DATAVOZ.

DEL GRUPO STATCOM

# Informe Final

# "Encuesta de trayectoria Doctorados en Chile"

## Licitación ID: 1098710-1-LE20

**Diciembre 2020**

www.datavoz.cl
Europa 1935 – Providencia – Santiago – Chile
Fono: 22 5994777

# INDICE

INDICE ..... 2
1. Antecedentes ..... 4
2. Lineamientos de la OCDE ..... 6
2.1 Lineamientos metodológicos ..... 6
2.2 Cuestionario modelo y manual de instrucción. ..... 7
2.3 Reporte de resultados e indicadores ..... 8
3. Definiciones y lineamientos metodológicos ..... 9
3.1 Marco Muestral ..... 9
3.1.1 Población objetivo ..... 9
3.1.2 Fecha de referencia ..... 9
3.1.3 Estratificación ..... 9
3.2 Consideraciones del diseño muestral ..... 9
3.3 Universo estimado ..... 10
3.3.1 Directorio de profesionales con grado de doctor - Actualización 2019 ..... 10
3.3.2 Directorio de profesionales con grado de doctor de LinkedIn ..... 12
3.3.3 Concatenación de bases en un directorio combinado ..... 14
3.3.4 Definición del marco muestral ..... 14
3.3.5 Definición del marco de selección ..... 14
3.4 Tamaño de muestra ..... 15
3.5 Selección de la muestra ..... 16
4. Actividades previas a la aplicación de la encuesta ..... 19
4.1 Recolección de datos de contacto del Directorio de LinkedIn ..... 19
4.2 Validación del instrumento y medio de recolección ..... 20
4.2.1 Metodología de Validación del Cuestionario ..... 20
4.2.2 Resultados de Validación del Cuestionario ..... 22
4.3 Definición del Instrumento final ..... 28
4.4 Definición de protocolos de levantamiento ..... 29
4.4.1 Antecedentes del proceso de definición de los protocolos de levantamiento ..... 29
4.4.2 Protocolo aplicado en levantamiento CAWI ..... 30
4.4.3 Protocolo aplicado en levantamiento CATI ..... 32
5. Resultados del trabajo de campo ..... 34
5.1 Medidas de gestión operativa ..... 34
5.2 Decisión de expansión de la muestra ..... 35

5.3 Seguimiento de las encuestas por modo de recolección ..... 36
5.4 Códigos de disposición final e indicadores de respuesta ..... 39
5.5 Evolución del levantamiento ..... 40
5.6 Supervisión y acompañamiento ..... 41
6. Aspectos metodológicos asociados a la elaboración de la base de datos ..... 44
6.1 Procesamiento y validación de la Base de Datos ..... 44
6.1.1 Homologación y deduplicación ..... 44
6.1.2 Validación de la Base de Datos ..... 45
6.1 Ponderadores y factores de expansión ..... 45
6.1.1 Cálculo de ponderadores ..... 45
6.1.2 Cálculo de factores de expansión ..... 46
6.1.3 Valores calculados de ponderadores y factores de expansión ..... 46
6.2 Error muestral final ..... 46
7. Análisis de Resultados ..... 48
7.1.1 Crecimiento de la población de personas con grado de doctor ..... 48
7.1.2 Estudios cursados ..... 52
7.1.3 Inserción laboral ..... 55
7.1.4 Similitudes y diferencias entre las personas que realizaron o no realizaron postdoctorados ..... 63
7.1.5 Similitudes y diferencias relevantes entre zonas geográficas ..... 66
8. Aprendizajes y recomendaciones para futuros levantamientos ..... 68
8.1 Aprendizajes sobre la población objetivo ..... 68
8.2 Aprendizajes sobre los medios de recolección ..... 69
8.2.1 Aprendizajes de levantamiento CAWI ..... 69
8.2.2 Aprendizajes de levantamiento CAWI ..... 69
8.2.3 Aprendizajes de la coordinación entre CAWI y CATI ..... 70
8.3 Recomendaciones para futuros levantamientos ..... 70
9. ANEXO 1 Cuestionario Aplicado ..... 71
10. ANEXO 2 Cálculo de indicadores AAPOR ..... 71
11. ANEXO 3 Modelo de carta de invitación ..... 72

# 1. Antecedentes

Entre las claves más importantes para impulsar la productividad de los países se encuentra el fomento de sus capacidades de investigación, innovación, progreso tecnológico y generación de conocimiento. No obstante, en Chile, la inversión en investigación, desarrollo e innovación (I+D+i) aún es baja. Por ejemplo, según reporta la Encuesta sobre Gasto y Personal en I+D (2018) del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, sólo un 0,35% del PIB chileno se destinó a actividades de I+D, mientras que tan solo hay 1,1 investigadores(as) por cada 1.000 trabajadores(as). En contraste, el promedio de inversión como porcentaje del PIB de los países OCDE es de 2,4% y el promedio de investigadores por cada 1000 trabajadores es 8,6.

