

Informe Final

Diseño e Implementación de una Evaluación de Impacto del Programa Presupuestario S191 “Sistema Nacional de Investigadores (SNI)”

Raymundo M. Campos Vázquez (El Colegio de México)

Índice Temático

1. Introducción	5
2. Estado del arte en evaluaciones de impacto de intervenciones en Ciencia, Tecnología e Innovación (CTI)	11
3. Hipótesis de impacto a probar con la evaluación	19
4. Efectos potencialmente atribuibles al SNI e indicadores asociados	20
5. Limitaciones de una evaluación en el área de Ciencia, Tecnología e Innovación	22
6. Fuentes de información disponibles	24
7. Metodología	26
a. Metodología cuantitativa	27
b. Metodología cualitativa	30
8. Construcción de base de datos final	31
9. Análisis de poder estadístico	39
10. Resultados para evaluación de impacto con metodología cuantitativa	41
a. Metodología 1: Regresión múltiple	42
b. Metodología 2: Método de apareamiento	48
11. Resultados para evaluación de impacto con metodología cualitativa	56
a. Frecuencias de palabras	57
b. Sentimiento en las preguntas	61
12. Retos y avenidas de mejora usando ambas evaluaciones	78
13. Conclusiones y recomendaciones	81
Referencias bibliográficas	86
Anexos	90

Índice de Figuras

Figura 1. Número de investigadores por año	32
Figura 2. Resultados por área de conocimiento.....	46
Figura 3. Densidades y balance en los diferentes métodos.....	53
Figura 4. Resultados por área de conocimiento.....	56
Figura 5. Nubes de palabras utilizadas en cada respuesta.....	57
Figura 6. Nube de palabras usando las 4 preguntas utilizadas.....	60
Figura 7. Número de apariciones en las preguntas para las 25 palabras más utilizadas	61
Figura 8. Número de investigadores que respondieron encuesta.....	62
Figura 9. Pregunta 1: Representa estímulo económico el SNI. (Porcentaje de respuesta)	63
Figura 10. Pregunta 1: Representa motivación el SNI. (Porcentaje de respuesta)	64
Figura 11. Pregunta 2: Experiencia en evaluación SNI. (Porcentaje de respuesta)	65
Figura 12. Pregunta 3: Influencia del SNI. (Porcentaje de respuesta).....	66
Figura 13. Pregunta 4: ¿Es adecuada la forma que incentiva el SNI? (Porcentaje de respuesta)	67
Figura 14. Pregunta 1-4: ¿Reglas claras en el SNI? (Porcentaje de respuesta)...	69
Figura 15. Pregunta 1-4: Tipo de queja mencionada en las respuestas. (Porcentaje de respuesta)	72
Figura 16. Pregunta 1-4: Palabra más mencionada en las respuestas (Porcentaje de respuestas).....	73
Figura A1. Resultados de método de apareamiento usando método 2.....	90
Figura A2. Resultados de método de apareamiento usando método 3.....	91
Figura A3. Resultados de método de apareamiento usando método 4.....	92

Índice de Tablas

Tabla 1. Investigadores por distinción. Años 2005, 2010, 2015 y 2018	33
Tabla 2. Investigadores evaluados y decisión de evaluación.....	33
Tabla 3. Muestra final y división por nivel antes y después de la evaluación, así como por tiempo entre evaluaciones.....	35
Tabla 4. Estadísticas descriptivas en la muestra final.	36
Tabla 5. Diferencia de medias entre grupos de tratamiento y control.	37
Tabla 6. Análisis de poder estadístico. Diferencia de medias (δ) asumiendo diferentes varianzas y poder estadístico.	40
Tabla 7. Resultados con método de regresión.	44
Tabla 8. Balance en características observables.	51
Tabla 9. Resultados de Diferencias en Diferencias con el método de apareamiento.....	54
Tabla 10. Respuestas más comunes	75

1. Introducción

El elemento clave que permite el desarrollo tecnológico, económico y social en un territorio es el capital humano. Este, asimismo, ayuda a conocer si una nación puede satisfacer sus procesos científicos, tecnológicos e innovadores en pro de otorgar beneficios a la población. En este sentido, y debido al avance sustancial para mejorar la ciencia y la tecnología, surge la necesidad de evaluar el impacto de los programas de apoyo que estimulan y fomentan el capital humano en materia de ciencia, tecnología e innovación (CTI).

México, como el resto de los países, en los últimos 40 años ha tomado medidas para mejorar su nivel científico. El Consejo Nacional de Ciencia y Tecnología (CONACYT) es el encargado de ejecutar los programas para el mejoramiento científico, pero, sobre todo, es su Programa Presupuestario S191- Sistema Nacional de Investigadores (SNI) que toma el liderazgo de los programas en general por la sencilla razón de que es el que estimula económicamente y fomenta la investigación científica nacional. Este programa es un sistema de incentivos económicos y no económicos por nombramiento (Nivel de distinción) para la investigación creado en 1984 con el objeto de mejorar el ingreso real de las personas dedicadas a la investigación que se vieron impactadas negativamente debido a las condiciones macroeconómicas y reconocer la labor de las personas dedicadas a producir conocimiento científico y tecnológico. En el primer año de operación del programa fueron reconocidos 1,396 investigadores, actualmente (2018) existen aproximadamente 30,000 mil investigadores registrados en el SNI. Sin lugar a duda, el SNI ha sido y es una de las políticas públicas en materia de apoyo hacia la investigación científica más importantes en los últimos 40 años.

En este contexto, y dado la importancia del SNI, se plantea que este trabajo tenga como objetivo elaborar el diseño e implementación de una evaluación de impacto para el programa. Es importante destacar que la evaluación continua de la investigación científica, así como, del otorgamiento responsable de los nombramientos e incentivos económicos por parte del SNI ha tomado la envergadura de un proceso de evaluación complejo. Como se discute en el Manifiesto Leiden (Hicks et al. 2015), la producción científica debe evaluarse en términos cuantitativos y cualitativos con visión de largo plazo que promueva la solución de problemas de la sociedad.

Por tanto, una evaluación de impacto al sector de CTI, y en particular el SNI, enfrenta retos y limitaciones importantes. Por un lado, la asignación del “tratamiento” a los investigadores no se realizan mediante diseños experimentales por lo que es complicado obtener el impacto causal de un nivel SNI en resultados (cualesquiera que sean). Por otro, no es posible tomar un enfoque de largo plazo o bien tomar todas las variables importantes a considerar. Por ejemplo, sería ideal saber si cierto tipo de producción académica ocasionó una mejora de bienestar en la sociedad, pero no es posible recuperar esa información. En la práctica se tienen múltiples variables no observadas que pudieran tener repercusiones importantes. En este sentido es prácticamente imposible realizar una evaluación de impacto integral con visión de largo plazo y que incluya múltiples indicadores. Sin embargo, a pesar de sus retos y limitaciones, una evaluación de impacto al programa es deseable e importante, aunque sea muy focalizada. Dentro de lo que es posible evaluar puede dar luz sobre posibles avenidas para mejorar el programa de tal forma que los investigadores y la sociedad se beneficien.

El CONACYT tiene información sobre la producción académica de los investigadores en el SNI. Esta información es incluida en el currículum vitae único (CVU). El CONACYT compartió esa información para poder realizar la evaluación de impacto. Dado los insumos, en esta evaluación se decide tomar en cuenta como variables de resultados los artículos producidos, capítulos de libros, libros y adicionalmente se calcula el número de artículos producidos ponderados por el factor de impacto. Estas variables son importantes, pero también tienen limitaciones. No es posible observar ex-post el número de citas que tenía cada artículo antes de que el investigador sea evaluado. Por tanto, no es posible analizar citas de cada unidad de producción académica. Tampoco es posible observar la significancia social y científica de cada producto académico. Por tanto, hay que tener estas limitaciones en mente cuando se analicen resultados de la evaluación.

Utilizamos la información de investigadores evaluados en el periodo 2005-2017 y explotamos toda la producción académica en este periodo. Al igual que CONACYT, dependemos de que la información en el CVU sea verídica. Lo único que es posible evaluar cuasi-experimentalmente dadas las restricciones de la información es el cambio de nivel del SNI. El SNI tiene tres categorías: Candidato a Investigador Nacional, Investigador Nacional e Investigador Nacional Emérito. La categoría de Investigador Nacional tiene tres niveles: 1, 2 y 3. Dentro de los cambios, el trabajo se enfoca en los investigadores que subieron de nivel 1 a 2, así como los que subieron de nivel 2 a 3. Este grupo se define como el grupo de tratamiento. El grupo de control se refiere a los investigadores que iniciaron en los mismos niveles, pero no recibieron la promoción a un nivel superior. En este sentido, la evaluación se refiere a cambios en la estructura dentro del SNI, y no puede

estimarse el impacto del SNI contra no tenerlo. Dado que la evaluación considera dos periodos de evaluación, no incluye efectos de largo plazo.

Para poder obtener una visión más completa de lo que significa el SNI, se hace una evaluación cualitativa. Se envían más de 800 invitaciones a contestar una encuesta a investigadores incluidos en la muestra final de este estudio. Se trata de analizar qué significa el SNI, su influencia en la producción académica, y el método que se usa para calificar a los investigadores para sus promociones. Esto permite entender de forma cualitativa los resultados encontrados en la parte cuantitativa.

Usando diferentes métodos econométricos, en general se encuentra que subir un nivel en el SNI impacta positivamente en la producción académica. El efecto para los diferentes productos varía, pero en general está entre 5-10% de una desviación estándar. Si bien este efecto no es sustancialmente grande, sí es parecido a lo encontrado en otro tipo de investigaciones de incentivos monetarios y aprovechamiento académico (Fryer 2016). Los resultados indican que en general el impacto es mayor en número de artículos que en artículos con factor de impacto. Es decir, no se aumenta tanto la producción de artículos con factor de impacto al subir un nivel en el SNI como en el impacto de número de artículos. Las áreas con el mayor impacto en número de artículos son Biología, Química y Ciencias de la Vida (área 2), Medicina y Ciencias de la Salud (área 3), Biotecnología y Ciencias Agropecuarias (área 6) e Ingeniería (área 7). Las áreas donde impactan más el número de libros son Humanidades y Ciencias de la Conducta (área 4). Es de llamar la atención que en la mayoría de las áreas (con excepción del área 5 de Ciencias Sociales) el impacto disminuye o se hace cero al considerar el factor de impacto de la revista donde se publica el artículo.

Los resultados cualitativos señalan que en general el SNI es motivo de reconocimiento. Lo cual a su vez desemboca a que sea un estímulo, sin considerar la parte monetaria, para pertenecer en el sistema. Sin embargo, también una parte considerable de los investigadores contactados menciona que el incentivo económico es muy importante. También se considera en general que las evaluaciones son justas, pero muchas veces los criterios no son claros. Muchos investigadores también mencionaron que dado el periodo tan corto entre evaluaciones no se fomenta investigación de largo plazo significativa o bien buscar publicar resultados en revistas de alto factor de impacto. Esto es consistente con la evaluación cuantitativa pues se observa claramente que la producción en revistas con factor de impacto no aumenta tanto como todos los artículos en cualquier revista. También se mencionó como aspecto negativo las diferencias en infraestructura y en tiempo en actividades administrativas entre instituciones como un factor fundamental a considerar.

Con base en los resultados encontrados en los métodos cuantitativos y cualitativos se pueden obtener potenciales avenidas de mejoras y retos para el SNI. Primero, resulta importante igualar el impacto de número de artículos con el de número de artículos en revistas con factor de impacto. Esto se podría lograr si se hicieran más salientes los lineamientos para la promoción dentro del SNI, así como considerar la idoneidad del periodo de vigencia de cada nivel SNI. Estos cambios podrían disminuir las quejas que se tienen de varios investigadores (72% de la muestra para el análisis cualitativo) sobre el enfoque en la cantidad de artículos y no en la calidad, así como las quejas de que existe “simulación” de los investigadores al incluirse como autores en trabajos que no contribuyeron.

Segundo, otra preocupación real para investigadores en edad avanzada es el tema del retiro y dejar obtener el ingreso del SNI. Para generaciones más jóvenes, y con el objetivo de evitar este problema nuevamente en el futuro, se podría diseñar un mecanismo para que un porcentaje del ingreso SNI sea depositado directa y automáticamente en la Afore del investigador, de forma opcional y sin compromisos laborales o legales para Conacyt. Finalmente, el posicionar la labor científica en la sociedad siempre será un reto importante. Sin embargo, para el 49% de los investigadores ser reconocido por parte del SNI es el mayor aliciente a la producción académica.

El orden de este trabajo es el siguiente. En la sección 2 se incluye una discusión sobre el estado del arte en evaluaciones de impacto en CTI. En la sección 3 se discuten las hipótesis a probar empíricamente. En la sección 4 se discuten los efectos potencialmente atribuibles al SNI. En la sección 5 discutimos un poco más las limitaciones a este tipo de evaluaciones. En la sección 6 se presentan los datos disponibles. En la sección 7 se explica la metodología cuantitativa y cualitativa. En la sección 8 se presentan los pasos para la construcción de la muestra final así como algunas estadísticas descriptivas. En la sección 9 se presenta un análisis de poder estadístico con esta muestra final. En la sección 10 se presentan los resultados principales del estudio para la metodología cuantitativa. En la sección 11 se hace lo mismo que la sección 10 pero para la metodología cualitativa. En la sección 12 se presentan avenidas de mejora para el programa usando los resultados de ambas metodologías. Finalmente, la sección 13 concluye.

2. Estado del arte en evaluaciones de impacto de intervenciones en Ciencia, Tecnología e Innovación (CTI)

Las políticas y/o programas en materia de Ciencia, Tecnología e Innovación (CTI) se han convertido en un tema de seguridad nacional para la mayoría de los países. Debido a que estas en su implementación están diseñadas para conseguir resultados como, por ejemplo, el desarrollo de un país, en materia de la investigación científica y tecnológica, así como de su gestión e innovación en función de la sociedad y de los seres humanos que conviven en ella (Ortiz et al., 2010; Ortiz, 2015; Reyes-Ruiz & Suriñach-Caralt, 2015; Hernández & Jiménez, 2017). Por ello, para los gobiernos, universidades, institutos científicos y líderes de la industria, la evaluación de programas que estimulan la producción científica, así como los impactos en investigación científica se han convertido en una prioridad debido a que erige una necesidad estratégica como vía de verificación de los efectos que tiene una política y/o programa de CTI.

Inmersos en este marco, en México, la evaluación y rendición de cuentas de la investigación científica y la educación superior tiene su base y sustento en el Sistema Nacional de Investigadores (SNI) del Consejo Nacional de Ciencia y Tecnología (CONACYT) (FCCyT, 2014; Sánchez, 2010). El cual emergió con una característica peculiar: dotar de ingresos adicionales (concebidos como estímulos) a los investigadores, previa evaluación de pares, con el objetivo de compensar la pérdida del poder adquisitivo de los años ochenta (Gil & Contreras, 2017). En concreto, había que *evaluar* para diferenciar las características propias de un investigador internacional con el fin de hacerles llegar recursos para que obtenga una capacidad de laborar a un nivel digno. Por tanto, el SNI inauguró dos formas de

proceder en el mejoramiento de la investigación científica: (1) proceder en la distribución de recursos; y (2) proceder en una evaluación en los rasgos propios del investigador.

Para estar en condiciones de abordar la temática de evaluación de impacto conviene, en primer lugar, exponer en forma sintética, la evolución del SNI y, en segundo lugar, los retos de crecimiento del programa. Su génesis del programa data a la crisis que vivió México en la década de los ochenta y sus efectos sobre el sistema de educación superior. En ese marco, se intentaba de revertir la caída de los ingresos de los académicos, reducir la deserción de los investigadores y evitar la fuga de cerebros. Con ello y en más de sus 30 años de creación se creó una cultura de la evaluación periódica del desempeño académico, se estimularon los estudios de posgrado y la producción científica de calidad, y favoreció la producción de artículos científicos y su difusión a nivel internacional (FCCyT & ACM, 2005; FCCyT, 2014). No obstante, todo lo anterior afronta, a la vez, un proceso continuo de mejoramiento, y en ello están los retos para alcanzar el crecimiento del SNI en México (Silva et al., 2016).

Ahora, entrando en materia sobre la evaluación de la investigación científica y su relación con el SNI, toma relevancia el concepto de la categoría del impacto, puesto que se distingue como una unidad importante dentro de los procesos evaluativos, porque su noción se asocia con los efectos que pueden producir la ciencia y la tecnología en determinado sistema social (Lozano et al., 2011). Dentro de este contexto, el impacto de la ciencia, tecnología e innovación puede ser evaluado comúnmente en la dimensión del conocimiento, es decir, el impacto en conocimiento. Y las técnicas habituales para medir este impacto son las

bibliométricas.¹ Las cuales se basan en citas recibidas por el trabajo de investigación en otros documentos. Asimismo, el conocido Factor de Impacto² de las revistas científicas o artículos se basa en estas técnicas. No obstante, la manera en cómo cuantificar esta medición del impacto es cuestionada por muchos autores, puesto que es muy difícil cuantificar el impacto integral de la investigación científica.

Debido a las constantes divergencias en opiniones sobre lo complicado que es cuantificar el impacto de las investigaciones científicas o lo superficial que resulta cuantificarlas, surge en el 2015 “The Leiden Manifesto for research metrics”, el cual propone ciertos criterios para evaluar de una mejor manera la investigación científica y contrarrestar este tipo de evaluación en términos del factor de impacto de las revistas o de medidas numéricas que dejan de lado la pluralidad de un ambiente cualitativo, como por ejemplo, el impacto de una investigación en el contexto nacional que no es apremiante a nivel internacional. Y uno de estos criterios es tender un puente o camino hacia la supresión del interés de los guardianes³ del alto impacto (revistas en inglés) y promover el pluralismo y la relevancia social de la investigación nacional. Por otra parte, existe también otro tipo de preocupaciones que también es recomendable tratarlas para contrarrestar la influencia en la medición de una evaluación a la investigación científica. En este

¹ Los indicadores bibliométricos permiten averiguar el impacto causado por un trabajo científico cualquiera a partir de la literatura científica y tecnológica publicada. Los indicadores bibliométricos permiten manejar, clasificar y analizar grandes volúmenes de publicaciones científicas (Abramo & D’Angelo, 2011)

² De acuerdo con Lozano et al. (2011), generalmente se considera que un artículo que haya sido publicado en una revista con elevado factor de impacto tiene por ello impacto científico.

³ El entendimiento de guardianes de alto impacto toma referencia en que la investigación científica y su respectiva medida métrica están sesgadas a las necesidades de países desarrollados. Por lo cual, dentro de los criterios del Manifiesto Leiden es tomar mayor relevancia la medición métrica de la investigación local de países en desarrollo. Todo ello conllevaría probablemente a que la investigación científica para dichos países sea más beneficiosa y sobre todo de mayor calidad.

sentido, Hicks et al. (2015) mencionan una preocupación por tomar como relevante a determinar el valor de la contribución de la generación del conocimiento que se produce a partir del trabajo de investigación, diferenciándolo por métricas que no alcanzan a expresar resultados significativos y que por tanto se han prestado a toda clase de simulaciones. Pero este es un camino que recién empieza su curso hacia la obtención de una mejor evaluación. Como se ha mencionado, una evaluación a una intervención en CTI es compleja porque no todo lo importante y relevante se puede medir (impactos sociales) así como el periodo de tiempo de evaluación se encuentra necesariamente restringido.

Tomando en cuenta lo anterior, la evaluación de impacto en CTI se realiza a través de diferentes indicadores bibliométricos para determinar la novedad y el aporte teórico de los nuevos conocimientos producidos por las investigaciones (Ortiz, 2015; Saavedra-Fernández et al., 2005). Lo cual refleja únicamente una dimensión de un proceso complejo como es el impacto en investigación científica de los estímulos del SNI. En este sentido, los resultados de un análisis de indicadores bibliométricos pueden explicar tópicos como, por ejemplo, su relación con el rendimiento y la productividad de los investigadores en términos cuantitativos, además, se puede ver su relación con la elección de qué tipo de factores (bibliométricos) son relevantes para cuantificar una investigación científica de alto impacto, y por supuesto, también su relación con la productividad académica diferenciado por género.

Con respecto a este primer tema de la medición del impacto de la investigación científica referenciada a la cuantificación (medidas bibliométricas) y su relación con el rendimiento y productividad académica de los investigadores,

destaca la investigación de Frixione et al. (2016), los cuales investigan la medida eficaz y eficiente a largo plazo para la evaluación de investigaciones de académicos en México. Los autores para llevar a cabo esta práctica aplican una encuesta limitada sobre la operación y el resultado del Sistema Nacional de Investigadores (SNI) de 30 años. En este proceso se selecciona una muestra estadística aleatoria de investigadores del área de Humanidades y Ciencias de la Conducta, dado que es una de las áreas más consolidadas del SNI y, asimismo, se aplica este procedimiento a los miembros de un Comité Dictaminador y de Revisión.⁴ De acuerdo con los autores, esta área es una de las más consolidadas debido a que es una de las primeras y comprende una amplia gama de disciplinas de investigación, desde filosofía hasta pedagogía, arqueología, y la investigación cerebral experimental. El procedimiento para tratar la base de datos consistió en realizar una comparación de rangos individuales o "Niveles de distinción" con el cumplimiento de los propios criterios de evaluación del SNI. Frixione et al. encontraron para ambos grupos una gran disparidad de la productividad de los investigadores dentro del mismo Nivel⁵ de distinción del SNI. Esta disparidad se puede deber a que evaluar la investigación científica es compleja, ya que mezcla conceptos cuantitativos y cualitativos, y necesariamente se tiene subjetividad. No obstante, este resultado no es generalizable debido a que no es comparable ante

⁴ El CONACyT establece en base al programa del SNI que un Comité Dictaminador y de Revisión sea integrado por investigadores de alto nivel pertenecientes a áreas generales de conocimientos, que son encargados de evaluar y clasificar a sus colegas.

