and Comparative Criminology*, 57 (10), 1189-1205.
- Hester, M. & Lilley, S.-J. (2014). Preventing violence against women: Artículo 12 del Convenio de Estambul. Una serie de artículos del Convenio del Consejo de Europa sobre la prevención y la lucha contra la violencia contra la mujer y la violencia doméstica. Estrasburgo: Consejo de Europa.
- Hester, M., Lilley, S.-J., O'Prey, L. & Budde, J. (2014). Overview and analysis of research studies evaluating European perpetrator programmes. Obtenido en http://www.webseiten.cc/fileadmin/WWP_Network/redakteure/IMPACT/Daphne%20III%20IMPACT%20Working%20Paper%202%20Overview%20and%20analysis%20of%20research%20studies%20Evaluating%20European%20perpetrator%20programmes.pdf

- Johnson, H. (1998). Rethinking survey research on violence against women. In R. E. Dobash & R. P. Dobash (eds.), *Rethinking violence against women*. Thousand Oaks, CA: Sage.
- Kutrovátz, K. (2013) *Work with Perpetrator – European Network national report Hungary*. Obtenido en http://wwp.webseiten.cc/fileadmin/WWP_Network/redakteure/Reports_2013/National_report_Hungary.pdf
- Patulny, R., Pini, B. (2013) *Counting men: quantitative approaches to the study of men and masculinities*. In B. Pini & B. Pease (eds.), *Men, masculinities and methodologies*. Nueva York: Palgrave Macmillan.
- Payarola, M. A. (2011) *Description of a voluntary attendance program for abusive men*. In C. Blazina & D. S. Shen-Miller (eds.), *An international psychology of men. Theoretical advances, case studies, and clinical innovations*. Nueva York: Routledge.
- Phillips, R., Kelly, L. & Westmarland, N. (2013) *Domestic violence perpetrator programmes: A historical overview*. Londres y Durham: London Metropolitan University y Durham University.
- Price, B. J. & Rosenbaum, A. (2009). *Batterer intervention programs: A report from the field*. *Violence and Victims*, 24 (6),747-770.
- Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). *Evaluation. A systematic approach (7ª ed.)*. Thousand Oaks: Sage.
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston Houghton Mifflin Company.
- Silverman, D. (2013). *Prowadzenie badań jakościowych*. Varsovia: Wydawnictwo Naukowe PWN.
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco: Jossey-Bass.
- Westmarland, N. & Kelly, L. (2012) (con Kelly). *Why extending measurements of ‘success’ in domestic violence perpetrator programmes matters for Social Work*. *British Journal of Social Work*, 43(6), 1092-1110.
- Westmarland, N., Kelly, L. & Chalder-Mills, J. (2010). *What Counts as Success?* Londres: Respect.
- Organización Mundial de la Salud (2002). *World report on violence and health. Summary*. Obtenido en http://www.who.int/violence_injury_prevention/violence/world_report/en/summary_en.pdf
- WWP European Network (2008). *Guidelines to develop standards for programmes working with male perpetrators of domestic violence*. Obtenido en <http://www.work-with-perpetrators.eu/index.php?id=86>