Estas brechas dan cuenta del gran desafío que enfrenta Chile para impulsar las actividades de I+D+i. Entre otras materias, para poder diseñar políticas públicas eficaces para acometer ese desafío es necesario poseer un diagnóstico claro de las condiciones de inserción laboral de personas con alta calificación, específicamente, personas con grado de doctor, el más alto grado académico. Adicionalmente, conocer las características del conjunto de personas con grado de doctor contribuye tanto a diagnosticar como a fortalecer las capacidades de Chile para aprovechar sus esfuerzos de I+D. Por este motivo, durante los años 2012 y 2015 se aplicaron en Chile las encuestas "Trayectoria de Profesionales con estudios de doctorado" (de aquí en adelante encuestas CDH). Estas encuestas tuvieron como objetivo conocer los patrones que guían las carreras y movilidad de las personas con grado de doctor.

Ambas encuestas se desarrollaron en línea con el proyecto "*Career of Doctorate Holders*" (CDH) de la Dirección de Ciencia, Tecnología e Industria de la OCDE, el Instituto de Estadísticas de la UNESCO y la Oficina de Estadística de la Comisión Europea (EUROSTAT). En cada ocasión, se construyeron tres directorios de personas con grado de doctor que identificaban (i) profesionales con grado de doctor residentes en Chile (directorio 1), (ii) profesionales chilenos con grado de doctor residentes en el extranjero (directorio 2) y (iii) profesionales chilenos que se encuentran estudiando un programa de doctorado en Chile o el extranjero (directorio 3).

Posteriormente, a fines del 2019 y principios del 2020, se actualizaron los directorios de personas con grado de doctor al año de referencia 2019. Utilizando el Directorio 1 como marco muestral, la Oficina de Estudios y Estadísticas del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación solicitó elaborar y aplicar una nueva versión de las encuestas CDH, denominada "Encuesta de trayectoria Doctorados en Chile", la que también debía ser realizada de acuerdo a los lineamientos de la OCDE. El concurso público fue adjudicado por Datavoz Statcom (Licitación ID: 1098710-1-LE20).

Los objetivos de la licitación fueron los siguientes:

Objetivo General: El objetivo general de este estudio es la realización de una encuesta sobre la trayectoria de profesionales con estudios de doctorado en Chile utilizando como año de referencia el 2019. Esto debe realizarse levantando información, a través de una muestra representativa, de los doctores y doctoras que residen en el país.

Objetivos específicos:

a. Elaborar un diseño muestral que garantice representatividad en los dominios de estudio para los que se requiere evidencia sobre la trayectoria de las personas con estudios de doctorado.
b. Actualizar el cuestionario para la muestra mencionada en el punto anterior, de acuerdo con lineamientos OCDE.
c. Elaborar un informe final del trabajo de campo y de análisis de los resultados de la situación actual en la que se encuentra este sector del capital humano con estudios de doctorado en Chile.

El presente informe resume y documenta el trabajo realizado por Datavoz Statcom entre los meses de junio y diciembre del 2020 en el marco de dicha licitación. Siguiendo los lineamientos planteados en los Términos de Referencia, este informe está organizado en seis secciones. En la sección 2 se introduce y exponen los lineamientos internacionales del estudio. En la sección 3 se entregan definiciones y lineamientos metodológicos. En la sección 4 se describen las actividades previas a la aplicación de la encuesta y los aprendizajes de los procesos de pre-testeo y otras actividades realizadas para afinar el cuestionario aplicado. En la sección 5 se describen los protocolos utilizados para contactar a las personas encuestadas. En la sección 6 se presentan aspectos metodológicos asociados a la elaboración de la base de datos. En la sección 7 se resumen las principales estadísticas descriptivas obtenidos en la investigación. En la sección 8 se presentan algunos aprendizajes relevantes para futuras aplicaciones de la encuesta CDH.

# 2. Lineamientos de la OCDE

Los detalles de los lineamientos técnicos para la ejecución de la encuesta CDH se encuentran disponibles en el sitio web de la Dirección de Ciencia, Tecnología e Innovación de la Organización para la Cooperación y el Desarrollo Económicos (OCDE)[1]. Particularmente en el documento titulado "Mapping Careers and Mobility of Doctorate Holders"[2].

La documentación para el levantamiento de estadísticas de CDH dispuesta por la OCDE, sugiere una recopilación de datos en modo mixto adaptada a las posibilidades disponibles dentro de cada país. Esto incluye el uso de datos existentes, encuestas postales e Internet, entrevistas telefónicas y entrevistas personales. La experiencia internacional en el levantamiento de la encuesta CDH con año de referencia 2014 muestra una preferencia por encuestas auto aplicadas vía web y/o papel en los países Chile, Portugal y Canadá. Estados Unidos incluye además recolección de encuestas telefónicas con asistencia computacional (CATI); Corea además de CATI también llevó a cabo en última instancia encuestas personales (CAPI) cara a cara; sólo Rusia llevó a cabo encuestas cara a cara (CAPI) como método primario de recolección. Los lineamientos OCDE señalan las ventajas de los métodos CATI y CAPI, frente a otros sistemas auto aplicados donde los encuestadores pueden jugar un rol activo en la recolección de información de los encuestados. También se advierte sobre su mayor costo, el cual es justificado por las mayores tasas de respuesta que normalmente alcanzan. Y recomiendan su uso para complementar y mejorar las tasas de respuesta de levantamientos que tengan como método primario la recolección por correo electrónico o postal.