⁵ Los niveles de distinción que confiere el SNI se clasifican en tres categorías, de acuerdo con los requisitos fundamentales que se establecen en el reglamento: (1) Candidato a Investigador Nacional; (2) Investigador Nacional, con tres niveles (Nivel I, Nivel II, y Nivel III); e (3) Investigador Nacional Emérito.

áreas y además porque presenta algunos posibles errores metodológicos, y uno de estos principalmente es que el total de observaciones es de solo 72 académicos examinados de un universo de más de 2500 registrados en el Área IV para el momento del estudio, es decir, alrededor del 2.9% del total. Esto último anterior, es importante debido a las controversias sobre la importancia estadística de las pruebas relacionadas con la evaluación de la investigación. No obstante, es importante añadir a este resultado controvertido por parte de Frixione et al. (2016), que evaluar el impacto de la investigación científica es complicada precisamente por combinar elementos cualitativos y cuantitativos, y que cada evaluador puede ponderar esos elementos de forma distinta.

En el sentido del contexto anterior, se pueden destacar distintos determinantes que cooperan con esta disparidad de la productividad de los investigadores y su distinción en el SNI. Como, por ejemplo, la edad y la relación entre colegas. En este contexto, González-Brambila & Veloso (2007) y Costas et al. (2010) encontraron que el determinante o factor edad juega un papel heterogéneo en los resultados de productividad de los investigadores. Es decir, que la edad no tiene un rol importante en términos de productividad de la investigación. Por su parte, Zinovyeva & Bagues (2014) encontraron que cuando los solicitantes son evaluados por sus colegas los sesgos de evaluación dominan el impacto potencial de la información. Por el contrario, cuando hay una relación débil (o distancia social) entre candidatos y sus evaluadores colegas se mejora la eficiencia del proceso de evaluación. Por otra parte, es relevante detallar, de acuerdo con los criterios de evaluación, que cuando existe una relación entre un miembro de alguna comisión y

un solicitante, el primero se excusa de evaluarlo y que si existiera este sesgo con una persona, existen otros 13 integrantes, todos ellos expertos, que lo eliminan.

Siguiendo la misma lógica anterior, los factores o determinantes bibliométricos son relevantes para evaluar la investigación científica. Sobre todo, porque sirven como herramientas rentables, principalmente porque dan información cuantitativa sobre la investigación y el impacto de esta. En este aspecto se han pronunciado autores como Abramo & D'Angelo (2011) en el ámbito de buscar parámetros (determinantes bibliométricos) esenciales para una implementación adecuada en la medición: precisión, robustez, validez, funcionalidad, tiempo y costos; Bollen et al. (2009), Didegah & Thelwall (2013) y Carpenter et al. (2014) sobre el uso de métricas – basadas en datos que incluyen medidas como el número de publicaciones, el número de citas, el puntaje del factor de impacto de la revista y así como las métricas emergentes basadas en un nivel micro del documento – para la productividad académica y su impacto utilizando datos de publicación y Seglen (1998), Pontille & Torny (2010), Eyre-Walker & Stoletzki (2013) y Wang et al. (2013) acerca de los métodos para evaluar el mérito de un artículo científico como, por ejemplo, la revisión por pares tiene subjetividad posterior a la publicación, al número de citas obtenidas por un artículo y el factor de impacto de la revista en la que se publicó el artículo.

Finalmente, el último aspecto a tratar es el tema de diferencias de género y su implicación en la caracterización de los programas de CTI. En esta dirección, Bukstein & Gandelman (2016) realizan una investigación sumamente interesante y retadora para un país como Uruguay. Los autores combinan dos enfoques: las diferencias de género en la economía laboral y la caracterización de programas

públicos en materia de Ciencia y Tecnología. Los investigadores utilizan métodos econométricos de descomposición y relacionan cuestiones de género con las promociones y jerarquizaciones en el campo de la ciencia y la tecnología. Bukstein & Gandelman encuentran evidencia de que las mujeres investigadoras tienen 7.1 puntos porcentuales menos de probabilidades de ser aceptadas en el mayor programa nacional de apoyo a la investigación en Uruguay, el Sistema Nacional de Investigadores. También tienen una menor productividad de investigación que sus homólogos masculinos y que la brecha de género es mayor en los rangos superiores del SNI. Por tanto, esta falta de mujeres en la posición más alta de investigación puede afectar a mediano plazo al grupo de investigadores al afectar las decisiones que los jóvenes están tomando hoy sobre sus estudios y su carrera profesional.

En conclusión, el valor agregado sobre la evidencia empírica y teórica de la evaluación de impacto en investigaciones científicas es que constituye uno de los principales canales para la aplicación de diferentes indicadores para determinar el grado de concordancia entre lo previsto por los investigadores y el aporte logrado (Ortiz, 2015). Sin embargo, hay que considerar que necesariamente esas evaluaciones son limitadas por la información incluida en los estudios. Resulta imposible tener todos los datos deseados con la periodicidad idónea para poder tener una evaluación de impacto integral. En esta forma, la mayoría de estudios limita el alcance de la investigación a aspectos medibles y verificables, con la salvedad que no son todos los indicadores que nos gustaría tener para saber el impacto integral de una política en CTI.

3. Hipótesis de impacto a probar con la evaluación

El objetivo que persigue esta investigación es evaluar qué tan importante es el Sistema Nacional de Investigadores – con relación al desempeño de incentivos a la investigación – para la productividad académica posterior del investigador. Con ello es posible identificar y evaluar el impacto de subir un nivel en el SNI en comparación con no subir, restringido a los investigadores pertenecientes al sistema. Es decir, no es posible estimar el impacto de investigadores dentro del sistema contra los que no están.

Se define productividad académica como el número de artículos, capítulos de libro, libros, y número de artículos ponderados por el factor de impacto. Esta definición es limitada. No se observan citas a cada artículo o capítulo al momento de la evaluación, ni tampoco se observa o es posible cuantificar el impacto social de cada investigación. Sin embargo, al menos ofrecen medidas estables y que pueden compararse entre grupos de investigadores en un periodo de tiempo determinado (solo corto plazo).

En este trabajo de investigación la hipótesis a probar es analizar si una promoción o incentivo (ascenso del Nivel de distinción) al investigador por parte del SNI fomentará una mayor investigación científica posterior al incentivo. Esto se realiza con métodos cuantitativos, pero también se usan métodos cualitativos para dar luz sobre el significado de las estimaciones. El trabajo se implementa con miras a servir a CONACYT para que pueda mejorar el programa, así como de apoyo a evaluadores de las siete Comisiones Evaluadoras del Sistema Nacional de Investigadores.

Para probar empíricamente esta hipótesis se utilizan dos métodos econométricos: regresión múltiple y método de aparejamiento con diferencias en diferencias. Es decir, se puede apreciar el potencial productivo en materia de investigación posterior a la promoción asignada al investigador usando métodos econométricos.

4. Efectos potencialmente atribuibles al SNI e indicadores asociados

El Programa S191–Sistema Nacional de Investigadores (SNI) es un Programa de Fomento a la Investigación Científica, establecido por el Gobierno Federal, cuya conducción y operación, así como el establecimiento de sus objetivos y funciones, organización, y reglamentación interna están a cargo del Consejo Nacional de Ciencia y Tecnología (CONACYT). En consecuencia, el SNI, de acuerdo con su reglamento, tiene por objeto reconocer, a través de los resultados de una evaluación, la calidad de la investigación científica, tecnológica e innovativa que se produce en México o por mexicanos en el extranjero (CONACYT, 2017).

De acuerdo con el reglamento interno del SNI, y conociendo la diversidad de disciplinas científicas y estilos de investigación este Programa S191 se divide en siete diferentes áreas generales de conocimiento: (I) Físico-Matemáticas y Ciencias de la Tierra; (II) Biología, Química y Ciencias de la Vida; (III) Medicina y Ciencias de la Salud; (IV) Humanidades y Ciencias de la Conducta; (V) Ciencias Sociales; (VI) Biotecnología y Ciencias Agropecuarias; e (VII) Ingenierías. Asimismo, estas áreas del conocimiento le corresponden a cada uno sus propios Comités Dictaminadores y Revisores; los cuales tienen por objeto el diseño y revisión de los criterios específicos para evaluar el desempeño de sus colegas en las respectivas áreas de

conocimiento, de acuerdo con sus propios métodos de trabajo y tradiciones (CONACYT, 2017). Con base en estos criterios específicos, los Comités Dictaminadores y Revisores debaten y evalúan las solicitudes individuales recibidas para cada período de evaluación. Los solicitantes son designados como investigadores nacionales en una de las tres categorías y tres Niveles de distinción (Frixione et al., 2016). A través de este proceso, un poco más de 28,634⁶ académicos actualmente inscritos en SNI son evaluados periódicamente, con intervalos de revisión obligatorios establecidos de acuerdo con su Nivel de SNI y la fecha de su nombramiento (Nivel de distinción) más reciente.

Al ser admitidos los investigadores a cada uno de estos Niveles de distinción se hacen acreedores de un estímulo económico libre de impuestos por parte del SNI. Adicionalmente, tienen la oportunidad que sus investigaciones sean financiadas por CONACYT y otras agencias de apoyo a la CTI, además del reconocimiento no pecuniario otorgado por la sociedad por ser miembro del SNI.

Por tanto, se requiere analizar qué efectos, si es que los tiene, genera el tener una distinción SNI. Dada la información disponible, solo es posible analizar la comparación de una distinción mayor a una menor, y no tener la distinción de estar en el SNI y no estar en el sistema.

El indicador principal del estudio es la productividad académica del investigador medida en el periodo posterior a la evaluación para la obtención del Nivel de distinción. Productividad académica se define como número de artículos, artículos ponderados por factor de impacto de revista, capítulos de libros y libros.

⁶ Cifra del año 2018, con datos propios otorgados por el SNI del sistema PeopleSoft.

En este aspecto, esta variable es fundamental puesto que un valor mayor de esa variable conlleva un nivel mayor en el sistema. Es decir, que para cada Nivel de distinción un requisito primordial es mantenerse en la vanguardia de la investigación científica, que conlleva a aumentar las publicaciones en revistas de calidad o, en dado caso, en revistas arbitradas de calidad internacional, indexadas en el JCR (Journal of Citation Reports) u otro tipo clasificación con calidad internacional (Frixione et al., 2016). Esa variable de productividad académica es relevante porque se usa la información que observan también los comités de evaluación para decidir una promoción o permanencia en el mismo nivel.

5. Limitaciones de una evaluación en el área de Ciencia, Tecnología e Innovación

Como se ha mencionado, la política pública en Ciencia, Tecnología e Innovación es compleja de evaluar. Esto simplemente porque no es posible incluir todos los indicadores posibles de forma cuantitativa. En este trabajo solo se evalúa cómo la estructura del SNI, y en particular la promoción del nivel 1 al 2 o bien del 2 al 3, fomenta la investigación científica. Esta se mide únicamente con número de artículos de revista, capítulos de libros y libros. Por lo tanto, el alcance de la investigación es limitado.

El Manifiesto de Leiden hace énfasis en que los indicadores para medir cuantitativamente la investigación científica han proliferado (Hicks et al. 2015). Y esta proliferación conlleva en muchos casos a que los resultados sean bien intencionados Aunque, también sucede que no siempre son bien informados, y usualmente tienden a ser mal aplicados en el análisis. Todo ello sucede por el

simple hecho de que la evaluación cuantitativa no se relaciona con la valoración cualitativa de expertos. Es decir, no todos los indicadores pueden señalar el impacto social o valoración completa de la investigación científica.

En este contexto, cuando no se tiene una relación entre un análisis cualitativo y cuantitativo de la medición de la investigación científica, resulta complicado medir el impacto social de la misma. Este problema se puede agravar si solo se utiliza el factor de impacto de una revista. Esto se debe a que el factor de impacto (FI) de un documento de investigación no es garantía de que un artículo se cite o de la relevancia para la sociedad. En este sentido, se puede agregar que no es posible medir el impacto social debido a que es complicado cuantificar los impactos en el largo plazo. Por tanto, si bien se reconoce que la medida de productividad académica usada en este estudio es valiosa e importante y refleja cómo se comportan los investigadores en esas medidas, no permite estimar el impacto económico o social del SNI.

Otra limitación importante del estudio es su enfoque en el corto plazo. No es posible estimar impactos de la investigación científica acumulados en el largo plazo, principalmente debido a la falta de un grupo de comparación adecuado. Tampoco es posible estimar impactos para investigadores fuera del SNI, simplemente porque no se observa su producción académica.

Esta variable contiene un subconjunto de la información contenida en la evaluación de pares, de acuerdo con los criterios de evaluación respectivos a las siete áreas de conocimiento del Programa presupuestario S191. Por tanto, se usa información que los evaluadores académicos toman en cuenta para decidir sobre una promoción o permanencia de nivel. Y esa decisión de promoción o

permanencia tiene consecuencias futuras sobre el comportamiento de los investigadores en el corto plazo acerca de su producción académica.

En resumen, no es posible evaluar el SNI integralmente en todas las variables deseadas como impacto social o impacto en producción académica en el largo plazo. Pero aun así esta evaluación es importante y valiosa.

6. Fuentes de información disponibles

Se tiene datos de los investigadores del SNI para los años 2005-2018,⁷ las solicitudes presentadas al SNI, así como sus dictámenes (únicamente decisión final del comité de qué nivel es asignado el investigador, no contamos con información adicional como argumentos a favor o en contra de la promoción del nivel de cada candidato), y la información histórica proveniente del Sistema de Currículum Vitae Único (CVU). De acuerdo con González-Brambila & Veloso (2016) es importante destacar que los investigadores pertenecientes al SNI no están distribuidos al azar y no representan a toda la población de investigadores en México. Por el contrario, dadas las características del SNI, se espera que los investigadores más productivos estén en el sistema. Por lo tanto, el efecto encontrado no puede generalizarse a toda la población de investigadores en México.

Los datos se clasifican en las siguientes categorías:

1) Variables de control (con información anterior a la evaluación)

- Género

⁷ Para ser incluidos en el SNI, los investigadores aplican al Programa Presupuestario S191, luego son evaluados por comités mediante un mecanismo de evaluación por pares que determinan qué investigadores pueden recibir los beneficios del programa y el nivel apropiado (González—Brambila & Veloso, 2016).

- Edad
- Años entre evaluaciones.
- Área y disciplina de conocimiento⁸
- Artículos, libros, capítulos de libros.
- Factor de impacto de publicaciones – ISSN

2) Variable de tratamiento:

- Nivel inicial de distinción de los investigadores (ya sea Nivel 1 o 2)
- Nivel final de distinción después de la evaluación (Nivel 1, 2 o 3).
- Grupo de tratamiento se refiere a investigadores que subieron de nivel en su evaluación.
- Grupo de control se refiere a investigadores que permanecieron en el mismo nivel.

3) Variables dependientes o de resultados para el periodo post-evaluación:

- Número total de publicaciones por investigador. Estas pueden ser número de artículos, número de artículos ponderados por factor de impacto de revista, número de capítulos de libros, número de libros.

⁸ En este trabajo se sigue la clasificación por área de conocimiento dada por el Reglamento del SNI, el cual ya se describió anteriormente. Por otro lado, se entiende como disciplina del conocimiento a un campo del área del conocimiento; por ejemplo, el área de Físico-Matemáticas y Ciencias de la Tierra incluye Matemáticas, Física, Astronomía, Ciencias de los Materiales, etc.

7. Metodología

Dada la complejidad de una evaluación al sector de CTI, resulta imprescindible implementar una evaluación de impacto tanto cuantitativa como cualitativa. Para la evaluación cuantitativa, se utilizan dos tipos de herramientas de análisis de datos: (a) método de regresión múltiple; (b) método de diferencias en diferencias combinado con apareamiento por puntajes. Se consideran estos algoritmos de datos porque permiten tener en cuenta cualquier diferencia constante en el tiempo entre los grupos de tratamiento y de comparación; además, permite detectar el nivel de asociación de las variables involucradas. Asimismo, estos métodos se han consolidado por sus eficientes resultados en cuanto a la evaluación de impacto y clasificación de datos (Reyes-Ruiz, 2018; Reyes-Ruiz & Suriñach-Caralt, 2012). Lo que estos métodos realizan es tratar de comparar investigadores similares para el periodo pre-evaluación, algunos investigadores obtendrán una promoción para un nivel superior en el SNI, mientras que otros permanecerán en el mismo nivel. Entonces dado que la producción académica es relativamente similar entre ambos grupos, es posible calcular el efecto de la promoción en la productividad posterior.

No obstante, y como se mencionó anteriormente, estos métodos solo utilizan variables cuantitativas. Dado el Manifiesto Leiden se debe tener claro que la parte cualitativa importa. Además de que la evaluación cualitativa complementa a la cuantitativa para transparentar y dar sentido a los resultados encontrados.⁹

⁹ Dos motivos muy importantes existen para la complementariedad del análisis cualitativo: (1) retroalimentación del proceso de evaluación del SNI y (2) retroalimentación del proceso de empuje o estímulo para el trabajo científico.

a. Metodología cuantitativa

Se implementan dos métodos: a) regresión múltiple y b) diferencias en diferencias con apareamiento por puntajes. El método de regresión múltiple, utilizado para esta investigación, opera con dos tipos de variables explicativas – el primer tipo de variable explicativa está en términos de la variable dependiente antes de la evaluación y el segundo tipo son variables de control. La relación funcional¹⁰ se define como:

$$Y_{i;t+j} = \alpha + \gamma T_i + \pi Y_{i;t} + \vartheta_K X_{i;t_K} + \varepsilon_{i;t+j} \quad (1)$$

donde:

$Y_{i;t+j}$: número de publicaciones por investigador posterior a la obtención del Nivel de distinción deseado. Pueden ser número de artículos, número de artículos ponderador por factor de impacto, capítulos de libro, y libros.

T_i : variable dicotómica indicando si el investigador i recibió una promoción de nivel de distinción (0 si permaneció en el mismo nivel). Solo se analizan aquellos con distinción inicial 1 o 2 y puede subir un nivel de 1-2 o 2-3.

$Y_{i;t}$: número de publicaciones por investigador antes de la evaluación para la obtención del Nivel de distinción deseado.

$X_{i;t_K}$: variables explicativas: área de conocimiento, años entre evaluaciones, género, año de evaluación, edad.

¹⁰ La ecuación detrás de la forma funcional de regresión, para este primer método, está dada por una simple diferencia (γ) dentro del grupo de tratamiento entre los resultados después y antes de la evaluación: $E[Y_{i;t+j}/X; Y; T = 1] - E[Y_{i;t+j}/X; Y; T = 0] = \gamma$.

$\varepsilon_{i; t+j}$: variable aleatoria, asociada al error de la estimación.

$i : 1, 2, 3, \dots, n$ – es el total de investigadores pertenecientes al SNI.

$K : 1, 2, 3, \dots, m$ – es el número de variables explicativas.

$j : 1, 2, 3, \dots, J$ – es el número en años posterior al periodo de evaluación para la obtención por parte del investigador del Nivel de distinción deseado.

Es claro que la variable dependiente $Y_{i; t+j}$ es el número de publicaciones por investigador con respecto al Nivel de distinción dado una solicitud aprobada por las Comisiones Evaluadoras del SNI, para el periodo comprendido después de la evaluación. Esta variable depende del periodo de evaluación siguiente ($t + j$) y de lo que se produjo académicamente en ese periodo. Si las variables de control son las adecuadas, y se compara investigadores relativamente similares, entonces el coeficiente de interés es γ . Este coeficiente se interpreta como la diferencia en productividad posterior de tener la promoción relativo a permanecer en el mismo nivel. Se hace una diferencia en este método porque se asume que la diferencia en el periodo pre-tratamiento es igual a cero, por lo que solo se requiere una diferencia. En la práctica se estiman cuatro modelos que cambian el número de variables explicativas o bien se estima un modelo tobit para considerar que la variable dependiente está censurada en 0.

El siguiente método se refiere a diferencias en diferencias con aparejamiento por puntajes. De acuerdo con Gertler et al. (2017), cuando se tienen datos de línea

de base¹¹ sobre los resultados, se puede combinar el apareamiento para reducir el riesgo de sesgo en la estimación. El primer paso se refiere a estimar un puntaje. Este puntaje se puede estimar con un modelo logit, componentes principales (discriminante) o bien una razón de momios. Lo importante es que este puntaje incluya las variables más relevantes para predecir el tratamiento. En este sentido, este método opera bajo el supuesto que cualquier característica no observable entre los dos grupos sea constante a lo largo del tiempo (Gertler et al., 2017).