De acuerdo con la última información disponible[3], la recolección de datos mediante encuesta no es la única forma reconocida por la OCDE para la publicación de datos relevantes para la iniciativa CDH. Hay países que utilizan datos de censos nacionales para cubrir parte de la información como Canadá, Estonia y Rusia. Otros países utilizan información de encuestas nacionales de empleo (Labor Force Survey, LFS) como Bélgica, República Checa, Alemania, España, Reino Unido, Corea del Sur, Portugal y Rusia. Brasil, Colombia, Estonia, Finlandia, Países Bajos y Noruega, usan también datos de Registros civiles nacionales. Chile es uno de los tres países que utilizaron una encuesta dedicada como único medio de recolección de datos CDH junto con Letonia y Estados Unidos.

A continuación, se enumeran los contenidos principales de esta publicación y que guían el proceso de levantamiento y reporte de datos de este estudio. Son tres las secciones principales de los lineamientos OCDE: (i) Una sección referida a los lineamientos metodológicos, (ii) una sección focalizada en un cuestionario modelo junto con su manual de instrucción, y finalmente, (iii) una guía para el reporte de resultados e indicadores.

## 2.1 Lineamientos metodológicos

1 Sitio web OCDE – CDH https://www.oecd.org/innovation/inno/careers-of-doctorate-holders.htm
2 Documento de lineamientos técnicos https://www.oecd-ilibrary.org/science-and-technology/mapping-careers-and-mobility-of-doctorate-holders_5k4dnq2h4n5c-en
3 Disponible en los metadatos de la encuesta CDH 2017 se encuentra la información del tipo de encuesta utilizada para reportar las estadísticas CDH utilizado por cada país https://www.oecd.org/sti/oecd-careers-of-doctorate-holders-2017-data-release.zip

De acuerdo a los lineamientos metodológicos de la OCDE[4], el propósito de las estadísticas de trayectoria de doctores es la generación de datos armonizados para atender necesidades de políticas nacionales e internacionales en ámbitos como el rol de los doctores y las doctoras en innovación y en la economía del conocimiento, equilibrar oferta y demanda en el mercado laboral, formación enfocada en la inserción laboral, y movilidad de las personas con grado de doctor, entre distintos sectores económicos, ocupaciones y países.

La metodología para la generación de las estadísticas CDH establecidas por la OCDE permite y fomenta, cuando sea posible, el uso de estadísticas nacionales ya existentes. Estas pueden provenir por ejemplo de los resultados de encuestas nacionales de empleo, registros nacionales de educación, registros de empleo, estadísticas de I+D, organizaciones estudiantiles, censos de población entre otros.

En el caso que un país decida realizar la encuesta CDH por sí misma, los lineamientos OCDE ofrecen distintas alternativas como estudios de cohortes y estudios transversales, y establecen también criterios de estratificación y muestreo entre los que se considera el criterio de género utilizado en este estudio (ver sección 3). A su vez, la OCDE describe los aspectos metodológicos necesarios para la armonización de la información reportada por los distintos países. Y al mismo tiempo proporciona a los países que introducen la encuesta CDH por primera vez, los lineamientos metodológicos necesarios para llevarla a cabo.

Por último, la OCDE define la población objetivo de la encuesta CDH como aquellas personas que han obtenido el grado de doctor, que residan en el país, y sean menores de 70 años (69 o menos). El grado de doctor corresponde al mayor grado académico que una universidad puede otorgar, lo que equivale al nivel 6 del estándar de Clasificación Internacional Normalizada de la Educación (ISCED 1997) de la UNESCO.[5]

## 2.2 Cuestionario modelo y manual de instrucción.

El cuestionario modelo consiste en seis módulos que abordan la educación de las personas que poseen el grado de doctor (EDU), los puestos de investigación postdoctoral (ECR), la situación laboral (EMP), movilidad internacional (MOB), experiencia relacionada con el doctorado (CAR) y características personales (PER).

El cuestionario modelo propone la siguiente secuencia de módulos: EDU-ECR-EMP-MOB-CAR-PER. Sin embargo, se alienta a los países a organizar el cuestionario de acuerdo con sus propias necesidades y prioridades de políticas nacionales, ya sea mediante el cambio del orden de los módulos y preguntas, agregando nuevas preguntas y/o dejando fuera las preguntas del cuestionario modelo que no son centrales para las estadísticas CDH. La Tabla 1 indica la cantidad de preguntas totales, opcionales y requeridas por cada módulo.

*Tabla 1: Número de preguntas por módulo*

| Módulo | Total de preguntas | Preguntas opcionales | Preguntas requeridas |
|---|---|---|---|
| EDU | 12 | 6 | 6 |
| ECR | 9 | 9 | 0 |

4 Ver nota al pie n°2.

5 https://unece.org/fileadmin/DAM/env/esd/inf.meeting.docs/EGonInd/ISCED.97.manual.pdf

| | | | |
|---|---|---|---|
| EMP | 11 | 5 | 6 |
| MOB | 7 | 1 | 6 |
| CAR | 12 | 11 | 1 |
| PER | 8 | 0 | 6 |
| Total | 59 | 32 | 25 |

En conformidad con lo anterior, para esta versión de la encuesta CDH en Chile se incorporaron preguntas sobre temas de desigualdad de género y se acotaron varios módulos, tal como se puede observar en el Cuestionario final aplicado disponible entre los documentos de la encuesta publicados en la web (el trabajo realizado para esas definiciones se encuentra en el documento entregado en el marco del contrato entre Statcom Datavoz y la Subsecretaría de Ciencia llamado "Informe Previo de Trabajo de Campo").

La OCDE establece también la necesidad de disponer de material de apoyo para completar el cuestionario CDH, en el caso particular de las encuestas administradas por un encuestador o encuestadora este último deberá contar con un manual de instrucciones para completar el cuestionario. En el caso de que la encuesta sea auto administrada se debe procurar entregar toda la información necesaria para poder responder de forma exitosa.

## 2.3 Reporte de resultados e indicadores

Los métodos de recopilación de información nacional pueden ser heterogéneos entre países reflejando la diversidad de los sistemas nacionales de estadísticas. Esta sección detalla la definición de conceptos e indicadores para el reporte armonizado de los resultados a nivel internacional, y la forma en que estos deben ser tabulados.

Los lineamientos de la OCDE establecen los siguientes aspectos principales para la elaboración de los reportes finales:

- Lista de tablas de resultados de indicadores de resultados
    - Características personales
    - Características educacionales
    - Situación y percepción laboral
    - Movilidad internacional: interior y exterior
    - Producción científica
- Fuentes y definiciones de las variables propuestas para la tabulación de resultados

En las secciones siguientes, estas directrices de la OCDE son aplicadas y adaptadas para la realidad chilena.

# 3. Definiciones y lineamientos metodológicos

En esta sección se presentan las estadísticas esenciales del marco muestral en términos de edad y género, las decisiones metodológicas para la selección de la muestra y los errores muestrales estimados.

## 3.1 Marco Muestral

### 3.1.1 Población objetivo

La población objetivo total de las estadísticas de CDH consiste en todas las personas que, de acuerdo a la codificación OCDE ISCED 1997 (ver sección 2.1), poseen un nivel de educación nivel 6; lo que equivale al grado de Doctor. Esta población mundial de doctores y doctoras se divide en poblaciones nacionales, que se consideran las poblaciones objetivo de las encuestas nacionales en cada país. Las poblaciones objetivo a nivel país consisten en individuos que en la fecha de referencia cumplen con lo siguiente criterios:

- Tener una educación equivalente al nivel 6 de acuerdo al estándar ISCED1997 obtenida en cualquier parte del mundo. En Chile este nivel equivale al grado de Doctor.
- Ser residente (permanente o no permanente) dentro de las fronteras nacionales del país encuestador.
- Ser menor de 70 años (69 años de edad y menos).

Con estas definiciones y el supuesto (teórico) de que todos los países están realizando la encuesta y utilizando la misma fecha de referencia, toda la población de doctorados residentes en todos los países es cubierto sin ninguna superposición.

### 3.1.2 Fecha de referencia

Para todos los efectos del cuestionario la fecha de referencia considerada debe ser el 1 de diciembre del año inmediatamente anterior. En este caso el 1 de diciembre del 2019.

### 3.1.3 Estratificación

En encuestas probabilísticas, la cantidad de dominios necesarios a menudo determina la cantidad de estratos. Sin embargo, para estadísticas multipropósito como las estadísticas de la encuesta CDH, el número de dominios a menudo excederá el número deseable de estratos. Dado que un número demasiado alto de estratos podría generar un problema con muy pocas observaciones en algunos estratos, es importante mantener la estratificación a un nivel razonable. La estratificación más obvia y significativa es desglosar la población de acuerdo con los marcos de muestreo aplicados. Dado que los estratos deben ser lo más homogéneos posible con respecto a las principales variables analizadas, los desgloses recomendados adicionales están de acuerdo con las características año de graduación, campos de conocimiento de Ciencia y Tecnología y género.

## 3.2 Consideraciones del diseño muestral

En conformidad con los lineamientos metodológicos de la Encuesta CDH, el diseño muestral implementado consideró los siguientes elementos claves:

- **Población objetivo:** Todas las personas con grado de doctor que residen en Chile (de forma permanente o no) que sean menores de 70 años (69 años de edad y menores).
- **Marco muestral:** Directorio de doctores y doctoras residentes en Chile entregado por la contraparte técnica. Este incluye un total de 18.352[6] registros con y sin datos de contacto.
- **Tamaño muestra**: 2.400 casos.
- **Tipo de muestreo:** aleatorio estratificado, siendo los estratos definidos en función del género.

En relación al universo al que hace referencia la población objetivo, de acuerdo a estimaciones realizadas por la contraparte, es posible asumir que el tamaño del directorio obtenido en su última actualización el año 2019 posea una cobertura alta respecto del universo de estudio. Esto se deduce del hecho que, de acuerdo con lo observado en el año 2014, las cifras de becarios reportados por ANID se corresponden en tamaño con el directorio elaborado en esa oportunidad, de aproximadamente 10.500 PhD. Para el 2018 ANID reporta 14.229 PhD, en tanto el directorio elaborado el año 2019 alcanza a 18.352 registros.

En consecuencia, para efectos de lo que se desarrolla en adelante se asumirá que el directorio del año 2019 posee una cobertura lo suficientemente grande del universo, de modo que, para efectos del cálculo de precisiones muestrales (errores), se considerará que el tamaño del universo a nivel total y de segmentos de interés es el que resulta de dicho marco.