Es decir, primero se calcula una probabilidad de ser tratado ($P(T = 1/X)$) usando las variables de control especificadas anteriormente. La forma funcional de esa estimación puede cambiar. Como segundo paso se calculan diferencias en productividad antes y después del tratamiento para obtener un efecto de diferencias en diferencias. La ventaja del método es que permite comparar investigadores con la misma probabilidad de ser promovidos en el SNI, pero unos en realidad sí fueron promovidos y otros permanecieron en el mismo nivel.

La gran desventaja de este método es que solo usa información de investigadores muy similares entre sí. Por ejemplo, se compara únicamente investigadores dentro del área 1 entre aquellos que iniciaron en nivel 1 y tardaron 3 años de evaluación, y así para todas las áreas, niveles iniciales (1 o 2), y años entre evaluaciones (3-4 para el nivel 1 y 4-5 para el nivel 2). Entonces se pueden perder muchas observaciones de investigadores tanto en el grupo de tratamiento como de control. Esto lleva a problemas de tratar de generalizar el estimador a toda la

¹¹ Se entiende como datos de línea de base a una estructura de datos que se recopilan antes de implementar el programa o evaluación, tanto en el grupo de tratamiento como en el grupo de control (Gertler et al., 2017).

población, pero tal vez lo más problemático es que se llega a falta de poder estadístico. Al perderse muestra se corre el riesgo que puede estar estimado el efecto correctamente, pero no se tendrá poder estadístico para diferenciarlo de un efecto 0.

En la práctica se presentan estimaciones con diferentes modelos para predecir el tratamiento. Se usa un modelo logit, componentes principales, y razón de momios. Luego también se varía el método de apareamiento. En este trabajo se usan vecinos cercanos restringidos a una vecindad.

b. Metodología cualitativa

Es importante que la evaluación cuantitativa se complemente con los resultados obtenidos con el análisis cualitativo a obtener por parte de pequeñas encuestas a investigadores pertenecientes del SNI. El combinar la metodología cuantitativa con la cualitativa es un tema de frontera y cada vez más usado por investigadores (Honig 2018; Mele y Belardinelli 2018). Como menciona Honig (2018), el análisis cualitativo en métodos mixtos se usa para “proveer más interrogaciones precisas de enunciados realizados en la parte cuantitativa” (p. 2). Para esto se diseñó un breve cuestionario que fue enviado vía correo electrónico a investigadores seleccionados de manera aleatoria en las diferentes áreas del SNI. Se determinó una muestra de aproximadamente 800 investigadores (se calculó una tasa de respuesta del 10-15%) para que sea representativa de las diferentes áreas de conocimiento y niveles de distinción. Esta muestra de investigadores se encuentra en la muestra cuantitativa, de tal forma que se puede relacionar el área de conocimiento y si la persona está en el grupo de tratamiento o control.

El objetivo de la evaluación cualitativa es conocer de primera mano de los investigadores lo que significa el SNI, su influencia, y la percepción de la evaluación para obtener diferentes niveles en el SNI. Se envía un correo electrónico a cada uno de los investigadores y se les solicita que por el mismo medio respondan a las siguientes preguntas:

1. ¿Qué significa el Sistema Nacional de Investigadores para usted?
2. ¿Cómo fue su experiencia cuando fue evaluado por el SNI para su reciente nombramiento?
3. ¿Cómo siente que el SNI ha influenciado su productividad científica?
4. ¿Considera que la forma en como incentiva el SNI la investigación científica es la adecuada? ¿Sí o No? ¿Por qué?

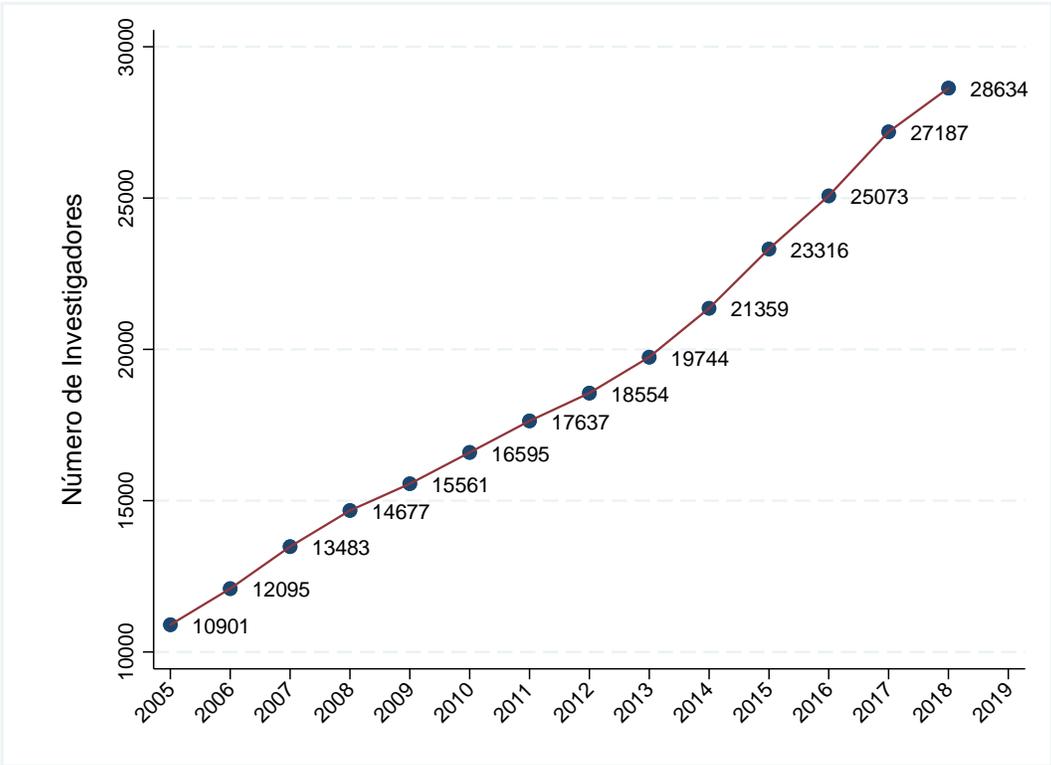
Una vez obtenidas las respuestas se procesan para analizarlas. Por un lado, se analizan las frecuencias de palabras más repetidas en las respuestas de cada pregunta. Por otro, se revisan todas las respuestas y se organizan en sentimientos los diferentes tipos de respuestas. Esta metodología de analizar el texto en forma cuantitativa es cada vez más utilizada en las ciencias sociales y humanidades: en economía (Gentzkow et al. 2017), en ciencia política (Wilkerson y Casas 2017), y en las humanidades (Jänicke et al. 2016). Finalmente, se presentan respuestas típicas o relevantes que den sentido a los resultados encontrados con la metodología cuantitativa.

8. Construcción de base de datos final

El primer paso para la construcción de base de datos final es determinar el número de investigadores y quiénes son evaluados. De acuerdo con los datos del

padrón de 2005-2018 se obtiene el número de investigadores vigentes para cada año. La Figura 1 muestra esta evolución y señala el crecimiento que ha tenido el SNI desde 2005. Desde ese año hasta el 2018 se tiene un aumento de aproximadamente 17,700 investigadores, lo que representa un aumento cada año de aproximadamente 1,267 investigadores. Es decir, los investigadores vigentes prácticamente se han triplicado en el periodo de estudio.

Figura 1. Número de investigadores por año



Nota: Cálculos por el autor.

Con los datos disponibles también se puede representar cómo se distribuyen los investigadores en distintos niveles. La Tabla 1 muestra cómo se dividen los investigadores por distinción para algunos años. El nivel 3 tiene menos del 10% de

todos los investigadores, mientras que el nivel 1 es el que tiene más del 50% de todos los investigadores.

Tabla 1. Investigadores por distinción. Años 2005, 2010, 2015 y 2018

Nivel	2005	2010	2015	2018
C	1876	3046	4574	6549
1	5980	8970	12775	15144
2	2076	3172	3964	4573
3	942	1372	2002	2368
Total	10901	16595	23316	28634

Nota: Cálculos por el autor. Eméritos incluidos en el total.

De la misma forma en la Tabla 2 se muestra el dictamen de la evaluación de cada investigador. En la última columna se muestra que del stock de investigadores vigentes, algunos no son evaluados en ese año calendario porque simplemente no les toca ser evaluados. Ellos representan alrededor del 70% de los investigadores vigentes de ese año. Algunos investigadores evaluados son de recién ingreso y otros investigadores son removidos. No se puede usar esos investigadores para la evaluación porque no se puede construir un grupo de control apropiado. La muestra que nos interesa son los investigadores que son evaluados. En esta muestra (las 3 columnas de la tabla) se observa que la mayoría subieron de nivel o no cambiaron de nivel. Son relativamente pocos los investigadores que bajaron de nivel. Por tanto, no es posible usar la muestra de investigadores que bajaron de nivel al ser muy pequeña. Se restringe la muestra final a aquellos investigadores que se mantuvieron en el mismo nivel (grupo de control) o bien que subieron de nivel (grupo de tratamiento).

Tabla 2. Investigadores evaluados y decisión de evaluación.

Periodo	Invest. evaluados	Invest.	Invest.	Invest.
---------	-------------------	---------	---------	---------

	vigencia establecida			evaluados recién ingreso	evaluados removidos	no evaluados
	Subieron de nivel	Bajaron de nivel	No cambiaron de nivel			
2005	894	33	2174	1802	608	7192
2006	418	1	981	1830	442	10253
2007	782	21	1468	1867	673	10539
2008	883	49	2356	1858	974	10415
2009	907	52	2804	2083	1049	10749
2010	635	47	2256	2017	975	12682
2011	743	56	2378	1996	1079	13381
2012	1057	102	2826	2424	1234	13335
2013	1232	66	3102	2788	1173	14171
2014	1078	73	2951	3008	1051	16206
2015	1084	51	3250	3152	1395	17536
2016	1238	74	3759	3572	1458	18544
2017	1262	115	3780	3270	1823	20207

Nota: Cálculos por el autor.

Adicionalmente, se requiere restringir el periodo que tuvieron entre evaluaciones y que tengan al menos dos evaluaciones. Se restringe a que sea lo especificado en los lineamientos del programa, y por tanto solo se utilizan a investigadores que tienen el Nivel 1 y tienen un tiempo de evaluación de 3 o 4 años, o bien tienen el Nivel 2 y tienen un tiempo de evaluación de 4 o 5 años. Se requieren dos evaluaciones para poder calcular la producción académica antes y después de dicha evaluación. Asimismo, dado que las variables de control y de resultados dependen de la producción académica, si un investigador no tiene producción académica (conjuntamente en artículos, capítulos de libros o libros) no se toma en cuenta para esta evaluación. Por tanto, la muestra final quedaría como se señala en la Tabla 3.

Tabla 3. Muestra final y división por nivel antes y después de la evaluación, así como por tiempo entre evaluaciones.

Duración Años	Nivel de distinción antes y después de evaluación			
	1-1	1-2	2-2	2-3
3	5,545	606	0	0
4	5,134	2,340	1,975	366
5	0	0	857	479
Tratamiento		2946		845
Control	10679		2832	

Notas: Cálculos por el autor.

La tabla indica que se tienen más investigadores en el grupo de control que en el tratamiento, así como más investigadores con nivel inicial 1 que 2. Esto es congruente con las tablas 1 y 2. En total se tienen 17,302 observaciones en la muestra final, de las cuales 10,715 son investigadores únicos. Esto se debe a que varios investigadores pueden tener 3 o 4 evaluaciones en el periodo referido.

Las estadísticas descriptivas de la muestra final se incluyen en la Tabla 4. El porcentaje de mujeres evaluadas que están en la muestra final es de 35% con una edad promedio de 49.5 años. La gran mayoría de las observaciones son para aquellos que empezaron en el Nivel 1 del SNI, y en menor medida para los que empiezan con un nivel 2. Nuestros años de evaluación van de 2005 a 2015, y se tiene que en general la muestra está dividida equitativamente por año de evaluación. En términos de áreas de conocimiento, se tiene que las áreas 1 y 2 tienen 16 y 18% respectivamente, el área 3 tiene 11%, el área 4 y 5 tienen porcentajes similares a la 1 y 2, el área 6 tiene 9% y el área 7 tiene el 12%. El número de artículos en promedio está entre 7-9, en menor medida los artículos

ponderados por factor de impacto se tienen entre 6-8 capítulos de libro por periodo, así como aproximadamente 4 libros.

Tabla 4. Estadísticas descriptivas en la muestra final.

Variables	N	Media	Desv. Est.
Mujer	17302	0.35	0.48
Edad	17302	49.47	8.39
Empezó en Nivel 1, 3 años de evaluación	17302	0.36	0.48
Empezó en Nivel 1, 4 años de evaluación	17302	0.43	0.50
Empezó en Nivel 2, 4 años de evaluación	17302	0.14	0.34
Empezó en Nivel 2, 5 años de evaluación	17302	0.08	0.27
Evaluación Año 2005	17302	0.10	0.30
Evaluación Año 2006	17302	0.03	0.16
Evaluación Año 2007	17302	0.07	0.26
Evaluación Año 2008	17302	0.11	0.31
Evaluación Año 2009	17302	0.12	0.33
Evaluación Año 2010	17302	0.09	0.28
Evaluación Año 2011	17302	0.10	0.30
Evaluación Año 2012	17302	0.13	0.33
Evaluación Año 2013	17302	0.14	0.35
Evaluación Año 2014 y 2015	17302	0.11	0.31
Área 1: Físico-Matemáticas	17302	0.16	0.37
Área 2: Química, Biología	17302	0.18	0.39
Área 3: Ciencias de la Salud	17302	0.11	0.31
Área 4: Humanidades	17302	0.16	0.36
Área 5: Ciencias Sociales	17302	0.18	0.38
Área 6: Biotecnología	17302	0.09	0.29
Área 7: Ingeniería	17302	0.12	0.32
# Artículos pre-evaluación	17302	6.87	8.77
# Artículos post-evaluación	17302	9.04	10.71
# Artículos ponderados por factor de impacto pre-evaluación	17302	2.95	10.00
# Artículos ponderados por factor de impacto post-evaluación	17302	4.90	17.06
# Capítulos de libro pre-evaluación	17302	6.34	13.47
# Capítulos de libro post-evaluación	17302	8.43	17.80
# Libros pre-evaluación	17302	3.89	9.57
# Libros post-evaluación	17302	4.76	10.51

Notas: Cálculos por el autor.

La Tabla 5 muestra una diferencia de medias simple entre los grupos de tratamientos y control para las variables principales del estudio. La tabla muestra que en general los dos grupos son diferentes entre sí. El grupo de tratamiento incluye menos mujeres que el grupo de control. Este punto requiere más estudio y profundidad de lo que se puede discutir en la presente investigación. Es importante que más mujeres puedan tener acceso a niveles más altos del SNI, pero la tabla señala que hay menos mujeres en términos relativos que logran subir que las que se quedan en el mismo nivel. Por otro lado, la edad es muy parecida entre 49-50 años, aunque sí es estadísticamente significativa esa diferencia como en el resto de las variables. Un aspecto a preocupar es que en general los grupos en términos de producción académica, como son artículos, artículos ponderados por factor de impacto, capítulos de libros y libros, son distintos. El grupo de tratamiento tiene una mayor producción académica en todos los casos. Esto no es de sorprender porque precisamente obtienen un nivel superior en el SNI por su producción académica. Sin embargo, resalta la importancia de utilizar métodos estadísticos para controlar por dichas diferencias.

Tabla 5. Diferencia de medias entre grupos de tratamiento y control.

VARIABLES	Control	Tratamiento	Diferencia	P-valor
Mujer	0.37	0.32	0.05	0.00
Edad	49.37	49.97	-0.60	0.00
# Artículos pre-evaluación	3.33	5.51	-2.18	0.00
# Artículos post-evaluación	5.27	7.85	-2.58	0.00
# Artículos ponderados por factor de impacto pre-evaluación	2.71	5.67	-2.96	0.00
# Artículos ponderados por factor de impacto post-evaluación	4.66	8.88	-4.22	0.00
# Capítulos de libro pre-evaluación	6.93	8.19	-1.27	0.00
# Capítulos de libro post-evaluación	9.16	10.92	-1.76	0.00

# Libros pre-evaluación	4.34	4.96	-0.62	0.00
# Libros post-evaluación	5.19	6.45	-1.26	0.00

Notas: Cálculos por el autor. Diferencia de medias con diferentes varianzas.

Antes de pasar a la siguiente sección, vale la pena reflexionar sobre la magnitud de lo encontrado en las tablas anteriores. Los investigadores están produciendo 6-9 artículos en cada periodo de evaluación, pero el número se reduce a la mitad cuando es ponderado por el factor de impacto. De la misma forma, se producen entre 7-10 capítulos de libro y 4-6 libros en cada periodo, en promedio. Sin embargo, la realidad para la mayoría de los investigadores es distinta. Si hay muchos investigadores con mucha producción se podría afectar el promedio. Cuando se analiza la mediana (en lugar de la media) se tiene que el número de artículos producidos entre periodos es de 5-6, pero con factor de impacto es 0-1, y capítulos de libros es 2-4, mientras que libros es 0. Es decir, la gran mayoría de los investigadores en el SNI, por la naturaleza de las disciplinas que conforman las áreas, no producen libros, y la producción se concentra en algunos investigadores, principalmente en área de humanidades. La gran mayoría de artículos producidos se publican en revistas que no tienen factor de impacto¹² (el 78% en artículos del periodo pre-evaluación y el 73% en artículos post-evaluación). Entonces si el investigador típico (en la mediana) produce cada periodo 5 artículos de investigación, ninguno o si acaso 1 con factor de impacto, y 3-4 capítulos de libro, el CONACYT podría enfatizar ciertos criterios para que se favorezca más la producción en revistas con factor de impacto en todas las áreas de conocimientos,

¹² List of latest Impact Factors released Selected JCR Year: 2017 Selected Editions: SCIE, SSCI Selected Category Scheme: WoS.

ya que el conocimiento que se tiene por los criterios de evaluación es que lo anterior se requiere solamente para las áreas 1, 2, 3, 6 y 7.

9. Análisis de poder estadístico

Un uso importante de las pruebas de significancia estadística en la investigación científica es probar hipótesis de impacto en experimentos controlados. Un componente importante de tal prueba es la noción de poder estadístico, que se define como la probabilidad de que una prueba estadística rechace correctamente la hipótesis nula o, en otras palabras, la probabilidad que detectemos un efecto cuando sí existe un efecto medible de nuestra intervención. Una prueba sin poder estadístico suficiente no podrá proporcionar al investigador suficiente información para sacar conclusiones con respecto a la aceptación o el rechazo de la hipótesis nula (Dybbå et al., 2006).

En este trabajo, se requiere calcular el poder estadístico para conocer la magnitud del efecto que es necesario encontrar para que se pueda rechazar la hipótesis nula de acuerdo con la construcción de la muestra final mostrada en la sección anterior.

La hipótesis nula es que el SNI no tiene ningún efecto sobre la productividad académica del investigador después del nombramiento o promoción, utilizando un nivel de significancia de 0.05. Las observaciones, para la realización del poder estadístico y para su posterior utilización en los métodos econométricos, se refieren a la muestra final obtenida con las restricciones mencionadas en la sección anterior: 17,302 observaciones con una clasificación de 13,510 observaciones para el grupo de control y 3,791 para el grupo de tratamiento.

Tabla 6. Análisis de poder estadístico. Diferencia de medias (δ) asumiendo diferentes varianzas y poder estadístico.

Poder Estadístico	# Obs	# Obs Control	# Obs Tratamiento	Efecto Estimado	Desv. Estándar en Control	Desv. Estándar en Tratamiento
80%	17302	13510	3791	0.051	1	1
80%	17302	13510	3791	0.058	1.5	1
90%	17302	13510	3791	0.067	1.5	1
80%	17302	13510	3791	0.072	1	1.5
90%	17302	13510	3791	0.06	1	1
90%	17302	13510	3791	0.084	1	1.5

Notas: Prueba t de Satterthwaite, suponiendo varianzas desiguales. Se fija un nivel de significancia estadística de 5%.

La Tabla 6 muestra los resultados del ejercicio fijando un nivel de significancia estadística al 5%. Se calculan efectos estimados para un poder estadístico de 80 o 90%, y se varían las desviaciones estándar para que sean iguales o bien 50% mayores para el grupo de tratamiento o de control. Para que el análisis sea más sencillo y transparente ponemos las estadísticas de medias en términos de desviaciones estándar, por lo que la desviación estándar propuesta es con valor 1, o bien para ser más conservadores consideramos una desviación estándar en el grupo de control de 1.5 veces la del tratamiento, o bien viceversa. La variable clave aquí es el efecto estimado con los parámetros de poder estadístico y de desviación estándar. El efecto estimado no es más que la diferencia entre los grupos de tratamiento y control para la variable de interés.