## 3.3 Universo estimado

Dado los supuestos en la sección 3.2, para la estimación del tamaño del universo, es necesario la construcción de un marco muestral lo más amplio (cobertura) y comprehensivo posible, a partir de las bases de datos proporcionada por la contraparte, particularmente:

- Directorio de profesionales con grado de doctor actualización 2019
- Directorio de profesionales con grado de doctor de LinkedIn

Para la construcción de este marco muestral se requiere concatenar ambas bases de datos considerando tres aspectos metodológicos:

- Evitar casos duplicados entre ambas bases
- Incluir el total de casos que pertenecen a la población objetivo
- Identificar los casos que pertenecen a cada estrato de interés

A continuación, se detallan las acciones de manejo de datos que se realizaron en ambas bases, necesarias para la concatenación y construcción del marco muestral.

### 3.3.1 Directorio de profesionales con grado de doctor - Actualización 2019

6 La obtención del tamaño del marco muestral se describe en la Sección 3.5

Para la construcción del marco muestral se utilizó el Directorio de personas con grado de Doctor en Chile entregado por la contraparte. Este directorio ha sido actualizado periódicamente con el doble objetivo de mantener actualizados los datos de contacto de los doctores y doctoras registradas en él, y, al mismo tiempo, incorporar a las personas con grado de Doctor que aún no se encuentran en sus registros. El último proceso de actualización se hizo a fines del año 2019 y comprendió tres etapas:

La primera etapa consistió en la construcción de una base de datos de contacto de doctores y doctoras mediante el cruce y homologación de registros del Sistema de Información de Educación Superior (SIES) del Ministerio de Educación (MINEDUC), de la Agencia Nacional de Investigación y Desarrollo (ANID) del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (MINCIENCIA), de centros de investigación públicos y de la versión anterior del Directorio. Los datos recolectados durante la primera etapa se concentran principalmente en doctores y doctoras que se desempeñan en el sector de Educación Superior y que financiaron su doctorado con becas del Estado de Chile.

En una segunda etapa se construyó una base de datos de contacto de profesionales que se desempeña en el área de I+D+i con un alto potencial de poseer el grado de Doctor. Esto tiene como objetivo complementar los datos recolectados y aumentar la exhaustividad de la búsqueda en el sector empresarial. Las fuentes de información para la construcción de esta base provienen principalmente de registros de beneficiarios de programas vinculados a la I+D+i de entidades públicas, como la Corporación de Fomento de la Producción (CORFO), el Instituto Nacional de Propiedad Intelectual (INAPI) y la Agencia Nacional de Investigación y Desarrollo (ANID). Una vez consolidada la base de datos de contacto de doctoras y doctores con la de profesionales con alto potencial de poseer el grado de doctor se llevó a cabo la tercera etapa. Esta consiste en la recolección de datos actualizados mediante un formulario en línea que fue enviado por correo a los profesionales con datos de contacto disponibles en la base consolidada tanto de doctores y doctoras (que viene de la primera etapa) como en la de potenciales personas con grado de doctor (que viene de la segunda etapa) para que actualicen sus datos de contacto o se registren como persona nueva en el directorio. En total 9.994 personas respondieron el formulario, ya sea actualizando sus datos o registrándose como nuevas personas en el directorio de doctores.

El Directorio de Doctores en Chile 2019 tiene un total de 27.521 registros los que se pueden clasificar en 3 subdirectorios según lugar de residencia (en Chile o el extranjero) y etapa de estudios (estudiante o graduado) tal como se muestra en la siguiente tabla.

*Tabla 2:Número de casos por subdirectorio*

| Subdirectorios | Número de casos |
|---|---|
| Directorio 1 | 21.665 |
| Directorio 2 | 794 |
| Directorio 3 | 3.653 |
| Sin información | 1.409 |
| Total general | 27.521 |

Elaboración propia a partir de Directorio de Doctores 2019

El Directorio 1 corresponde a profesionales con grado de doctor residentes en Chile; el Directorio 2 son las personas con grado de doctor de nacionalidad chilena residentes en el extranjero; y el Directorio 3 a estudiantes de programas de doctorado con ciudadanía chilena. Los profesionales para los cuales no hay suficiente información disponible para su clasificación se agruparon en la categoría "Sin información".

Para identificar los registros que pertenecen a la población objetivo se llevó a cabo una serie de filtros aplicados respecto a la pertenencia al directorio 1, la validez del RUT y la pertenencia al rango de edad establecido para el proyecto CDH. El número de observaciones resultantes luego de aplicar cada uno de los filtros en el orden indicado, se especifica en la columna "Número de casos" de la Tabla 3. De esta manera se obtiene un total de 16.491 casos. Los filtros aplicados y los universos resultantes fueron los siguientes:

*Tabla 3:Número de casos que cumplen las condiciones de la población objetivo*

| Condiciones | Número de casos |
|---|---|
| Casos que pertenece al Directorio 1 | 21.665 |
| Y que además tiene un RUT válido | 18.543 |
| Y que además se encuentran en el rango de edad entre 18-69 años | 16.491 |

Elaboración propia a partir de Directorio de Doctores 2019

### 3.3.2 Directorio de profesionales con grado de doctor de LinkedIn

Tal como se mencionó en la sección anterior, el proceso de actualización del Directorio de personas con grado de Doctor en Chile depende principalmente de registros de instituciones públicas vinculadas a la educación superior y al financiamiento público de las I+D+i. Sin embargo, estas fuentes de información no tienen el mismo nivel de cobertura sobre doctores y doctoras que se desempeñan en el sector empresarial, que no participan de actividades financiadas I+D+i que hayan postulado a fondos públicos que financien esas actividades y/o que no hayan financiado su doctorado mediante becas del Estado de Chile.

Es por eso que, en una licitación previa a la presente, con el objeto de ampliar el marco muestral hacia el sector empresarial, el Ministerio de Economía a través de la Subsecretaria de Economía y Empresas de Menor Tamaño, había encargado una búsqueda y descarga de perfiles con grado de doctor y residentes en Chile en la red social LinkedIn, utilizando herramientas computacionales de "Raspado de datos" (*Web scraping*). Los datos recolectados incluían, entre otras cosas, nombre completo, descripción del perfil, un resumen o extracto, una breve reseña de su experiencia laboral, nombre del grado, lugar de obtención del grado, fecha de inicio y de término del doctorado.

*Identificación del sector empresarial*

A partir de la información disponible en la base de datos obtenida de LinkedIn, es posible estimar un modelo de clasificación capaz de identificar, con un cierto margen de error, si un doctor o doctora se desempeña o no laboralmente en el sector empresarial. Para ello se utilizó en dicha licitación un modelo de análisis de texto que transforma las descripciones y reseñas de los perfiles en variables numéricas, y que a su vez alimenta un modelo de clasificación de regresión logística.

Esta diferenciación permitió seleccionar los casos identificados como pertenecientes al sector empresarial y utilizarlos como insumo en la construcción del marco muestral.

*Construcción de variables de población objetivo*

Los datos extraídos de LinkedIn presentaron ciertas limitaciones para su inclusión en el marco muestral. En particular, el directorio de LinkedIn no tenía disponibles los datos de edad, necesarios para la definición de la población objetivo, y tampoco sobre el sexo, necesario para la estratificación. Para solucionar estos problemas se procedió a la creación estimaciones de ambas variables edad y sexo, a partir de información contenida en los mismos registros:

- **Edad** se estimó a partir de los años de inicio y/o finalización del doctorado, considerando una fecha de obtención de grado de 36 años. Así, la edad aproximada se puede calcular como 36 años más la cantidad de años desde la fecha de obtención de grado hasta el año 2019.[7]
- **Sexo** se estimó a partir del nombre de cada persona. Los nombres en español suelen estar asociados a un sexo por lo que se utilizó este dato como proxy. Para nombres ambiguos y/o de procedencia extranjera se utilizó como criterio las características de género expresadas en la foto de perfil de la cuenta de LinkedIn asociada.

La base de doctores y doctoras construida a partir de LinkedIn tiene un total de 4.390 registros de los cuales 2.096 casos fueron identificados como registros existentes en la base del Directorio de Doctores 2019, basado en un criterio de pareo por coincidencia ("matching") entre los nombres completos en ambas bases de datos.

Este proceso de unión por coincidencia de nombre utilizó un método llamado Índice de Jaccard el cual es utilizado para cuantificar la similitud de dos conjuntos en función de sus elementos. Aplicado a la similitud entre nombres, el índice se calcula como cociente entre el número de palabras comunes entre dos nombres (la intersección) y el número total de palabras únicas entre ambos nombres (unión). El índice de Jaccard entre dos nombres se calcula de la siguiente manera:

$$Jaccard(nombre_1, nombre_2) = \frac{|nombre_1 \cap nombre_2|}{|nombre_1 \cup nombre_2|}$$

El pareo por coincidencia entre ambas bases de datos consistió en seleccionar sólo los pares con uniones confiables. El criterio de confiabilidad considerado es un índice de Jaccard igual o mayor a 0.65, y se consideró que no existe coincidencia si el índice de Jaccard es menor que 0.65.

Para la concatenación de la base de datos de LinkedIn y Directorio de Doctores 2019 fue necesario excluir los casos que se encuentran en ambas bases de datos (aquellos con coincidencia de nombres) y además aplicar los filtros correspondientes a la población objetivo. En consecuencia,

7 Esta metodología es una variación de la utilizada para crear la variable de edad en el Directorio de Doctores 2019 en el que se usó el año de ingreso al programa de doctorado y una edad promedio de 30 años. En el caso del Directorio de Doctores de LinkedIn se optó por utilizar el año de obtención del grado, debido a que el dato tiene mayor cobertura que el año de ingreso al programa de doctorado. La edad utilizada es 36 años, correspondiente a la edad promedio de egreso del doctorado.

se obtuvo un total de 2.126 casos. Este número de observaciones resulta de aplicar cada filtro en el orden indicado en la Tabla 4.

*Tabla 4:Número de casos que cumplen con las condiciones de la población objetivo*

| Condiciones | Número de casos |
|---|---|
| Total de casos del Directorio de Doctores de LinkedIn | 4.390 |
| Sin coincidencia con el Directorio de Doctores en Chile 2019 (Índice de Jaccard menor a 0.65) | 2.276 |
| Se encuentran en el rango de edad entre 18-69 años | 2.126 |

Elaboración propia a partir de Directorio de Doctores de LinkedIn

### 3.3.3 Concatenación de bases en un directorio combinado

A partir del resultado de la concatenación de ambos directorios en un solo directorio combinado, se generó una base de datos con 18.617 registros, que corresponden a personas con el grado de doctor residentes en Chile y que se encuentran entre los 18 y 69 años. La distribución de estos casos de acuerdo a su género es la siguiente tabla:

*Tabla 5:Distribución del Universo según Género*

| Género | Directorio 1 | LinkedIn | N | % |
|---|---|---|---|---|
| Femenino | 6.152 | 578 | 6.730 | 36,1% |
| Masculino | 10.298 | 1.324 | 11.622 | 62,4% |
| Otro | 15 | | 15 | 0,1% |
| Sin información | 26 | 224 | 250 | 1,3% |
| Total Universo | 16.491 | 2.126 | 18.617 | 100% |

Fuente: Elaboración propia a partir de Directorio 1 y base de LinkedIn.