En la tabla se muestra algo muy relevante para la investigación, si se considera una desviación estándar del grupo de control de 1.5 veces la del grupo de tratamiento (renglones 2 y 3), se debe tener un efecto estimado de 0.06-0.07

desviaciones estándar para que se rechace la hipótesis nula a un poder de 80-90% y un nivel de significancia de 5%. Por su parte, considerando una desviación estándar de 1 a 1 entre grupo de tratamiento y control, el efecto estimado es menor (0.05-0.06) que con una razón de desviación estándar mayor. Al cambiar la razón de desviaciones estándar, que la desviación estándar en el grupo de tratamiento sea mayor que la de control, se tiene que el efecto estimado sube a 0.072-0.084 desviaciones estándar. Es decir, si la varianza entre el tratamiento es mayor se requiere que el impacto estimado sea también mayor, esto para poder rechazar la hipótesis nula propuesta a un poder estadístico de 80-90% y un nivel de significancia de 5%. Si el efecto estimado es más pequeño, por ejemplo 0.05 o menor, dependerá de qué tanta varianza es absorbida por las variables de control para poder lograr significancia estadística a un poder de 80%.

El análisis de poder puede variar también si se restringe el número de observaciones para hacer aparejamiento, por ejemplo. En este método solo se usan las observaciones que sean muy similares en un puntaje. Por tanto, el número de observaciones en los grupos respectivos disminuye, y el análisis de poder estadístico cambia. Pero al mismo tiempo es posible que la variación disminuya. Un análisis más profundo se puede realizar al momento de tener las muestras finales con aparejamiento.

10. Resultados para evaluación de impacto con metodología cuantitativa

En esta sección se explican los resultados encontrados de acuerdo con la metodología establecida previamente. Para facilitar la lectura se explica de forma

breve cada método, así como las decisiones que se tomaron en términos de robustecer los resultados al estimar el modelo de diferentes formas.

a. Metodología 1: Regresión múltiple

En el caso de regresión múltiple se estima la ecuación (1) mostrada arriba: $Y_{i;t+j} = \alpha + \gamma T_i + \pi Y_{i;t} + \vartheta_K X_{i;t_K} + \varepsilon_{i;t+j}$. Se tienen cuatro variables dependientes con los mismos controles. Estas variables dependientes son el número de artículos, artículos ponderados por factor de impacto (si no hay factor de impacto encontrado se le da un valor de cero), número de capítulos de libros, y número de libros. Todas las variables son estandarizadas por área de conocimiento y por año de evaluación. Es decir, para cada área de conocimiento y año de evaluación se calcula la media y desviación estándar y la variable estandarizada resulta de la variable de interés (número de artículos, libros, etc) menos su media entre la desviación estándar.

Además, para conocer la sensibilidad de los resultados se usan cuatro tipos de modelos para cada variable dependiente, para un total de 16 estimaciones. El método 1 incluye 25 variables de control las cuales son además del tratamiento, edad, sexo, número de artículos, número de artículos ponderados por factor de impacto, número de capítulos de libros, número de libros para el periodo pre-evaluación, variables dicotómicas que controlan por el nivel inicial antes de la evaluación y el número de años entre evaluaciones, variables por área de conocimiento y por año de evaluación. Para el método 2 se incluyen polinomios de las variables de productividad e interacciones para un total de 35 variables de control. Si existen variables omitidas que afecten el tratamiento y la variable de

resultado al mismo tiempo y éstas están correlacionadas con esas interacciones entonces el estimador sobre el tratamiento debería cambiar sustancialmente. Para el método 3 utilizamos un procedimiento conocido como post-lasso (Belloni, Chernozhukov y Hansen 2014) donde se penalizan las variables que no están correlacionadas ni con la variable de resultado ni con el tratamiento. Este procedimiento ayuda para quedarse únicamente con las variables más importantes. Finalmente, el método 4 estima un modelo Tobit para permitir que la variable dependiente está censurada en cero. Es decir, tenemos muchos investigadores con producción de cero en diferentes componentes académicos, y el modelo Tobit ayuda a tomar ese aspecto en consideración.

Los resultados se muestran en la Tabla 7. En general, los resultados son muy consistentes entre modelos. En el método 1 que incluye 25 variables de control tenemos que ser promovido un nivel en el SNI (tratamiento) en comparación con no ser promovido (y mantenerse en nivel 1 o 2, control) eleva la productividad de los investigadores en 10% de una desviación estándar en número de artículos. Conforme ajustamos el método el resultado es sumamente estable entre 8-10% de una desviación estándar.

Este efecto no es sustancialmente grande cuando se compara con estimadores encontrados en la bibliografía de intervenciones en educación y aprovechamiento académico. Fryer (2016) revisa 196 experimentos de campo en educación para diferentes tipos de intervenciones. Los mayores impactos son encontrados en intervenciones en edad temprana de 0 a 5 años (0.12 desviaciones estándar) y en sesiones de práctica con tutor (0.39). Por otro lado, incentivos monetarios a los estudiantes o profesores tienen impactos de 0.04 a 0.05

desviaciones estándar. Entonces el resultado encontrado es relativamente similar al encontrado en pruebas de aprovechamiento cuando se otorga un incentivo monetario. Por tanto, el resultado es significativo en el área de productividad académica. El impacto es relativamente similar cuando se usan otros indicadores como artículos ponderados con factor de impacto, capítulos de libros, y libros.

Tabla 7. Resultados con método de regresión.

Variables	Método 1	Método 2	Método 3	Método 4
Artículos	0.101*** [0.016]	0.092*** [0.016]	0.100*** [0.016]	0.079*** [0.010]
Artículos con Factor de Impacto	0.073*** [0.017]	0.072*** [0.017]	0.073*** [0.017]	0.052*** [0.010]
Capítulos	0.057*** [0.017]	0.049*** [0.018]	0.059*** [0.017]	0.052*** [0.010]
Libros	0.097*** [0.020]	0.088*** [0.020]	0.097*** [0.020]	0.069*** [0.012]
N	17,302	17,302	17,302	17,302
# Variables control incluidas	25	35	15	25

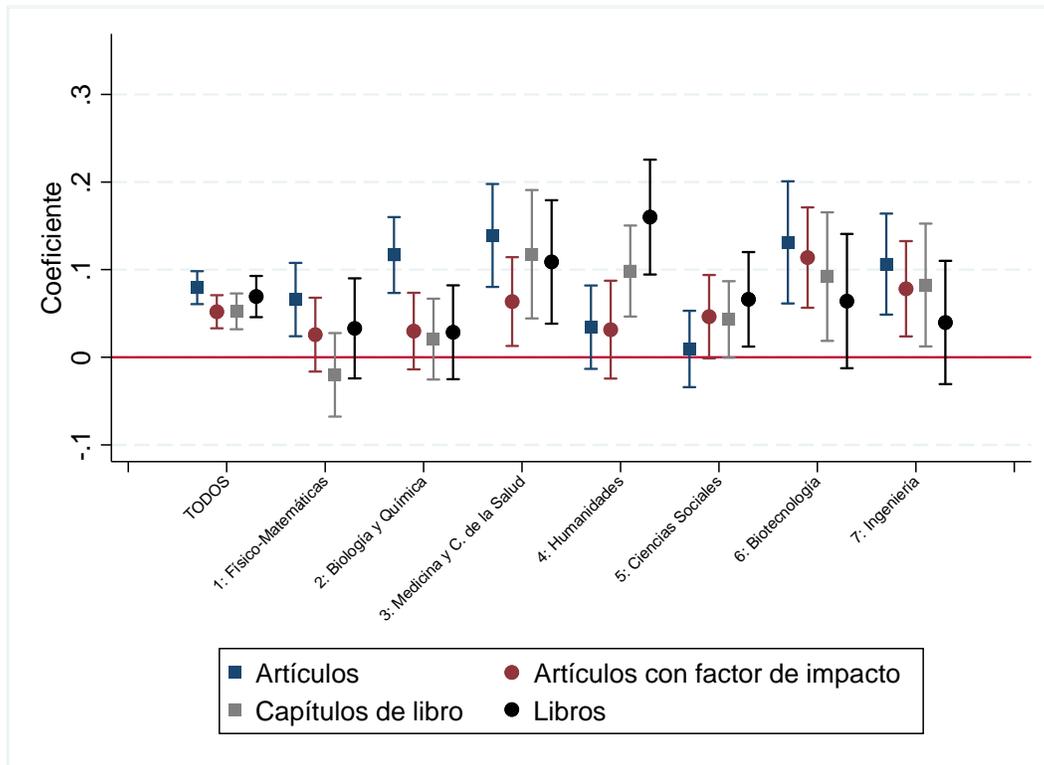
Notas: Cálculos por el autor. Errores estándar robustos. El método 1 incluye 25 variables de control las cuales son además del tratamiento, edad, sexo, número de artículos, número de artículos ponderados por factor de impacto, número de capítulos de libros, número de libros para el periodo pre-evaluación, variables dicotómicas que controlan por el nivel inicial antes de la evaluación y el número de años entre evaluaciones, variables por área de conocimiento y por año de evaluación. Para el método 2 se incluyen polinomios de las variables de productividad e interacciones para un total de 35 variables de control. El método 3 implementa post-lasso para el método 2. El método 4 es como el método 1 pero estimado con modelo Tobit.

Se podría pensar que los controles no son suficientes y se requiere tener controles más flexibles. Para ello utilizamos polinomios e interacciones de las variables de productividad académica. Los resultados son muy similares al método 1, aunque en general un poco más pequeños. Sin embargo, cuando realizamos la

regresión con el método post-lasso que selecciona las mejores variables a incluir se tiene un estimador muy similar al del método 1. Finalmente, cuando estimamos un modelo tobit el cual considera que la variable dependiente incluye ceros (censurada en ese valor) los estimadores son más pequeños que en el método 1, pero siguen siendo estadísticamente significativos y diferentes de cero. Considerando todos los resultados, aumentar un nivel en el SNI eleva la productividad académica más en números de artículos y libros, y menos (pero diferente de cero) en artículos ponderados por factor de impacto y capítulos de libros. Es de resaltar que siempre el impacto en números de artículos es mayor que en la variable de número de artículos ponderados por factor de impacto. Esto significa que los investigadores que son promovidos sí aumentan su productividad, pero más en artículos publicados en revistas no indizadas que en revistas con alto factor de impacto.

A la luz de los resultados, vale la pena analizar los impactos a nivel área de conocimiento. Estos se muestran en la Figura 2 para el modelo tobit (método 4, que es el más conservador en la tabla 7) con intervalos de confianza robustos al 95%. La figura incluye en el eje x el resultado para todos y para cada área de conocimiento. Al tener el mayor tamaño de muestra cuando se incluye a todos los investigadores se tiene los menores intervalos de confianza. Algunas áreas tienen bajo tamaño de muestra y por tanto los intervalos de confianza se amplían. En estos casos no tenemos suficiente poder estadístico a nivel área de conocimiento para rechazar la hipótesis nula de no efecto en subir en el SNI. Pero es de resaltar que para la mayoría de las estimaciones sí podemos rechazar la hipótesis nula de no efecto.

Figura 2. Resultados por área de conocimiento.



Notas: Cálculos por el autor. Intervalos de confianza robustos al 95%.

Para el caso de área 1, se tiene que los impactos en número de artículos son significativos, para número de artículos ponderados por factor de impacto y libros es positivo, con impacto bajo, pero imprecisamente estimado (es decir con errores estándar relativamente grandes o intervalos de confianza amplios). Para al área 2 algo similar ocurre, se tiene un impacto significativo en número de artículos, pero un impacto positivo relativamente bajo en el resto de las variables con estimación imprecisa. Para el área 3 se tienen impactos estadísticamente significativos para todas las variables, y los impactos son grandes, con un valor de 10% de una desviación estándar. Para el área 4 de humanidades solo se encuentran efectos significativos en libros y capítulos de libros. Para el área 5 de ciencias sociales se

encuentran efectos marginalmente significativos para artículos ponderados por factor de impacto, capítulos de libro, y para libros, aunque no se encuentran impactos en número de artículos. Para las áreas 6 y 7 se encuentran efectos marginalmente significativos al 10% de una desviación estándar para todas las variables y al 5% para todas excepto número de libros. Las áreas con el mayor impacto en número de artículos son Biología, Química y Ciencias de la Vida (área 2), Medicina y Ciencias de la Salud (área 3), Biotecnología y Ciencias Agropecuarias (área 6) e Ingeniería (área 7). Las áreas donde impactan más el número de libros son humanidades (área 4) y, medicina y ciencias de la salud (área 3).

Es de llamar la atención que en la mayoría de las áreas (con excepción del área 5 de ciencias sociales) el impacto disminuye o se hace cero al considerar el factor de impacto de la revista donde se publica el artículo. Esto es consistente con las estadísticas descriptivas mostradas anteriormente, donde la mediana de artículos publicados con factor de impacto es 0. En las áreas 1, 2 y 4 se reduce y se llega a cero, y en las áreas 3 y 7 el estimador de número de artículos con factor de impacto es casi la mitad del de sin considerar el factor de impacto. Entonces estos resultados nos están indicando que, si bien aumentar de nivel SNI aumenta la productividad académica en número de artículos para la mayoría de las áreas de conocimiento, no tiene un resultado recíproco al considerar el factor de impacto. En otras palabras, se logra aumentar el número de artículos, pero no tanto en revistas con alto factor de impacto.

Finalmente, es de destacar que en el área 4 de humanidades la promoción en el SNI no tiene impactos en artículos con o sin considerar el factor de impacto. Únicamente se tiene en capítulos de libros, pero principalmente en libros. Esto se

podría esperar ya que esta área se especializa más en estos últimos que en los artículos de revistas académicas.

b. Metodología 2: Método de aparejamiento

En esta subsección se analiza otro método para obtener el impacto de ser promovido en el SNI en comparación al no ser promovido. El método propuesto es el de aparejamiento por un puntaje. En este método se tienen dos pasos. En el primero se obtiene un puntaje obtenido con algún método en específico, y en el segundo paso se hace un aparejamiento con base en ese puntaje y ciertas restricciones en el aparejamiento. Es importante mencionar que todos los aparejamientos se realizan dentro de cada área de conocimiento y respetando el nivel inicial y años entre evaluación de cada investigador. Por ejemplo, se compara únicamente investigadores dentro del área 1 entre aquellos que iniciaron en nivel 1 y tardaron 3 años de evaluación, y así para todas las áreas, niveles iniciales (1 o 2), y años entre evaluaciones (3-4 para el nivel 1 y 4-5 para el nivel 2).

Para facilitar la comparación, así como ser más robustos en la estimación se siguen cuatro métodos que cambian la forma en que se estima el puntaje (primer paso) como la forma en que se realiza el aparejamiento (segundo paso). El método 1 y 2 estiman un puntaje de propensión usando un modelo logit. Se tiene como variable dependiente si el investigador o investigadora están en el grupo de tratamiento o control y como variables explicativas las variables en el modelo extendido del método 2 de regresión múltiple que incluye 35 variables. El método 3 estima, en lugar de un modelo logit, un modelo de componentes principales (análisis discriminante) para predecir un puntaje basado en las variables pre-evaluación de

artículos, artículos ponderados por factor de impacto, capítulos de libro y libros. El método 4 usa el puntaje estimado en los métodos 1 y 2 pero lo transforma usando la razón de momios (el logaritmo de la razón del puntaje sobre uno menos el puntaje). Este método se ha utilizado para evitar puntajes con valores muy cercanos a cero o a la unidad.

El método de aparejamiento en la primera propuesta de estimación usa hasta un máximo de cuatro vecinos más cercanos en una vecindad del puntaje de 0.005 unidades. Este método permite tener más observaciones del grupo de control y aumentar la muestra, pero al mismo tiempo se es muy exigente para encontrar vecinos pues el valor de la vecindad que busca es muy pequeño considerando que el puntaje tiene un rango entre 0 y 1. El segundo método utiliza todos los vecinos disponibles en una vecindad todavía más reducida de 0.0005 unidades en el puntaje de propensión. El tercer método vuelve a usar cuatro vecinos, pero dado que el valor del puntaje cambia se usa una vecindad de 0.025 unidades. Finalmente, el cuarto método usa el logaritmo de la razón de probabilidades con todos los vecinos disponibles en un rango de 0.001 unidades.

Como se discutió, el método de aparejamiento tiene ventajas y desventajas. La clara ventaja es que se está comparando investigadores tan similares como es posible pero un investigador sí recibió la promoción y otro no. Sin embargo, se tiene una desventaja fuerte en el sentido que al buscar los vecinos óptimos se reduce sustancialmente la muestra. Esto se debe a que no todos los individuos en el grupo de tratamiento tienen un individuo en el grupo de control que sea comparable, y también a que muchas observaciones en el grupo de control no son válidas para usar en el grupo de tratamiento. Todo esto lleva a que los errores estándar de la

estimación crezcan, y por tanto con esta metodología nos hace falta poder estadístico para rechazar la hipótesis nula de no efecto en productividad académica. Este problema se agrava, por supuesto, al estimar por área de conocimiento.

El primer aspecto que se debe de corroborar al usar aparejamiento por puntajes de propensión es que efectivamente se logre un balance en características observables. Es decir, una vez que se realiza el aparejamiento no debiera de existir ninguna diferencia estadísticamente significativa en variables clave entre los grupos de tratamiento y control. Mientras más diferencias haya significa que el aparejamiento es de peor calidad. Las pruebas que se presentan son diferencias en ciertas variables clave incluyendo su valor p de probabilidad, así como la prueba F con su valor p de probabilidad donde se incluye una regresión de la variable dummy de tratamiento contra las variables más importantes en el pre-tratamiento. Si el aparejamiento es de calidad entonces esta prueba no debe ser rechazada estadísticamente, lo que significa que las variables no ayudan a predecir si el investigador está en el grupo de tratamiento o control. Estas pruebas incluyen los ponderadores obtenidos por el aparejamiento para que se compare únicamente la muestra tratada y control.

Los resultados se muestran en la Tabla 8. En el panel A se incluyen las pruebas de diferencias de medias para variables claves en cada método utilizado. En el panel B se incluye la prueba F. Finalmente, el panel C incluye el número de observaciones finales que se logran al realizar el procedimiento de aparejamiento en los grupos de tratamiento y control. Debajo de cada renglón de variable se incluye el valor p de la prueba correspondiente.

Tabla 8. Balance en características observables.

Pruebas	Método 1	Método 2	Método 3	Método 4
A. Variables				
Mujer	0.004 [0.689]	-0.006 [0.610]	-0.053** [0.000]	-0.005 [0.749]
Edad	-0.041 [0.824]	-0.149 [0.497]	-0.526** [0.005]	-0.031 [0.913]
Artículos	-0.018 [0.477]	-0.015 [0.559]	0.091** [0.000]	-0.024 [0.421]
Artículos con Factor de Impacto	0.018 [0.456]	0.013 [0.547]	0.083** [0.000]	0.008 [0.764]
Capítulos	-0.019 [0.443]	-0.015 [0.547]	-0.047** [0.009]	-0.026 [0.389]
Libros	-0.013 [0.577]	-0.046* [0.067]	-0.081** [0.000]	-0.025 [0.419]
B. Prueba F				
Estadístico F	0.315	0.741	11.764**	0.320
P-Valor	[0.930]	[0.616]	[0.000]	[0.927]
C. # Observaciones				
Tratamiento	3466	2514	3392	1604
Control	6654	5609	6466	2108

Notas: Cálculos por los autores. Valor p en corchetes. * (**) estadísticamente significativo al 5% (1%).

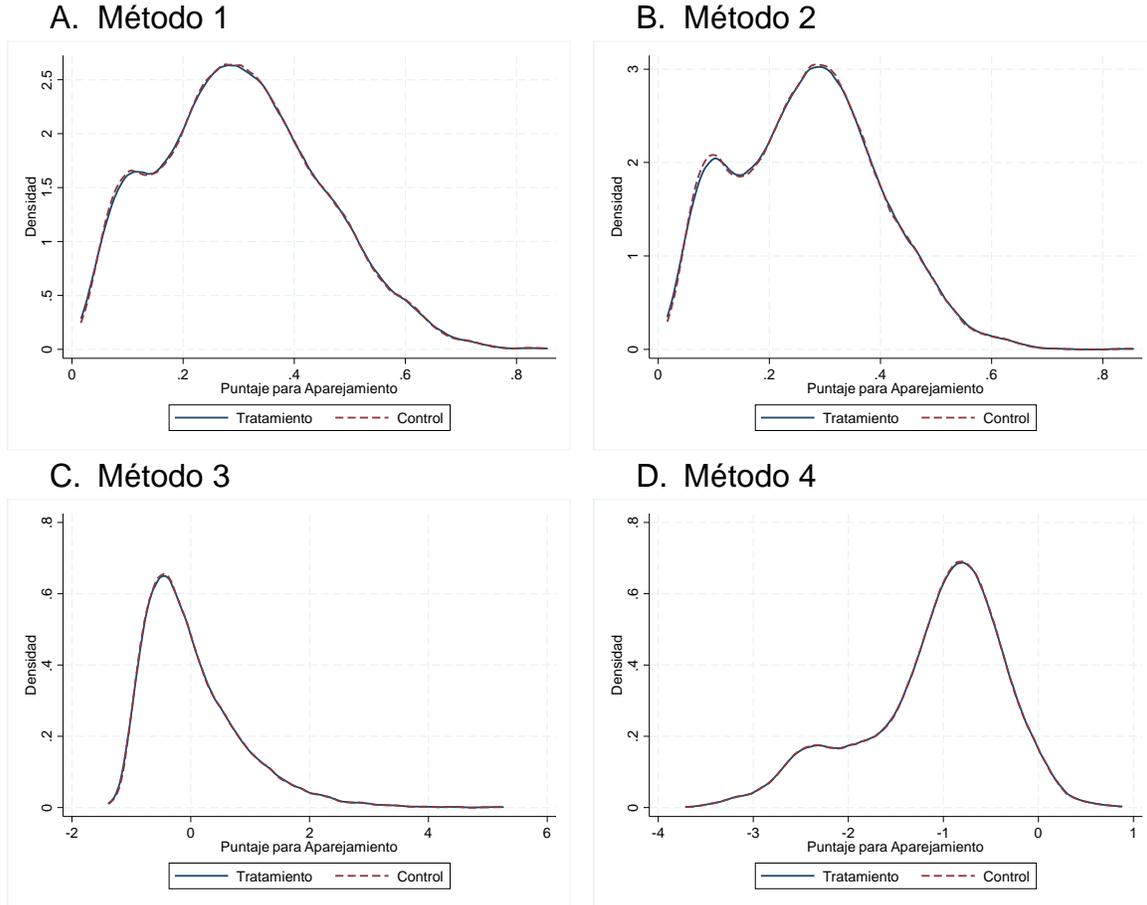
La tabla muestra que los mejores métodos son el 1, 2 y 4. El método 1 tiene una leve diferencia que no es significativa al 1% en artículos con factor de impacto en el periodo pre-evaluación pero tiene la ventaja que es el grupo con mayor número de observaciones en el grupo de tratamiento y de control. Por otro lado, el método 2 hace un muy buen trabajo también, pero tiene menos número de observaciones. Sin embargo, este tamaño de muestra es sustancialmente menor al utilizado en la metodología 1 de regresión múltiple. El apareamiento de menor calidad es el método 3 pues se observan diferencias estadísticamente significativas en variables clave, así como la prueba F se rechaza la hipótesis nula.