### 3.3.4 Definición del marco muestral

De acuerdo a las consideraciones de diseño muestral descritas en la sección 3.2, una vez estimado el tamaño del universo, el marco muestral y sus estratos se definen en función del género consolidando un **marco muestral de 18.352 personas**. La distribución del marco muestral según el género se muestra en la Tabla 6. Fuera del marco quedaron 265 personas para las que no se tiene información de su género o bien declaran una identidad de género no incluida en los géneros femenino y masculino.

*Tabla 6:Distribución del Marco Muestral según Género*

| Género | Directorio 1 | LinkedIn | N | % |
|---|---|---|---|---|
| Femenino | 6.152 | 578 | 6.730 | 36,7% |
| Masculino | 10.298 | 1.324 | 11.622 | 63,3% |
| Total Marco Muestral | 16.450 | 1.902 | 18.352 | 100% |

Si bien, algunos registros del marco muestral, provenientes de la base de datos de LinkedIn, no cuentan con datos de RUT o con información de contacto completa y actualizada, estos sí proporcionan la información necesaria sobre la distribución de la población para definir el tamaño de cada estrato, el cálculo de los errores y posteriormente los respectivos ponderadores y factores de expansión.

### 3.3.5 Definición del marco de selección

Para efectos de la selección de las unidades de la muestra es necesario definir un marco muestral alcanzable de acuerdo los siguientes criterios de inclusión señalados en las bases técnicas:

- Todos los registros del directorio elaborado el 2019 **que poseen datos de contacto**
- Todos los registros identificados a través de **LinkedIn el mismo año, que se desempeñan en empresas públicas o privadas y que no poseen datos de contacto**[8]

De los 16.450 registros pertenecientes al marco muestral y provenientes del Directorio 1, 3.431 fueron excluidos por no poseer con datos de contacto, cifra que se compone en términos de género por 2.078 personas de género masculino y 1.353 de género femenino. El marco de selección resultante está compuesto por 13.019 registros provenientes del Directorio 1 y 1.902 registros provenientes de la base de datos de LinkedIn.

Tal como se muestra en la Tabla 7 la distribución final del marco de selección según género tiene una diferencia de un 0,7% respecto a la distribución del marco muestral, por lo que es posible suponer que la disponibilidad de datos de contacto es aleatoria.

*Tabla 7:Distribución del Marco de Selección según Género*

| Género | Directorio 1 | LinkedIn* | N | % |
|---|---|---|---|---|
| Femenino | 4.799 | 578 | 5.377 | 36,0% |
| Masculino | 8.220 | 1.324 | 9.544 | 64,0% |
| Total Marco Selección | 13.019 | 1.902 | 14.921 | 100,0% |

* De los 1902 casos, 494 corresponden a personas que se desempeñan en el sector privado. Para una revisión de los esfuerzos de contacto realizados sobre ese subsegmento véase la sección 4.2

## 3.4 Tamaño de muestra

El tamaño de muestra final que se propuso alcanzar fue de **n = 2400 casos**, cuyo error muestral absoluto máximo es de **1,8%,** considerando varianza máxima, un nivel de confianza del 95%, corrección por finitud y un efecto de diseño deff[9] de 1,0. Este tamaño muestral fue superado en el levantamiento. Los detalles de la muestra efectiva se presentan en la sección 6 del presente documento.

El tipo de diseño que se utilizó fue el aleatorio estratificado, este implica la selección de una muestra aleatoria simple en cada uno de los estratos que define la variable correspondiente al género. Con los antecedentes anteriores, la distribución del tamaño de muestra final por género (considerando asignación proporcional) fue la que se presenta a continuación. Se incluye además los errores muestrales absolutos máximos en los estratos marginales.

*Tabla 8:Distribución del tamaño muestral según Género y errores muestrales absolutos máximos*

| Género | N | % | e.m.m. |
|---|---|---|---|
| Femenino | 880 | 36,7% | 3,08% |
| Masculino | 1520 | 63,3% | 2,34% |

8 En la primera etapa del estudio se realizó un procedimiento de búsqueda de datos de contacto para este grupo de interés.

9 Corresponde al incremento/disminución que se produce en la varianza de la estimación debido al uso de un muestreo diferente al muestreo aleatorio simple.

| | | | |
|---|---|---|---|
| Total Muestra | 2.400 | 100,0% | 1,87% |

Fuente: Elaboración propia.