Finalmente, la tabla en el panel C muestra las observaciones que se obtienen después de implementar el apareamiento. El número de observaciones en los grupos de tratamiento y control en todos los casos es menor que el número de observaciones al usar regresión. Esto va ocasionar que los errores estándar de los estimadores crezcan, que los intervalos de confianza sean más amplios y que se corra el riesgo de no tener poder estadístico para rechazar la hipótesis nula de no impacto. Esto será una mayor preocupación cuando se estimen impactos por área de conocimiento.

Otro aspecto a considerar al implementar el método de apareamiento de puntajes es qué tanto el soporte común es similar al aplicar el método entre los grupos de tratamiento y control. Lo óptimo es que la distribución del puntaje después del apareamiento sea tan similar como sea posible. De lo contrario no se podría lograr comparaciones similares entre investigadores.

Para ello, la Figura 3 incluye la densidad (no paramétrica) de los puntajes para cada método. En general, todos los métodos logran que los puntajes sean muy similares entre grupos de tratamiento y control. Se destacan los resultados para los métodos 1, 2, y 4, pero para el método 3 se observa ciertas partes con algunas diferencias. Por tanto, en resumen, los métodos más confiables son los métodos 1 y 2, y quienes tienen una mayor muestra. Se presentan resultados agregados para todos los métodos, mientras que los resultados por área de conocimiento se presenta únicamente el método 1. En el Anexo se incluyen los resultados para todos los métodos restantes.

Figura 3. Densidades y balance en los diferentes métodos



Nota: Cálculos por el autor.

La Tabla 9 incluye los resultados de diferencias en diferencias usando el método de apareamiento con puntajes. Al igual que en la subsección anterior, se tienen 4 variables donde se busca analizar si se tiene algún impacto de una promoción en el SNI. En cada renglón se tiene el estimador de diferencias en diferencias (controla por diferencias en el grupo de tratamiento y control antes de la evaluación) y debajo de cada estimador se tiene el error estándar respectivo. El error estándar se obtiene de un procedimiento de remuestreo (bootstrap) con 500 replicaciones.

Se resalta que con el método de aparejamiento se obtienen resultados muy similares al de regresión múltiple si consideramos los mejores métodos 1, 2 y 4. El método 1 es el más confiable en términos de balance en características observables y tamaño de muestra. Con este método se tiene un impacto positivo y significativo en las variables de artículos, artículos ponderados por factor de impacto, y libros. Los impactos están entre 0.05-0.10 de una desviación estándar, y son estadísticamente significativos. El método 2 que es el siguiente más confiable obtiene resultados muy similares al método 1 aun cuando las observaciones se reducen sustancialmente en los grupos de tratamiento y control. Estas estimaciones son consistentes con lo encontrado en la subsección anterior y dejan ver que los resultados son robustos al método elegido. Como se mencionó, dado que la muestra es más pequeña se tienen impactos más imprecisos, pero la magnitud es muy similar a la de regresión múltiple.

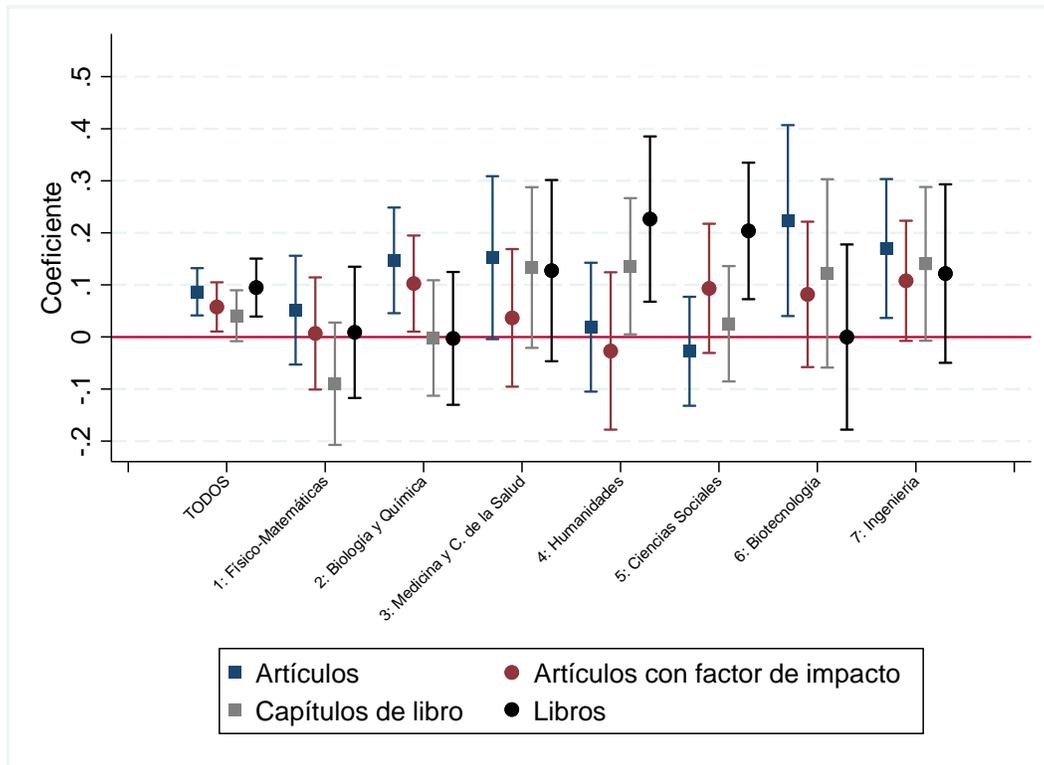
Tabla 9. Resultados de Diferencias en Diferencias con el método de aparejamiento.

Variable	Método 1	Método 2	Método 3	Método 4
Artículos	0.087*** [0.023]	0.078*** [0.030]	0.040* [0.023]	0.090** [0.043]
Artículos con Factor de Impacto	0.058** [0.024]	0.068** [0.029]	0.021 [0.024]	0.054 [0.043]
Capítulos	0.041 [0.025]	0.023 [0.033]	0.058** [0.023]	-0.013 [0.049]
Libros	0.095*** [0.028]	0.117*** [0.035]	0.129*** [0.030]	0.049 [0.053]

Notas: Cálculos por el autor. Error estándar debajo de cada estimador en corchetes. Errores estándar se obtuvieron por un procedimiento de remuestreo (bootstrap) con 500 replicaciones. *(**)(***) representa significancia estadística al 10%(5%)(1%).

Al igual que antes también se pueden calcular impactos por área de conocimiento. Sin embargo, como se mencionó, el tamaño de muestra se reduce sustancialmente con el método de aparejamiento por lo que los intervalos de confianza se amplían y en la mayoría de las áreas de conocimiento no se tiene el poder estadístico para rechazar la hipótesis nula. De cualquier manera, los resultados de la subsección anterior por lo general se mantienen, ver Figura 4. En primer lugar, casi todos los impactos de artículos ponderados con factor de impacto son menores a los impactos sin considerar el factor de impacto (con excepción de ciencias sociales). Asimismo, el número de libros es particularmente importante para el área 4 de humanidades. Es de llamar la atención que el coeficiente para los artículos ponderados por factor de impacto en las áreas 3 y 6 se reduce sustancialmente en comparación con el de número de artículos totales.

Figura 4. Resultados por área de conocimiento



Nota: Cálculos por el autor. Intervalos de confianza robustos al 95%. Se muestran resultados para el método de diferencias en diferencias con apareamiento en puntaje de propensión con modelo logit, 4 vecinos en una vecindad de 0.005 (método 1).

11. Resultados para evaluación de impacto con metodología cualitativa

En esta sección se presentan resultados de la evaluación de impacto con metodología cualitativa. Se enviaron más de 800 correos a una muestra de investigadores con las características de este estudio (evaluaciones de nivel 1 y 2). Los correos incluían 4 preguntas a contestar de forma abierta. Pregunta 1: ¿Qué significa el Sistema Nacional de Investigadores para usted? Pregunta 2: ¿Cómo fue su experiencia cuando fue evaluado por el SNI para su nombramiento actual? Pregunta 3: ¿Cómo siente que el SNI ha influenciado su productividad académica?

Pregunta 4: ¿Considera que la forma en como incentiva el SNI la investigación científica es la adecuada? ¿Sí o No? ¿Por qué?

Analizar cuantitativamente las respuestas a esas preguntas es complicado. Para realizarlo presentamos primero nubes de palabras y frecuencias de palabras más utilizadas en cada pregunta (esta metodología está motivada por los siguientes estudios: Gentzkow et al. 2017; Jänicke et al. 2016; Wilkerson y Casas 2017). Después analizamos el sentimiento de la respuesta a cada palabra, así como algunas respuestas típicas o llamativas.

a. Frecuencias de palabras

La Figura 5 muestra nubes de palabras utilizadas en cada respuesta. El tamaño de cada palabra se refiere a qué tanto es utilizada en cada pregunta, un mayor tamaño implica mayor uso de la misma. Para la pregunta 1 las palabras más utilizadas son investigadores, investigación, sni, reconocimiento, económico. En esta pregunta se tiene una dicotomía de si el SNI representa más reconocimiento entre la sociedad o bien un estímulo económico o apoyo al salario. En general, significa ambos, pero la palabra reconocimiento es más utilizada que la palabra salario o apoyo.

Figura 5. Nubes de palabras utilizadas en cada respuesta.

A. Pregunta 1

B. Pregunta 2

la pregunta 3 sobre influencia del SNI las palabras más repetidas son productividad, investigación, sni, estímulo, económico. En esta pregunta se menciona más estímulo y económico que la palabra reconocimiento. También una palabra muy mencionada fue motivación. En general en esta pregunta los investigadores se expresaron positivamente sobre la influencia del SNI. Finalmente, la pregunta 4 sobre incentivos entre las palabras más mencionadas son criterios, calidad, adecuada, revistas, entre otras. Un aspecto que preocupa a muchos investigadores es que los criterios de evaluación no son claros, y adicionalmente que el peso de las publicaciones en revistas con diferentes niveles de calidad no es conocido.

La Figura 6 conjuga todas las respuestas a las preguntas y muestra las palabras más utilizadas en ese conjunto. Entre las palabras más importantes están reconocimiento, productividad, estímulo, calidad. Un aspecto que preocupa a los investigadores es la evaluación de pares y la medición de calidad de las investigaciones, y a la vez que estar en el sistema les provee de un reconocimiento importante.

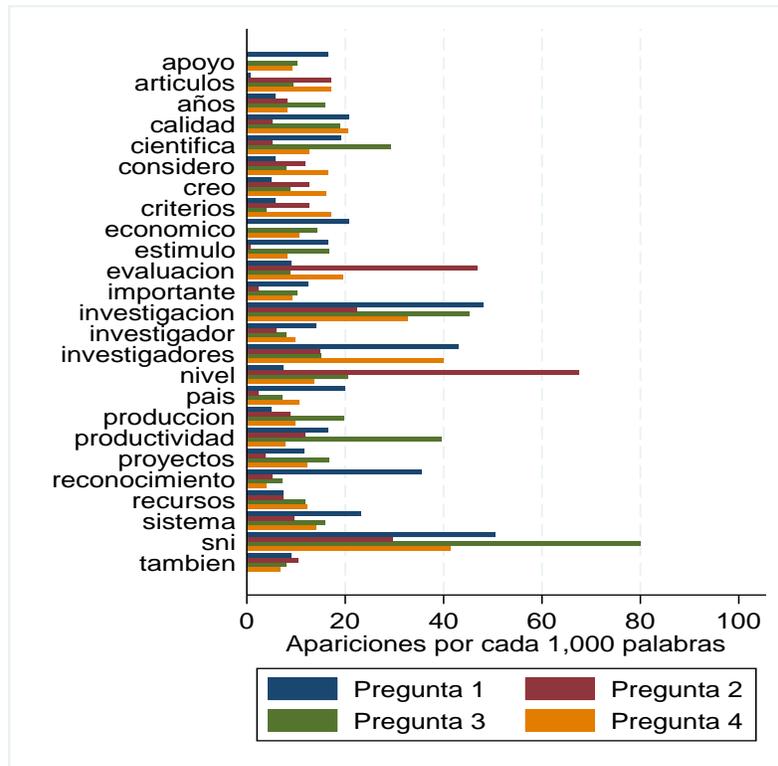
Figura 6. Nube de palabras usando las 4 preguntas utilizadas.



Nota: Cálculos por el autor. Palabras de un mayor tamaño quiere decir que son usadas más frecuentemente.

La Figura 7 resume lo encontrado anteriormente, pero de forma cuantitativa. Se incluyen las 25 palabras más utilizadas en total y se grafica el número de menciones por cada 1,000 palabras (del total del corpus) para cada pregunta. Las palabras más mencionadas son sni, evaluación, investigación, investigadores, nivel. Es de resaltar que en algunos casos las palabras de la pregunta 1 no aparecen o es muy bajo en la pregunta 2 (apoyo, económico y estímulo). La palabra calidad es mucho más relevante en las preguntas 1 y 4 que en las 2 y 3. La palabra productividad es más común para la pregunta 3 que en otras. La palabra reconocimiento es más importante para la pregunta 1 que otras.

Figura 7. Número de apariciones en las preguntas para las 25 palabras más utilizadas



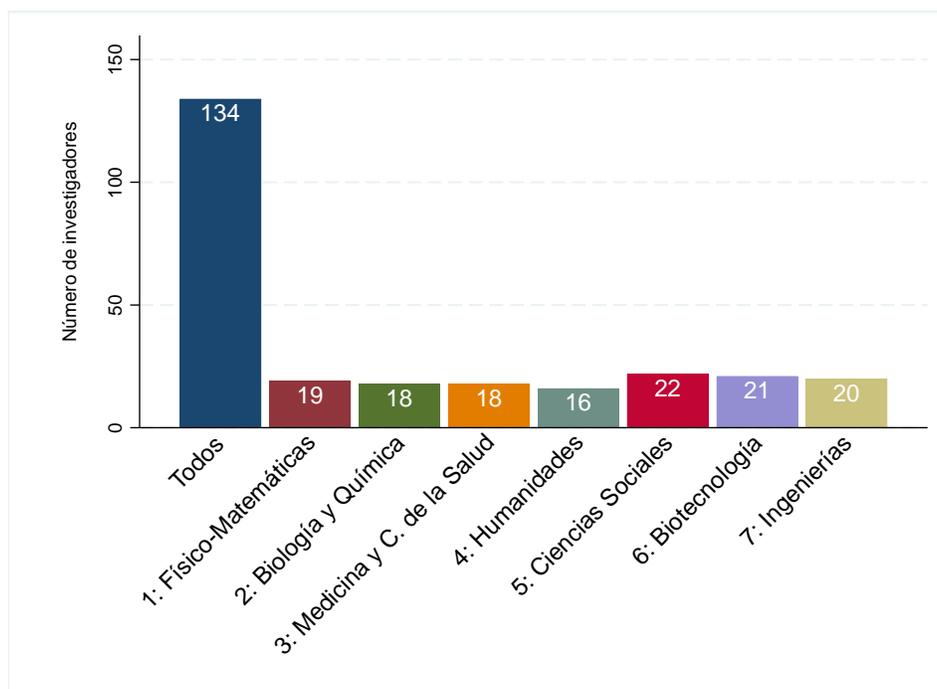
Notas: Cálculos por el autor. Datos recabados en la encuesta enviada por correo electrónico.

b. Sentimiento en las preguntas

En esta subsección se enfatizan los resultados generales obtenidos en la subsección anterior. Este análisis de sentimiento parte de la recopilación de respuestas de los investigadores sobre sus experiencias como resultado de una evaluación para subir o mantenerse de distinción. Esta metodología es cada vez más utilizada en ciencias sociales (Gentzkow et al. 2017; Jänicke et al. 2016; Wilkerson y Casas 2017). Este proceso se puede considerar como complementario al análisis cuantitativo, pues permite entender de mejor manera posibles canales

que expliquen los resultados encontrados, así como sentimientos asociados con el SNI (Honig 2018).

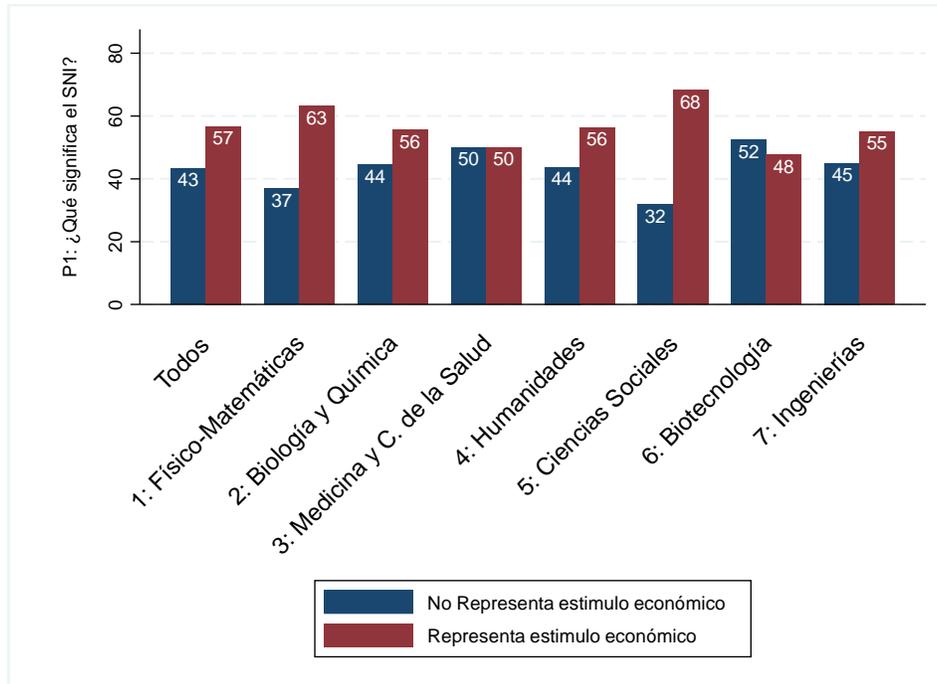
Figura 8. Número de investigadores que respondieron encuesta.



Notas: Cálculos por el autor.

En este contexto, se inicia el análisis de sentimientos partiendo con el número de investigadores que pudieron realizar la encuesta. Esto se refleja en la Figura 8. De los más de 800 correos enviados se tuvo un total de 134 investigadores que emitieron una respuesta a las preguntas de la encuesta. Ellos están distribuidos aproximadamente de manera equitativa para las siete áreas de conocimientos registradas en el Sistema Nacional de Investigadores. El área con mayor participación al momento de responder la encuesta es la de Ciencias Sociales, pero en general las respuestas de los investigadores están balanceadas entre áreas.

Figura 9. Pregunta 1: Representa estímulo económico el SNI. (Porcentaje de respuesta)

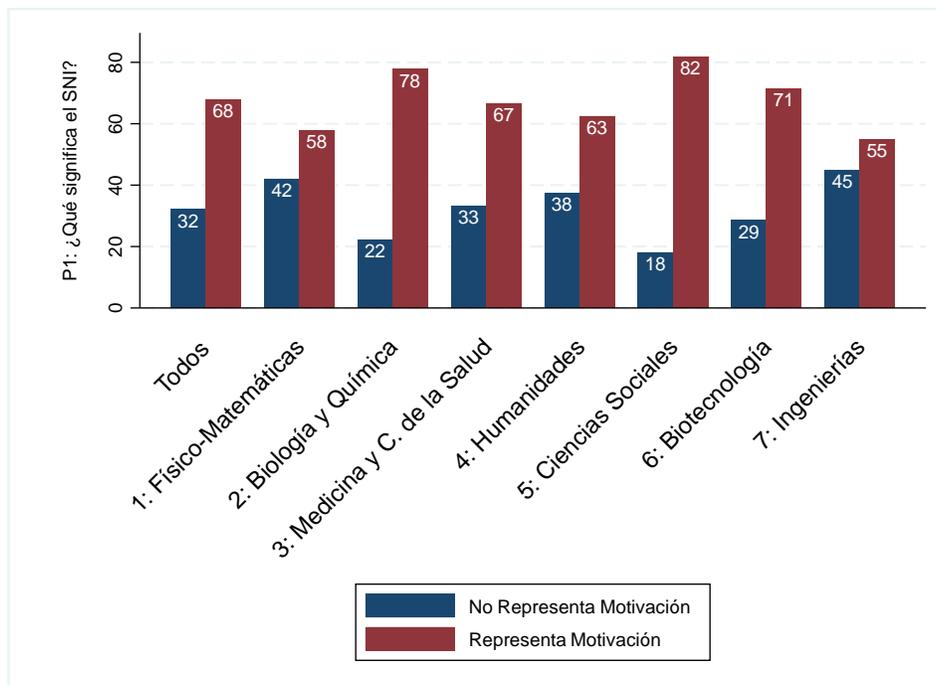


Notas: Cálculos por el autor.

El siguiente procedimiento del análisis de sentimiento, con respecto a las respuestas brindadas por los investigadores es la cuantificación de la expresión en variables categóricas. En este aspecto la pregunta 1 de la encuesta se puede apreciar de dos maneras: (1) desde una percepción económica y (2) desde una percepción de motivación. Recordemos que en esta pregunta no se dan opciones sino que son abiertas para que los investigadores expresen lo que crean más conveniente. Cada pregunta fue catalogada en diferentes sentimientos o grupos de respuesta. La Figura 9 muestra que en la mayoría de las áreas de conocimiento existe un porcentaje alto de investigadores que perciben al SNI como una fuente de estímulo económico (en promedio 57% para todos los investigadores). Además, aun sin desagregar por áreas se sigue manteniendo la predominancia de percibir al SNI

como algo económico. El área que presenta una mayor percepción de estímulo económico es el área 5 de ciencias sociales, y la que menor percepción tiene es el área 6 de biotecnología.

Figura 10. Pregunta 1: Representa motivación el SNI. (Porcentaje de respuesta)

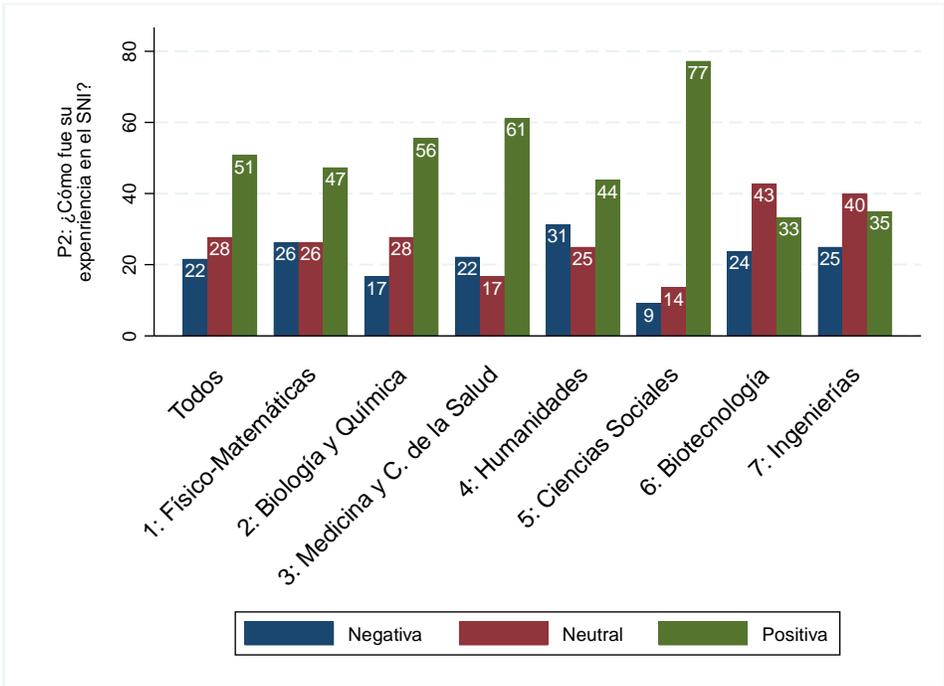


Notas: Cálculos por el autor.

La segunda forma de lo que significa el SNI se refleja en la figura 10. Aquí se cataloga si los investigadores piensan que el SNI representa motivación. En esta, como en la figura anterior, puede darse el caso de que al investigador el SNI le signifique una fuente de estímulo económico como una de motivación (no son mutuamente excluyentes). Se define motivación en el sentido de seguir investigando o ser más productivo. Con ello, la figura muestra que, para todas las áreas de conocimiento, la percepción sobre que el SNI motiva a los investigadores a seguir publicando más para su historial de producción académica (artículos, libros,

capítulos de libros, etc.). El área de conocimiento con menor percepción de que el SNI sea una motivación es la de Ingenierías. Por su parte el área con mayor representación de investigadores de que el SNI sea una motivación para su vida profesional es la de Ciencias Sociales. Es de resaltar que en general es más importante la motivación de ser parte del SNI (68%) que el estímulo económico (57%) cuando un investigador piensa sobre la significancia del mismo.

Figura 11. Pregunta 2: Experiencia en evaluación SNI. (Porcentaje de respuesta)

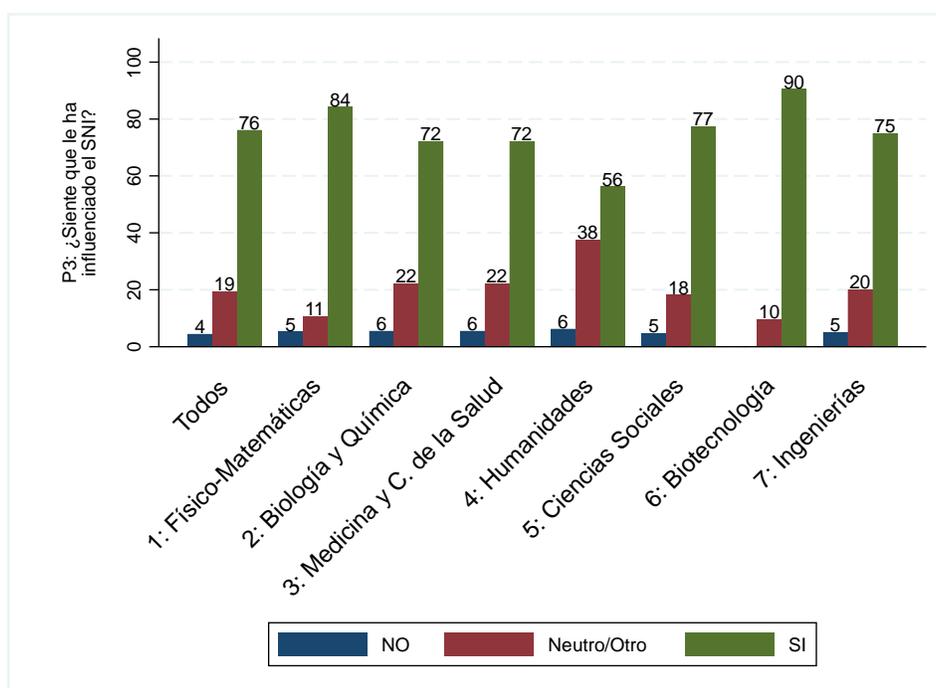


Notas: Cálculos por el autor.

La siguiente pregunta de la encuesta es responder sobre cómo fue su experiencia en el SNI cuando fue evaluado. Los resultados que se muestran en la Figura 11 representan una cuantificación de la respuesta en tres categorías: (1) la experiencia fue negativa, (2) la experiencia fue neutral y (3) la experiencia fue positiva. De la figura se percibe que los investigadores, para las cinco primeras

áreas de conocimiento, cuando fueron evaluados resultó ser una experiencia positiva. En cambio, para las áreas seis y siete se percibió que la experiencia del investigador fue neutral o justa. Cuando se toman a todas las áreas en conjunto, la predominancia es que existe una experiencia positiva para cuando se realizó la evaluación. El área con el sentimiento más negativo hacia la evaluación fue el área de humanidades.

Figura 12. Pregunta 3: Influencia del SNI. (Porcentaje de respuesta)



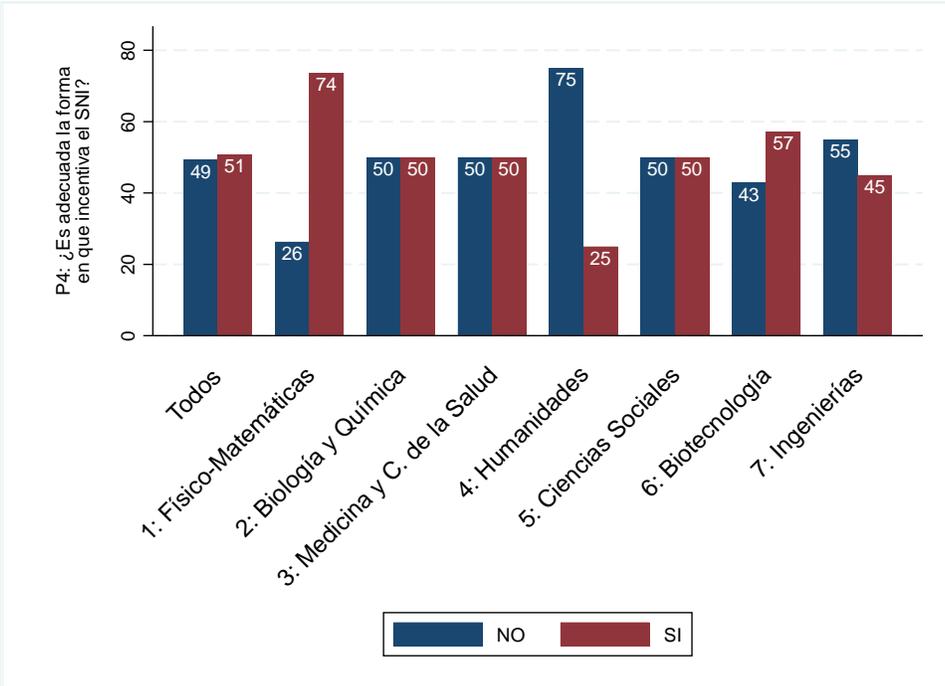
Notas: Cálculos por el autor.

El siguiente punto en la encuesta es sobre si el SNI ha sido una influencia para el investigador con respecto a su actividad académica. Los resultados de la pregunta se muestran en la figura 12. En ella se representa que, para todas las áreas de conocimiento, y también cuando se congregan todas las áreas, el investigador sí se vio influenciado por el SNI para incrementar su productividad académica ya que es apoyado económicamente y además percibe al SNI como un

reconocimiento que lo motiva a seguir siendo más productivo. Sin embargo, algunos investigadores mencionaron que el periodo tan corto de evaluaciones y el hecho de no perder el incentivo económico generaba mucho estrés y que su salud física y mental se deterioraba conforme se acercaba la fecha de evaluación.

Figura 13. Pregunta 4: ¿Es adecuada la forma que incentiva el SNI?

(Porcentaje de respuesta)



Notas: Cálculos por el autor.

El cuestionario de la entrevista se cierra con una cuarta pregunta que refleja si es adecuada la forma en como incentiva el SNI la productividad, y calidad, de los investigadores con respecto a sus investigaciones. Para esta pregunta solo se cuantifica de manera dicotómica, con una categorización de *SI* para cuando es adecuada la forma y de *NO* para cuando resulta que es lo contrario. Todo ello se muestra en la figura 13. Esta refleja resultados diversos para todas las áreas. Los

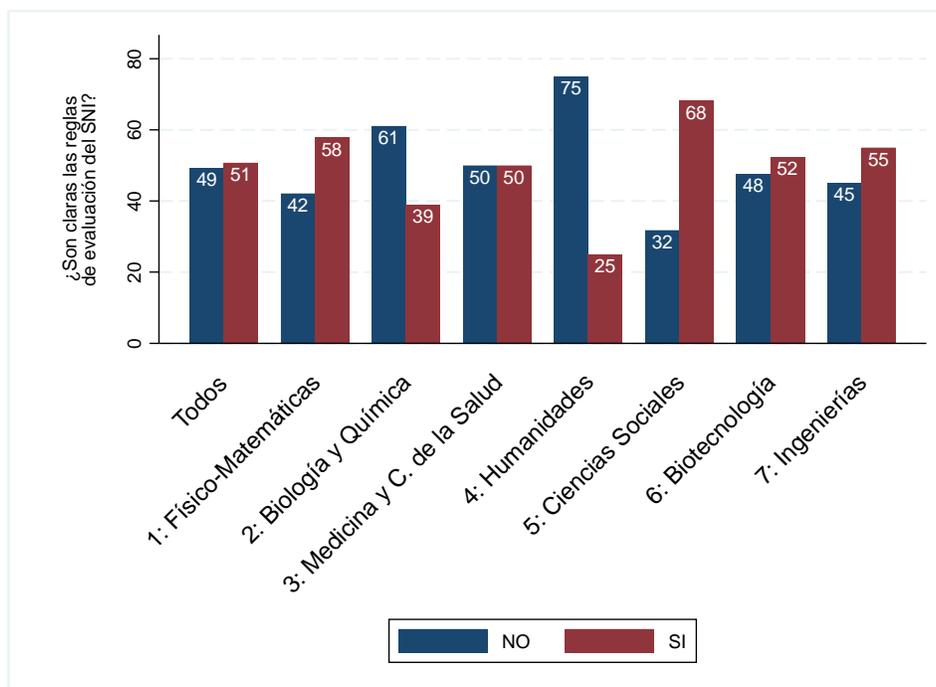
casos extremos son dos. El primero se muestra en el área de Físico-Matemáticas y Ciencias de la Tierra, donde hay un alto porcentaje de investigadores que perciben que la forma como incentiva el SNI es la adecuada. Por su parte, se tiene el caso opuesto con el área de Humanidades. Sin embargo, cuando se analizan todas las áreas como un solo conjunto, se tiene una distribución aproximadamente equitativa de investigadores que perciben ambas situaciones.

En el contexto de la pregunta cuatro, los investigadores hicieron comentarios sobre otros aspectos que se podrían tomar para incentivar la producción científica. El área con más comentarios es el de Humanidades. Los investigadores de esta área mencionan los siguientes aspectos: se deben incluir formas de incentivo diferente al económico (ampliar visibilidad a la sociedad sobre calidad de publicaciones, trascendencia de investigadores, entre otras); otra forma es fomentando encuentro entre sus miembros, reconociendo los saberes originarios, impulsando la transdisciplinariedad (por ejemplo, abrirse una nueva área donde entrarán los que laboran en el campo de las artes (teatro)); se menciona también que criterios de la evaluación tomen más la parte cualitativa que la cuantitativa y por periodos más largos.

Por otra parte, las otras áreas mencionan que otra forma que complementaría el incentivo actual sería diferenciar las formas de incentivos para investigadores que tienen que realizar actividades de investigación versus lo que se encuentran comprometidos con docencia directamente y que también realizan investigación. Esto último, porque mientras los primeros se encuentran en instituciones que sólo están enfocadas hacia la investigación científica, los segundos tienen que hacer

labores de docencia como parte de su compromiso laboral. Por supuesto que esta sugerencia no es cómo se estructura el SNI actualmente, pero es importante mencionar las preocupaciones de los investigadores sobre sus labores que finalmente afectan su productividad académica.

Figura 14. Pregunta 1-4: ¿Reglas claras en el SNI? (Porcentaje de respuesta)



Notas: Cálculos por el autor.

Habiendo analizado pregunta por pregunta resulta interesante analizarla estas como si fueran las cuatro una sola. Así se toma una observación por investigador que resume la posición del mismo en las cuatro respuestas que dio. En este caso, las respuestas por parte de los investigadores ponen en relevancia un punto central de los resultados del estudio. Este punto es con respecto a los criterios de evaluación. Es decir, los investigadores perciben que no se tienen reglas claras por parte del SNI al momento de adjudicar una distinción. En este caso, tanto investigadores que resultaron con la distinción que desearon obtener previa a la

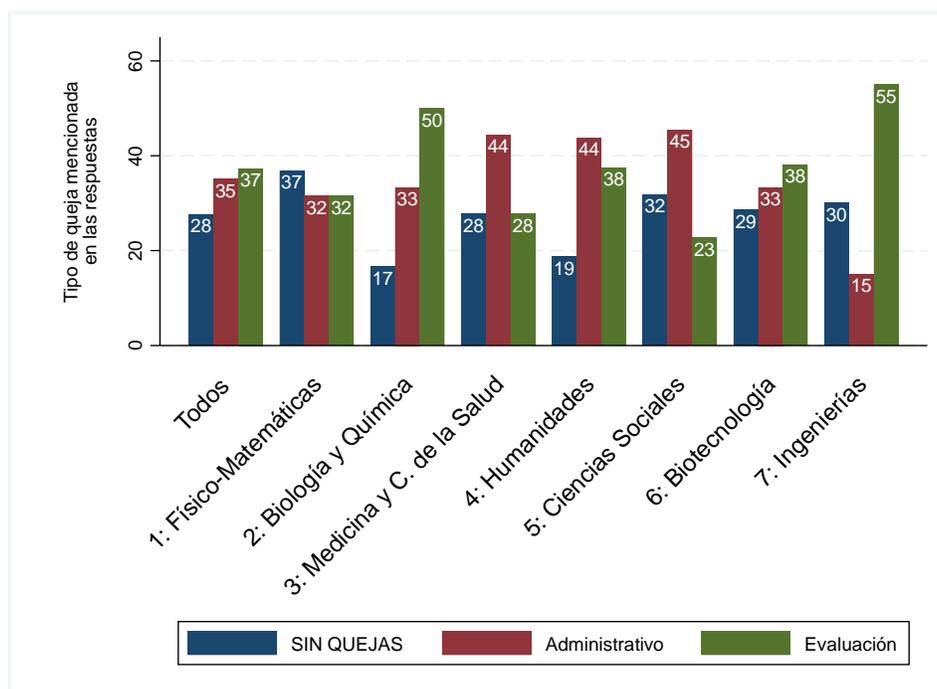
evaluación, como a los que se mantuvieron en el mismo nivel, emitieron que las reglas del SNI no son claras. Toda esta recopilación se debe a muchas situaciones particulares de cada investigador. Todo lo mencionado se muestra en la figura 14. Prácticamente la mitad de los investigadores piensa que las reglas son claras y la otra mitad que no son claras. Adicionalmente, la figura refleja que existen algunas áreas en donde los investigadores percibieron que cuando fueron evaluados las reglas no fueron claras. Estas áreas son la de Biología, Química y Ciencias de la Vida y la de Humanidades. Por su parte, el área con mayor predominancia de respuesta positiva con respecto a las reglas de evaluación es la de Ciencias Sociales. En ella, prácticamente el 68% de investigadores mostraron que las reglas son claras al momento en que se evalúa. Es posible que esto se deba a la predominancia de ciertos productos académicos en determinadas áreas de conocimiento. En ciencias sociales es más común artículos académicos en revistas especializadas, mientras que en humanidades es el libro, a su vez es más fácil tener un análisis bibliométrico de un artículo académico (número de veces de descarga, citas, vistas, etc) que de un libro, por lo que evaluación se vuelve más cualitativa.

Otro punto importante son las quejas por parte de los investigadores hacia el SNI como programa y hacia los criterios de evaluación para evaluar el historial académico del investigador. En este contexto se toman tres identificadores: (1) una queja administrativa; (2) una queja para los criterios de evaluación; y (3) no hay presencia de ninguna queja por parte del investigador. El primer aspecto corresponde a la valoración Administrativa y esta hace referencia a muchos aspectos del SNI como programa, tanto en su objetivo hacia la investigación científica como a su forma de recolectar información. En este sentido se abren

diferentes menciones de los investigadores como, por ejemplo, que “el objetivo del programa se ha desviado” y no está contemplando lo que debería ser, un programa que incentiva por impacto y lo acumulado de la investigación. Otro comentario que se toma como una valoración Administrativa es el sistema electrónico, el cual se menciona “no es amigable e incluso es confuso” e implica invertir demasiado tiempo en el llenado del historial académico; además también se toma como valoración Administrativa a los comentarios en donde se menciona que el pago o beca que realiza el SNI es insuficiente o preocupación sobre de qué manera el SNI apoyará económicamente para cuando se retiren del servicio académico o se jubilen. El segundo aspecto toma en relevancia la manera de evaluar a los investigadores, sobre todo se pone énfasis en los criterios de evaluación. Este aspecto hace referencia en que los criterios de evaluación “no siempre son los adecuados para investigadores que se encuentra en otros estados” que los que se encuentran en la capital; también en este aspecto se presenta preocupación “por la forma de evaluación de la productividad de investigadores que se encuentran en universidades con deficiencias en infraestructura”, lo cual no permite un desarrollo adecuado de la labor de investigación; asimismo, se hace referencia que la evaluación tiene un “favorecimiento con respecto a una preocupación mayor por la cantidad y no por la calidad”; además se toman los comentarios que corresponden a aquellos investigadores que cumplieron con los requisitos de evaluación pero aún así resultaron con un dictamen de evaluación desfavorable. Todo lo anterior se ve reflejado en la figura 15. Casi el 30% de los investigadores no presentan quejas ni de administrativo ni de evaluación. Luego el 35% presenta quejas de tipo administrativo, y 37% de tipo de evaluación. La figura muestra que para las áreas

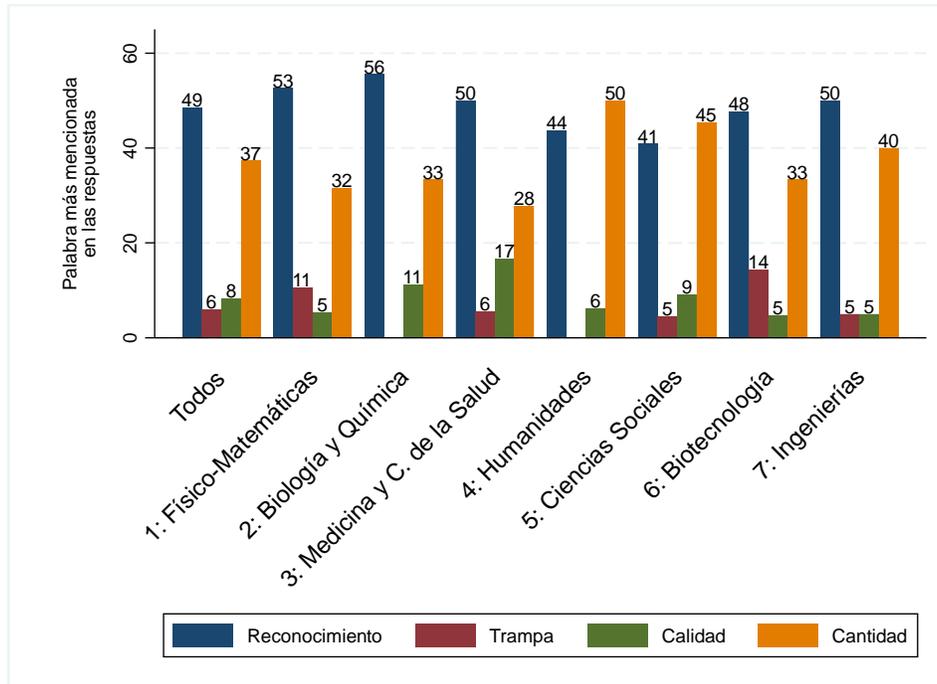
dos, seis y siete la queja principal es con respecto a los criterios de evaluación, por su parte para las áreas tres, cuatro y cinco predomina la queja administrativa. Por su parte, solo el área de Físico-Matemáticas y Ciencias de la tierra tienen una mayor presencia de investigadores que no se quejan.

**Figura 15. Pregunta 1-4: Tipo de queja mencionada en las respuestas.
(Porcentaje de respuesta)**



Notas: Cálculos por el autor.

**Figura 16. Pregunta 1-4: Palabra más mencionada en las respuestas
(Porcentaje de respuestas)**



Notas: Cálculos por el autor.

En el siguiente punto de este análisis conjunto de las cuatro preguntas resulta interesante identificar las palabras o el contexto más común de las cuatro respuestas a las cuatro preguntas del cuestionario. En este aspecto, se identificaron cuatro patrones de repetición de palabras o contexto: (1) reconocimiento: refiriéndose a esta palabra como la acción de reconocer la labor científica del investigador; (2) simulación o trampa: refiriéndose a la percepción que tienen los investigadores con respecto a sus colegas, en el aspecto que simulan a la hora de juntar el historial académico para la evaluación, es decir, se llenan de artículos por el hecho de que tienen un colega amigo que les puede poner sus nombres en un artículo sin haber participado en él o por el otro caso de que puede existir compadrazgo entre colegas; (3) calidad: refiriendo a aspectos positivos y negativos

de la prioridad que el SNI refleja con respecto a la investigación científica, si es de interés tener menor calidad que cantidad o en el caso de mantener un equilibrio en ambas partes; y (4) cantidad: refiriendo a aspectos positivos y negativos de la presión que realiza el SNI hacia los investigadores por producir el historial académico adecuado o incluso más para Niveles de distinciones altos, todo ello para cuando se realice la evaluación. Todo lo anterior se representa en la figura 16. En esta se muestra que en las áreas uno, dos, tres, seis, y siete tienen un mayor uso común de la palabra Reconocimiento, mientras que para las áreas cuatro y cinco la palabra con mayor valoración es la de Cantidad. Asimismo, la palabra Trampa o simulación solamente no se menciona en las áreas dos y cuatro. También la palabra Calidad presenta una baja percepción en el contexto de las respuestas para todas las áreas. Por su parte, cuando se analizan todas las áreas como un solo conjunto, se tiene una distribución aproximadamente mayoritaria para las palabras de Reconocimiento y Cantidad.

En resumen, la figura anterior presenta más aspectos positivos que negativos sobre los incentivos que genera el SNI. Primero, el valor que los investigadores otorgan al reconocimiento que brinda ser parte del sistema es alto. Prácticamente el 50% de todos los investigadores mencionan reconocimiento como un factor importante. Segundo, si bien algunos investigadores mencionan que existe simulación por parte de otros investigadores, el porcentaje que lo menciona es relativamente bajo (6%). Tercero, la preocupación de los investigadores no está tanto en calidad de los productos académicos, sino en la cantidad. Más de una tercera parte de los investigadores sienten que se pone mucho énfasis en cantidad, siendo el área de humanidades la que más lo menciona.

De manera complementaria al análisis cualitativo de las respuestas de la encuesta se tiene la tabla 10. Esta tabla muestra las respuestas más comunes a las preguntas realizadas. Y confirman lo mostrado en las figuras del 9 al 13. En donde el SNI representa: un estímulo económico, una motivación para seguir en la investigación, una experiencia buena o mala cuando se evalúa, que resulta ser una forma adecuada o no de incentivar la producción académica.

Todo lo anterior se puede resumir en dos palabras muy importantes: reconocimiento y productividad. Esto representa que el Programa presupuestario S191, en su trayectoria de más de 30 años, se establece como un reconocimiento para la mayoría de los investigadores vigentes registrados, lo cual produce una motivación hacia un emprendimiento de mayor productividad académica o investigación científica.

Tabla 10. Respuestas más comunes

Pregunta	Respuestas más comunes	
	Positivas	Negativas
¿Qué significa el SNI para Usted?	<p>El SNI significa un reconocimiento del trabajo que realizo diariamente, tanto en académico y científico; otro aspecto muy importante es el estímulo económico ya que los investigadores de instituciones públicas tenemos salarios modestos.</p> <p>Es una instancia que apoya a los investigadores nacionales con una distinción académica de alto impacto que sirve como indicador de la productividad, que además representa un estímulo económico para motivar al investigador.</p>	<p>Un aparato de Estado que, tras la careta del “apoyo” a los estudiosos que cumplen con ciertos requisitos, los sofoca cada día más, cargándolos de corsés, y cadenas burocráticas que, en lugar de facilitarles la labor, la obstaculizan de forma onerosa, insoportable, estresante, injusta. La última de esas piedras se llama CVU nuevo, una verdadera monserga y hoyo negro, al parecer diseñado para que quien no lo haya llenado sea excluido de manera sumaria del SIN (“sin”, gran traducción automática de la computadora de unas siglas que deberían indicar otra cosa).</p> <p>A mí el cambio de plataforma me tocó estando de sabático. No tiene usted idea de lo que significó para mí en tiempo perdido miserablemente, con resultados infructuosos; por si fuera poco, el SIN me negó la beca para estancias académicas en el extranjero a la que tengo derecho con sobradas razones y documentos probatorios luego de muchos años de pertenencia en el SIN. La respuesta fue desalentadora: “que no había enviado bien los papeles”. Como si no los tuvieran en mi expediente. Yo estuve en Alemania, en condiciones precarias, y aun así escribí un libro y comencé otro, en condiciones de sobrevivencia dada la carestía y sin esa beca que me negaron arbitrariamente. Eso significa para mí el SIN, estimado evaluador.</p>
¿Cómo fue su experiencia cuando fue evaluado por el SNI, para su reciente nombramiento?	<p>Fue una experiencia positiva. Me pareció importante que reconocieran mi trabajo del periodo y mi trayectoria académica y profesional.</p>	<p>Mala; mostré publicaciones de calidad y una transferencia tecnológica y me bajaron al nivel I, lo cual me enfureció por lo que apelé, pero no hubo manera. Muy muy injusto. Cómo son públicos los nombres de los evaluadores ya se quién me evaluó y en este año que me toco renovar pedí que dos de ellos no me evaluaran por conflicto de intereses y lo fundamenté</p>
¿Cómo siente que el SNI ha influenciado su productividad académica?	<p>Positivamente, creo que me ha forzado a incrementar mi productividad. Sin embargo, estimo que he desdeñado cosas útiles por lo cuantitativo.</p>	<p>En mi caso el SNI no ha representado un motor en mi carrera, siempre me ha motivado el querer hacer. Claro que el aporte económico por el SNI es importante. Sin embargo, el deseo de producir ha sido motivado solo por mi gusto y pasión por el trabajo.</p>

¿Considera que la forma en como incentiva el SNI la investigación es la adecuada? ¿Si o No? ¿Por qué?

Si, 1. ser un referente de evaluación de mi producción, lo cual me parece bien, porque ayuda a tener presiones externas a la producción. Digamos a mantener un ritmo constante de la docencia, investigación y publicación. Este es un estímulo importante, para quienes necesitamos presiones externas y de retroalimentación comunitaria. 2. ser un referente adscriptivo de una comunidad científica nacional, como ámbito de reconocimiento. Es un factor subjetivo, pero a mi parecer sumamente importante en el ejercicio profesional y académico. 3. ser una fuente de incremento a los ingresos como investigadores vía salarios. Obtener más recursos monetarios por el mismo trabajo por el que me pagan como profesora-investigadora constituye un estímulo económico sumamente importante.

Quando fui estudiante pensé que sí. Ahora veo que no, pues hay 'mafias' que solo quieren mantener o subir el nivel a toda costa sin realizar trabajo académico. En México, todos los investigadores deberíamos impartir clase, hacer divulgación y colaborar en el trabajo Colegiado y en el desarrollo de nuestras instituciones, sin embargo, esas actividades están castigadas en el SNI. Si de alguna manera y supuestamente los SNI II y III son los 'más calificados' (además de los mejor pagados y los que más estímulos económicos reciben), por qué no se les 'incentiva y vigila' para que realmente contribuyan con su 'experticia' a la formación de investigadores, el fortalecimiento institucional y la solución de problemas nacionales, en lugar de solo tenerlos en una 'esfera' para que publiquen 'lo que sea' y gradúen 'a toda costa' a cualquier 'estudiante'...todo sea por los puntitos. Aclaro hay excepciones, pero la regla es otra. Penosamente veo que muchos colegas terminan escribiendo las tesis y los artículos de sus estudiantes ¿Nos pagan para formar investigadores y solucionar problemas o, para juntar puntitos?

Otra debilidad del sistema de evaluación en que 'supuestamente' nos miden a todos por igual pero, hay investigadores que en sus laboratorios tienen técnicos (en muchos casos más de uno), auxiliares de investigación (en muchos casos más de uno), profesores invitados, posdocts (en muchos casos más de uno), preferencia al repartir alumnos, etc. mientras que otros ni siquiera tenemos laboratorio (apenas me entregaron mi laboratorio hace aproximadamente dos o tres meses, y en diciembre de este año cumpliré ocho años en Cinvestav) y apenas hace unos cinco años tuve accesos a estudiantes de posgrado porque como indiqué, cuando me contrataron, por tres años no tuve adscripción a posgrado alguno...hasta que aprobaron los posgrados que creamos. Pero eso sí, para el SNI todos somos, supuestamente, evaluados 'igual'. Claro, conseguir ese laboratorio me constó un gran problema porque reclamé al director que tenía muchos equipos, cristalería y reactivos en cajas y ya ni siquiera las cajas cabían en mi oficina. Creo que ahora quien visite el laboratorio, no creará que tiene tres meses, pero ahí está el recurso invertido de varios proyectos que he conseguido y finiquitado aún sin laboratorio y con escasos estudiantes, varios de ellos de ingeniería y TSU.

Nota: Las respuestas más comunes en los aspectos negativos y positivos.

Finalmente, es relevante detallar algunas particularidades de la encuesta, aunque resulten ser muy interesantes los resultados, estos no pueden plasmarse como una generalidad de patrón para los más de 28,000 investigadores vigentes registrados en el SNI en 2018. La muestra es representativa para la muestra final de este estudio (que presentan al menos dos evaluaciones en el periodo 2005-2017), pero no para el universo de investigadores asociados al sistema.

No obstante, resulta interesante comparar la encuesta realizada con aquellas que tengan el mismo tipo de características. Y precisamente en el 2015, el Foro Consultivo Científico y Tecnológico (FCCyT) realizó un trabajo de investigación similar que se denomina "*Encuesta a los Investigadores en el SNI*". El objetivo de la

investigación era determinar el peso que tienen las actividades más representativas de los investigadores en su quehacer diario, conocer su nivel de participación en los programas de apoyo a la investigación e identificar si establecen relaciones de colaboración para realizar proyectos de investigación, así como sus posturas con respecto a las iniciativas de acceso abierto que existen actualmente (Rodríguez, 2015a,b,c).

En este contexto, el tópico de la investigación del FCCyT, realizado por Rodríguez (2015c), con respecto a Actividades de los investigadores y apoyos a la investigación es la que más compagina con el presente análisis cualitativo de la investigación. En el presente trabajo y en el documento de trabajo de FCCyT (Rodríguez, 2015c) se perciben que la mayor parte de investigadores priorizan más el trabajo en investigación que el de docencia. Además, se percibe que hay una relación positiva entre la productividad en investigación y el ascenso en distinción del investigador. Por otra parte, en ambos documentos de investigación se presencia el obstáculo de no ser claros en los criterios de evaluación. Por último, un aspecto muy relevante en ambos documentos de investigación son los recursos económicos. En ambos se menciona que el obstáculo para realizar investigación es la falta de recursos económicos. Por ende, en muchos casos del análisis cualitativo de la presente investigación se menciona que la característica principal para seguir haciendo investigación es por el apoyo en un estímulo económico realizado por el SNI.

En conclusión, con esta subsección se realizó un análisis de sentimientos de las respuestas de los investigadores al cuestionario. En ellas se cuantificaron temas de particular interés para el presente trabajo de investigación, la cual complementa el análisis cuantitativo. En los resultados de esta subsección se han obtenido que el

49% de investigadores que respondieron la encuesta (134) presencian en el SNI un patrón de reconocimiento por la labor científica realizada. Asimismo, el 57% de investigadores muestran que para ellos el SNI significa un estímulo económico y además el 68% de investigadores el SNI le representa una motivación para seguir en la investigación científica. Sin embargo, el 37% de investigadores presenta quejas sobre los criterios de evaluación, siendo sus comentarios al respecto de que estos no son los adecuados al tener la percepción de que no se prioriza la calidad de la investigación. Y como otra parte importante, un poco más de 70 investigadores de los que respondieron la encuesta tienen la edad mayor a 60 años y una preocupación frecuente en este grupo de edad es perder el SNI en la jubilación, sugiriendo la posibilidad de un esquema que permita una estabilidad de ingresos a futuro (después de la jubilación). En resumen, el SNI significa reconocimiento, productividad, y se mencionan más aspectos positivos que negativos, sin dejar pasar algunos mecanismos potenciales para su mejora.

12. Retos y avenidas de mejora usando ambas evaluaciones

En esta sección se discuten brevemente los resultados encontrados, se relacionan los resultados de ambos tipos de evaluaciones, y se señalan retos y avenidas de mejora.

Los resultados de la evaluación cuantitativa son claros. La productividad académica aumenta cuando los investigadores son promovidos de nivel en comparación con los que permanecen en el mismo nivel. Esto se da particularmente en artículos de revistas y libros. Cuando se analizan los resultados

por área de conocimiento, se tienen efectos heterogéneos. Por ejemplo, para humanidades son más importante los libros que otros aspectos de la producción académica.

Un aspecto a resaltar es que el impacto en número de artículos disminuye cuando se incluye la ponderación por el factor de impacto. Esto se da en todas las áreas con excepción de ciencias sociales. Esto quiere decir que el efecto de la promoción del SNI en revistas de alto impacto no es tan alto como en las de sin impacto. Esta disminución se da más en las áreas 1-3 que en otras áreas. Este resultado es consistente con la evaluación cualitativa donde varios investigadores mencionaron que la promoción del SNI tiene incentivos a enfocarse en la cantidad de artículos no en la calidad. Adicionalmente, el enfoque en la cantidad podría llevar a que exista “simulación” donde algunos autores de un artículo no contribuyan a su realización (varios investigadores señalaron esto en la evaluación cualitativa).

Algunas recomendaciones concretas y/o retos para el futuro con la información cualitativa recabada (se toman en cuenta todos los comentarios de los investigadores a pesar que en algunos de ellos no es atribución directa del SNI):

- Continuar con la labor de divulgación que realizan el CONACYT y el SNI para resaltar el prestigio y orgullo que conlleva ser investigador SNI. Muchos investigadores han mencionado que esto es un aspecto fundamental en el programa. Por tanto, si CONACYT pudiera aumentar la visibilidad del programa en la sociedad, su importancia y relevancia, es posible que esto aumente la cantidad y/o calidad de la producción académica.

- En economía conductual se le conoce como “empujón” a una intervención que facilita la toma de decisiones hacia un óptimo social. Por ejemplo, un “empujón” en pensiones podría ser realizar el ahorro en la vida de forma más sencilla. El sector de CTI tiene margen para poder utilizar “empujones” para que los investigadores tomen decisiones a favor del óptimo social. Uno ejemplo de ellos sería enviar un correo electrónico a los investigadores con los lineamientos de la evaluación de pares. Se podría informar también qué tipo de producción académica se tuvo en años anteriores para brindar mayor información. Esto atendería dos tipos de quejas que se presentaron en la evaluación cualitativa. Por uno, se podría disminuir la queja de que los criterios no son claros. Por otro, incluso podría ayudar a que los investigadores no señalen que existe “simulación” en la evaluación de pares, como se menciona en algunos comentarios de la evaluación cualitativa.
- Un reto importante para el futuro es la idoneidad del tiempo de evaluación para cada investigador. En la evaluación cualitativa se mencionó que un aspecto negativo de la evaluación es que no es posible emprender proyectos de investigación de largo plazo. Si se llegan a cambiar los plazos se recomienda ser explícitos en los compromisos de publicaciones académicas que se esperan.
- Otro reto mencionado en la evaluación cualitativa es el tema del retiro y el ingreso monetario del SNI. El 35% de los investigadores que respondieron la encuesta están preocupados para no salir del SNI especialmente en edad avanzada, pues el ingreso del SNI es un fuerte complemento a su salario

mensual. Y con ello, la falta de jubilación podría implicar que jóvenes científicos no puedan participar o incorporarse al sistema universitario como investigadores. También algunos investigadores mencionaron que sería benéfico que el SNI se tome en el salario para que cuente al ahorro para el retiro. Este aspecto no es posible implementar con la normatividad vigente. Pero es posible diseñar un mecanismo que permita a los investigadores escoger voluntariamente si desean depositar una cantidad determinada de su estímulo SNI en su Afore de forma automática.

- Otro reto importante es cómo promover la investigación científica en universidades con menos recursos que las “grandes” universidades y centros de investigación. El 37% de investigadores mencionaron las carencias de recursos económicos, infraestructura, así como el tiempo invertido en la formación de capital humano y labores administrativas.

13. Conclusiones y recomendaciones

El Consejo Nacional de Ciencia y Tecnología (CONACYT) es el encargado de operar el Programa Presupuestario S191-Sistema Nacional de Investigadores (SNI). El SNI es un sistema de incentivos económicos y no económicos por nombramiento (Nivel de distinción) para la investigación. Actualmente (2018) existen casi 30,000 mil investigadores registrados en el SNI. El programa es una de las políticas públicas más importantes para el apoyo a la investigación científica.

En este contexto, resalta la importancia de tener una evaluación de impacto para el programa. Ahora bien, la política pública en Ciencia, Tecnología e

Innovación resulta complejo de evaluar. En particular, no es posible evaluar el SNI integralmente en todas las variables deseadas como impacto social y producción académica. Esto simplemente porque no es posible incluir todos los indicadores posibles de forma cuantitativa. En este trabajo solo se evaluó cómo la estructura del SNI, y en particular la promoción del nivel 1 al 2 o bien del 2 al 3, fomenta la investigación científica. Donde productividad académica se mide únicamente con número de producción académica en artículos de revista, capítulos de libros y libros.

A pesar de esas limitaciones, esta evaluación es valiosa e importante. La evaluación cuantitativa considera variables de producción académica que representan un subconjunto de la información contenida en la evaluación de pares, de acuerdo con los criterios de evaluación respectivos a las siete áreas de conocimiento del Programa presupuestario S191. Por tanto, se usa información que los evaluadores académicos toman en cuenta para decidir sobre una promoción o permanencia de nivel. Y esa decisión de promoción o permanencia tiene consecuencias futuras sobre el comportamiento de los investigadores en el corto plazo acerca de su producción académica.

En este trabajo se implementan metodologías cuantitativas y cualitativas. En lo cuantitativo, se utilizan diferentes estrategias como son regresión lineal con extensos controles, así como métodos de diferencias en diferencias con aparejamiento en puntajes. En lo cualitativo, se enviaron encuestas con preguntas abiertas a aproximadamente 800 investigadores para conocer lo que significa el SNI y su influencia en la producción académica.

Usando diferentes métodos econométricos, en general se encuentra que subir un nivel en el SNI impacta positivamente en la producción académica. El efecto

para los diferentes productos varía, pero en general está entre 5-10% de una desviación estándar. Si bien este efecto no es sustancialmente grande, sí es parecido a lo encontrado en otro tipo de investigaciones de incentivos monetarios y aprovechamiento académico (Fryer 2016). Los resultados indican que en general el impacto es mayor en número de artículos que en artículos con factor de impacto. Es decir, no se aumenta tanto la producción de artículos con factor de impacto al subir un nivel en el SNI como en el impacto de número de artículos. En resultados descriptivos se tiene que la gran mayoría de artículos producidos no tienen factor de impacto sus revistas (el 78% en artículos del periodo pre-evaluación y el 73% en artículos post-evaluación). Entonces el investigador típico (en la mediana) produce cada periodo de evaluación 5 artículos de investigación, ninguno o si acaso 1 con factor de impacto, y 3-4 capítulos de libro.

Los resultados cualitativos señalan que en general el SNI es motivo de reconocimiento. Lo cual a su vez desemboca a que sea un estímulo, sin considerar la parte monetaria, para pertenecer en el sistema. A pesar de esto, también el 57% de los investigadores contactados menciona que el incentivo económico es muy importante, especialmente en investigadores en edad avanzada.¹³ También se considera en general que las evaluaciones son justas, pero muchas veces los criterios no son claros. Muchos investigadores también mencionaron que dado el periodo tan corto entre evaluaciones no se fomenta investigación de largo plazo significativa o bien buscar publicar resultados en revistas de alto factor de impacto.

¹³ El total de investigadores que respondieron la encuesta son de 134 y dentro de ellos un poco más de 70 investigadores tienen una edad mayor a 60 años y para la mayoría de ellos en su larga trayectoria el SNI le ha representado como estímulo económico.

Con base en los resultados encontrados en los métodos cuantitativos y cualitativos se pueden obtener potenciales avenidas de mejoras y retos para el SNI. Primero, resulta importante igualar el impacto de número de artículos con el de número de artículos en revistas con factor de impacto. Esto se podría lograr si se puede hacer más saliente los lineamientos para la promoción dentro del SNI, así como considerar la idoneidad del periodo de vigencia de cada nivel SNI. Estos cambios podrían disminuir las quejas que se tienen del 37% de investigadores sobre el enfoque en la cantidad de artículos y no en la calidad, así como las quejas de que existe “simulación” de los investigadores al incluirse como autores en trabajos que no contribuyeron. Segundo, otra preocupación real para investigadores en edad avanzada es el tema del retiro y dejar obtener el ingreso del SNI. Para generaciones más jóvenes, y con el objetivo de evitar este problema nuevamente en el futuro, es posible diseñar un mecanismo que permita a los investigadores escoger voluntariamente si desean depositar una cantidad determinada de su estímulo SNI en su Afore de forma automática. Finalmente, el posicionar la labor científica en la sociedad siempre será un reto importante. Sin embargo, para la mayoría de los investigadores el prestigio y el orgullo de ser parte del SNI es el mayor aliciente a la producción académica.

Para finalizar, el proceso de evaluación continua del programa debe seguir. Algunas recomendaciones mencionadas aquí deben evaluarse para mostrar si tuvieron efecto y si conviene en términos costo-efectividad y en los objetivos del programa. Asimismo, sería benéfico ampliar esta evaluación en el futuro para considerar más indicadores de la producción académica como son citas de artículos

académicos, visibilidad de la investigación (por ejemplo, Altmetrics), presencia en la discusión nacional, entre otros. Esta información complementaría lo encontrado en este estudio y describiría con mayor precisión el panorama de la producción científica en México.

Referencias bibliográficas

- Abramo, G. & D'Angelo, C. A. (2011). Evaluating research: From informed peer review to bibliometrics. *Scientometrics*, 87(3), 499–514.
- Belloni, A., Chernozhukov, V. & Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28 (2): 29-50.
- Bensusán, G. & Valenti, G. (2018). La evaluación de los académicos. Instituciones y Sistema Nacional de Investigadores, aciertos y controversias. *FLACSO México*.
- Bollen, J., Van de Sompel, H., Hagberg, A. & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE*, 4(6).
- Bukstein, D. & Gandelman, N. (2016). Glass Ceiling in Research: Evidence from a National Program in Uruguay. *IDB Working Paper Series IDB-WP-798*.
- Carpenter, C. R., Cone, D. C. & Sarli, C. C. (2014). Using publication metrics to highlight academic productivity and research impact. *Academic Emergency Medicine*, 21(10), 1160–1172.
- Consejo Nacional de Ciencia y Tecnología [CONACyT]. (2017). Reglamento del Sistema Nacional de Investigadores. México. *Recuperado de <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/marco-legal>*
- Costas, R., Leeuwen, T. N. Van. & Bordins, M. (2010). A Bibliometric Classificatory Approach for the Study and Assessment of Research Performance at the Individual Level: The Effects of Age on Productivity and Impact. *Journal of the American Society for Information Science*, 61(8), 1564–1581.
- Dávalos-Sotelo, R. (2015). Una forma de evaluar el impacto de la investigación científica. *Madera y Bosques*, 21(especial), 7–16.
- Didegah, F. & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, 7(4), 861–873.
- Dybå, T., Kampenes, V. B., & Sjøberg, D. I. K. (2006). A systematic review of statistical power in software engineering experiments. *Information and Software Technology*, 48(8), 745–755.
- Eyre-Walker, A. & Stoletzki, N. (2013). The Assessment of Science: The Relative Merits of Post-Publication Review, the Impact Factor, and the Number of Citations. *PLoS Biology*, 11(10).

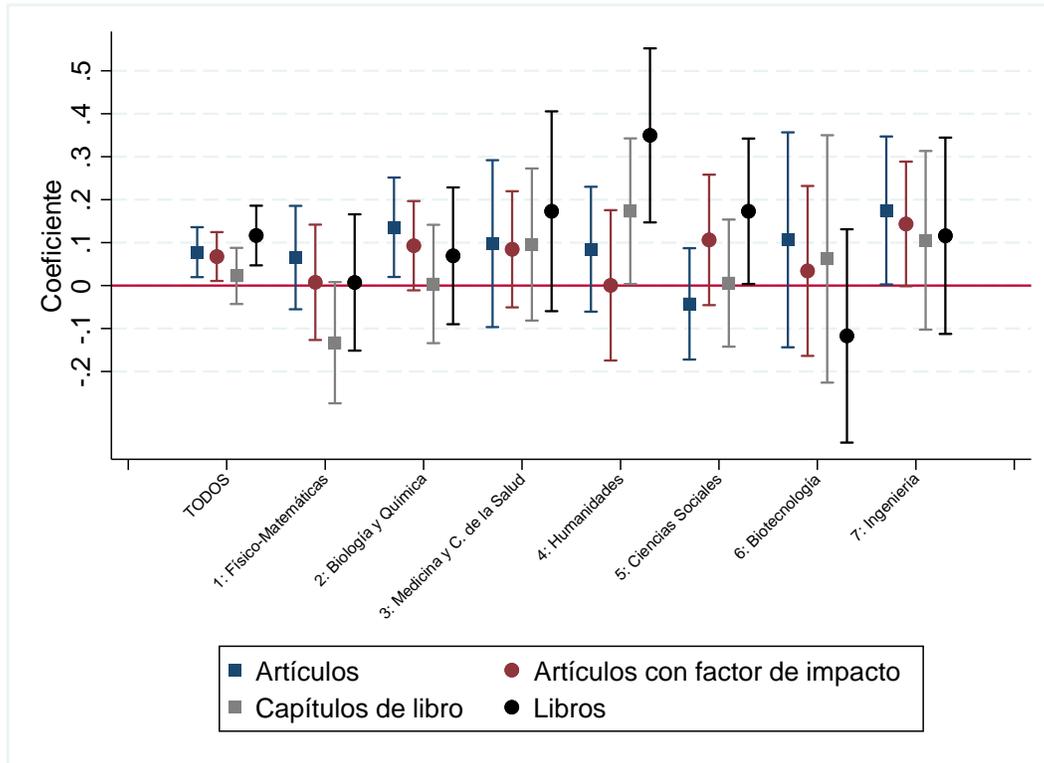
- Foro Consultivo Científico y Tecnológico [FCCyT] & Academia Mexicana de Ciencias [ACM]. (2005). Una reflexión sobre el Sistema Nacional de Investigadores a 20 años de su creación. México. *Documento de trabajo*. Disponible en: http://www.foroconsultivo.org.mx/libros_editados/20_sni.pdf
- Foro Consultivo Científico y Tecnológico [FCCyT]. (2015). Reporte: Reflexiones sobre la evaluación a investigadores: una mirada desde diferentes perspectivas. *Documento de trabajo*. Disponible en: http://www.foroconsultivo.org.mx/libros_editados/evaluacion_de_la_evaluacion_subgrupos_individuos.pdf
- Frixione, E., Ruiz-Zamarripa, L., & Hernández, G. (2016). Assessing individual intellectual output in scientific research: Mexico's national system for evaluating scholars performance in the humanities and the behavioral sciences. *PLoS ONE*, 11(5), 1–28.
- Fryer, Roland (2016). "The production of human capital in developed countries: Evidence from 196 randomized field experiments". En *Handbook of Field Experiments* editado por Abhijit Banerjee y Esther Duflo, Capítulo 12. Por publicarse.
- Gentzkow, M., Kelly, B., & Taddy, M. (2017). Text as data. *National Bureau of Economic Research*, No. 23276.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2017). Impact Evaluation in Practice. *The World Bank Publications*.
- Gesto, N. (2016). Análisis de Poder Estadístico y su Aplicación a Evaluaciones Experimentales. *Maestría en Ingeniería Matemática – Udelar*.
- Gil, M. & Contreras, L. (2017). El Sistema Nacional de Investigadores: ¿espejo y modelo? *Revista de La Educación Superior*, 46(184), 1–19.
- Gonzales-Brambila, C. & Veloso, F. (2007). The Determinants of Research Productivity: A Study of Mexican Researchers. *Research Policy*, 36, 1–54.
- Hernández, A. & Jiménez, M. (2017). Evolución de las Políticas del Sistema Nacional de Investigadores y su impacto en la trayectoria de los investigadores. *Debates en Evaluación y Curriculum, Congreso Internacional de Educación*, 3(3), 306-319.
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431.
- Honig, D. (2018). Case study design and analysis as a complementary empirical strategy to econometric analysis in the study of public agencies: deploying mutually supportive mixed methods. *Journal of Public Administration Research and Theory*, por publicarse.

- Jänicke, S., Franzini, G., Cheema, M., & Scheuermann, G. (2016). Visual text analysis in digital humanities. *Computer Graphics Forum*, 36(6), 226-250.
- Lin, S. W., & Chen, S. C. (2009). PSOLDA: A particle swarm optimization approach for enhancing classification accuracy rate of linear discriminant analysis. *Applied Soft Computing Journal*, 9(3), 1008–1015.
- Lozano, J., Saavedra, R., & Fernández, N. (2011). La evaluación del impacto de los resultados científicos. Metodologías y niveles de análisis. *Humanidades Médicas*, 11(1), 99–117.
- Mele, V. & Belardinelli, P. (2018). Mixed methods in public administration research: selecting, sequencing, and connecting. *Journal of Public Administration Research and Theory*, por publicarse.
- Ortiz, E. A. (2015). La evaluación del impacto científico en las investigaciones educativas a través de un estudio de caso. *Revista Electrónica de Investigación Educativa*, 17(172), 89–100.
- Ortiz, E., González, M. V., Infante, I. & Viamontes, Y. (2010). Evaluación del impacto científico de las tesis doctorales en Ciencias Pedagógicas mediante indicadores cuantitativos. *Revista Española de Documentación Científica*, 33(2), 279–286.
- Pomeranz, D. (2017). Impact Evaluation Methods in Public Economics. *Public Finance Review*, 45(1), 10–43.
- Pontille, D. & Torny, D. (2010). The controversial policies of journal ratings: Evaluating social sciences and humanities. *Research Evaluation*, 19(5), 347–360.
- Reyes-Ruiz, G. & Suriñach-Caralt, J. (2012). Un análisis crítico sobre las evaluaciones internas del SNI de México a partir de distintas técnicas de análisis de datos. *Memorias XVII Congreso Internacional de Contaduría, Administración e Informática. México: Universidad Nacional Autónoma de México (Facultad de Contaduría y Administración), octubre 3-5, 2012.*
- Reyes-Ruiz, G. & Suriñach-Caralt, J. (2015). Análisis sobre la Evolución del Sistema Nacional de Investigadores (SNI) de México. *Investigación Administrativa*, 44 (115), 55-69.
- Reyes-Ruiz, G. (2018). Las evaluaciones internas del sistema nacional de investigadores de México a través de un análisis clúster. *Revista Actualidades Investigativas en Educación*, 18(1), 1-32.
- Rodríguez, Carlos. (2015a). Encuesta a los investigadores en el SNI 2015. Módulo: Acceso abierto a la información científica. *Foro Consultivo Científico y Tecnológico*.
- Rodríguez, Carlos. (2015b). Encuesta a los investigadores en el SNI 2015. Módulo: Colaboración para la investigación. *Foro Consultivo Científico y Tecnológico*.

- Rodríguez, Carlos. (2015c). Encuesta a los investigadores en el SNI 2015. Módulo: Actividades de los investigadores y apoyos a la investigación. *Foro Consultivo Científico y Tecnológico*.
- Saavedra-Fernández, O., Quinteros-Carrillo, P., Saldaña-González, I., Rangel-Hernández, S., Zurita-Gómez, A. y Guzmán-Sánchez, M. V. (2005). El uso de los indicadores cualitativos y cuantitativos en la evaluación de las publicaciones periódicas: el caso del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional de México. *Revista Española de Documentación Científica*, 28(4), 500–517.
- Sánchez, G. (2010). 25 años del Sistema Nacional de Investigadores (SNI) y su impacto en las Políticas de Evaluación de la Educación Superior en México. *Revista*, 5(13), 219-236.
- Seglen, P. O. (1998). Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthopaedica Scandinavica*, 69(3), 224–229.
- Silva, M. P., García, V. & Aquino, S. P. (2016). Retos del crecimiento del Sistema Nacional de Investigadores (SNI) del Consejo Nacional de Ciencia y Tecnología (CONACyT) en México. *Revista Electrónica Actualidades Investigativas En Educación*, 16(2), 24.
- Solís, F. M., Milanés, Y. & Navarrete, J. (2010). Evaluación de la investigación científica. *El caso de Andalucía. Revista Fuentes*, 10, 83–100.
- Wang D., Song C. & Barabási A.L. (2013) Quantifying Long-Term Scientific Impact. *Science*, 342(6154), 127–132.
- Wilkerson, J., & Casas, A. (2017). "Large-scale computerized text analysis in political science: opportunities and challenges." *Annual Review of Political Science*, 20, 529-544.
- Zinovyeva, N. & Bagues, M. (2015). The role of connections in academic promotions. *American Economic Journal: Applied Economics*, 7(2), 264–292.

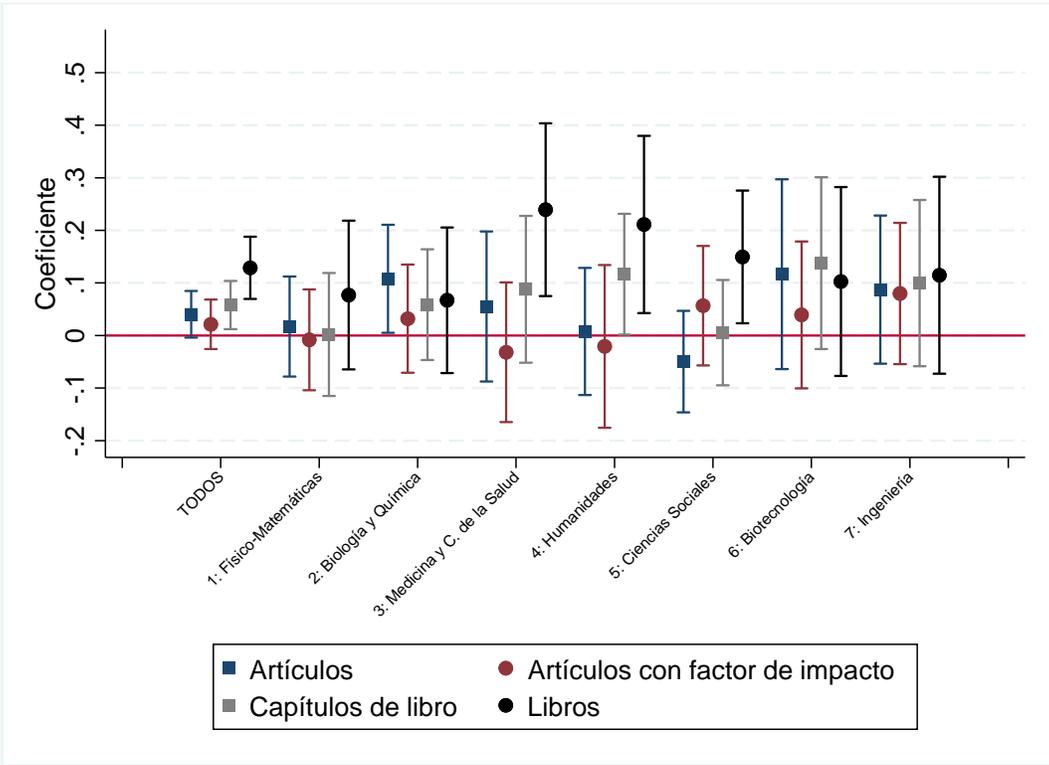
Anexos

Figura A1. Resultados de método de aparejamiento usando método 2



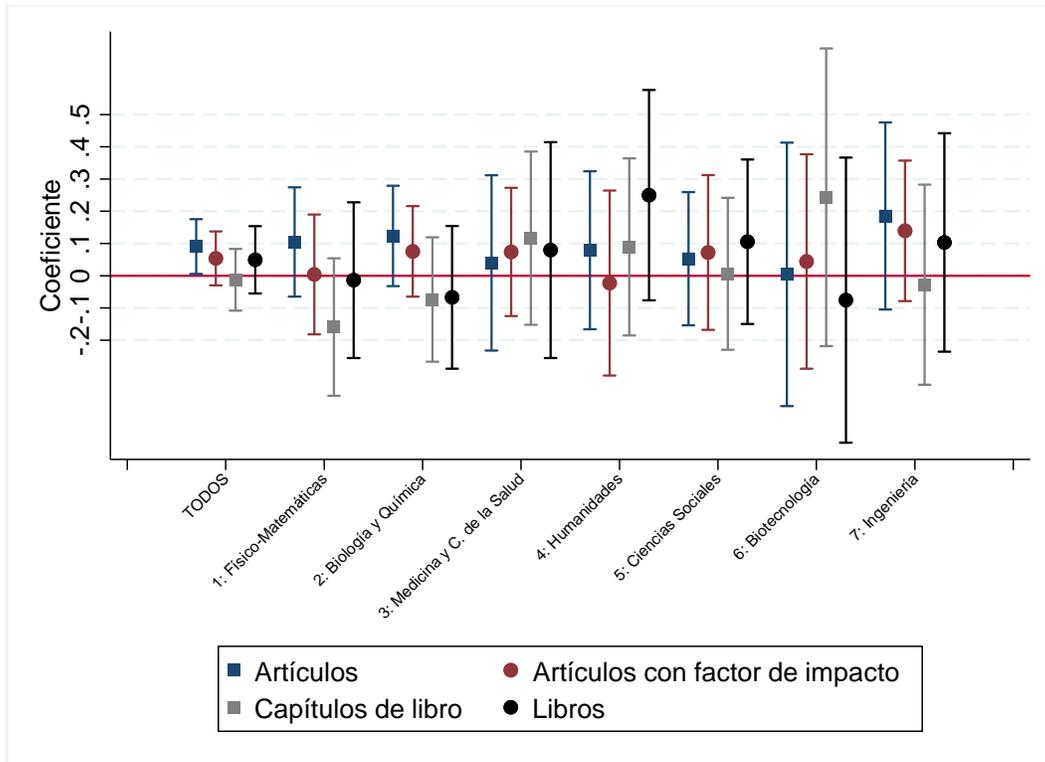
Notas: Cálculos por el autor. Método 2 estima el puntaje con un modelo logit y realiza el aparejamiento con todos los vecinos en una vecindad de 0.0005.

Figura A2. Resultados de método de aparejamiento usando método 3



Notas: Cálculos por el autor. Método 3 estima el puntaje con componentes principales y realiza el aparejamiento con 4 vecinos en una vecindad de 0.025.

Figura A3. Resultados de método de apareamiento usando método 4



Notas: Cálculos por el autor. Método 4 estima el puntaje como el logaritmo natural de la razón de probabilidades ($p/1-p$) y realiza el apareamiento con todos los vecinos en una vecindad de 0.001.