En dicho cuadro se observa que los márgenes de error no superan el 3,1% en cada segmento de interés. Respecto de estos errores muestrales absolutos máximos (e.m.m.) para cada muestra, ya sea a nivel total como en estratos de interés, se ha considerado para su cálculo los siguientes valores y criterios para los parámetros que utiliza:

- Muestreo aleatorio simple (m.a.s.[10])
- Varianza máxima considerando la estimación de una proporción ($p = 0{,}5$)
- Un nivel de confianza ($1 - \alpha$) de 95%
- Corrección por universo finito $\left(1 - \frac{n}{N}\right)$, en que "N" es el tamaño de universo y "n" el tamaño de la muestra

Dicho error se puede obtener a partir de la aplicación de la siguiente formula[11]:

$$error.muestral.absoluto = Z_{1-\alpha/2} * \sqrt{\frac{p \cdot (1-p)}{n} \cdot \left(1 - \frac{n}{N}\right) \cdot deff}$$

## 3.5 Selección de la muestra

En relación al procedimiento de selección de la muestra a ser levantada en cada estrato, como se indicó previamente se ejecutó mediante un *muestreo aleatorio estratificado*. Como se indica en la sección 0, esto significó 3.3.5 la selección de una muestra de 880 registros de los 5.377 registros del estrato de personas de género femenino del marco de selección mediante un *muestreo aleatorio simple*, y de igual manera para el estrato de género masculino se seleccionaron 1.520 registros de los 9.544 registros pertenecientes al marco de selección.

En total de los 14.921 registros del marco de selección se sorteó una muestra de 2.400 registros cuya distribución se detalla en la Tabla 9.

*Tabla 9:Distribución de la muestra seleccionada según Género*

| Género | Directorio 1 | LinkedIn | N | % |
|---|---|---|---|---|
| Femenino | 787 | 93 | 880 | 36,7% |
| Masculino | 1.296 | 224 | 1.520 | 63,3% |
| Total general | 2.083 | 317 | 2.400 | 100,0% |

Fuente: Elaboración propia.

Por otra parte, el procedimiento contempló la selección de una **muestra original y una de reemplazo** en cada estrato, esta última aplicando el mismo procedimiento de selección que la original. También se consideró reemplazos en los casos en que se detectaran respuestas que no pertenezcan a la población objetivo.

10 Muestreo aleatorio simple.

11 En que $Z_{1-\alpha/2}$ =1,96 (percentil 97,5 de la distribución normal estándar al considerar un nivel de confianza del 95%)

En relación al tamaño de la muestra de reemplazo, se propuso **el doble de la muestra original**, esto implicó asumir que de cada 3 contactados 1 debería responder. Es importante señalar, que la muestra de reemplazo sólo se utilizó en el caso que no se cumpliera con el tamaño propuesto en cada estrato (lo que efectivamente ocurrió, ver sección 5.2, por lo que no corresponde considerarla como una sobre-muestra).

*Tabla 10:Distribución de la muestra de reemplazo seleccionada según Género*

| Género | Directorio 1 | LinkedIn | N | % |
|---|---|---|---|---|
| Femenino | 1.560 | 200 | 1.760 | 36,7% |
| Masculino | 2.615 | 425 | 3.040 | 63,3% |
| Total general | 4.175 | 625 | 4.800 | 100,0% |

Fuente: Elaboración propia.

Se puede observar en la Tabla 9 y la Tabla 10 que, de la base de datos de personas con grado de doctor obtenida de LinkedIn 317 registros fueron seleccionados en la muestra original y 625 fueron seleccionados en la muestra de reemplazo. Debido a que la base de datos de LinkedIn no posee, por su origen, disponibilidad de datos de contacto, una de las actividades mínimas definidas en las bases técnicas es la recolección de datos de contacto para lograr una mayor cobertura de los doctores y doctoras que se desenvuelven en el sector empresarial. Este procedimiento se describe en mayor detalle en la sección 4.1

Los resultados de la selección de la muestra se encuentran en el documento "Directorio_2020_Muestra.xlsx" entregado por Datavoz Statcom a la contraparte técnica. El documento consiste en un subconjunto de datos del directorio concatenado de 7200 casos, estos se desglosan en la muestra original de 2400 casos y la muestra de reemplazo de 4800 casos.

Los resultados del proceso de levantamiento consideraron el reporte de los indicadores de rendimiento de la muestra. Para cada contacto realizado, se dejó registro de la incidencia del contacto según una adaptación de los códigos de disposición final de la AAPOR[12], los que además serán usados para el cálculo de las tasas que se señalan más adelante y para el análisis de la NO respuesta.

*Tabla 11:Códigos de Disposición Final adaptados*

| **CÓDIGOS DE DISPOSICIÓN FINAL PARA SEGUIMIENTO CATI** | |
|---|---|
| **1. Entrevistado, elegible**<br>110. Entrevista completa<br>120. Entrevista parcial<br>**2. No entrevistado, elegible**<br>211. Se rechazó la entrevista<br>212. Se interrumpió la entrevista<br>224. No responde teléfono/teléfono fuera de servicio<br>225. Informante no ubicable o no puede atender<br>236. Otra razón elegible | **4. No elegible**<br>410. Fuera de muestra<br>451. Empresa, oficina de gobierno u otra organización<br>452. Instituciones (hospital, cárcel, asilo de ancianos, etc.)<br>463. Otra razón no elegible |

Fuente: Elaboración propia.

12 https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

En base a los códigos anteriores de la AAPOR se obtienen los indicadores de disposición de casos que se solicitan en los términos de referencia:

- Número de entrevistas completas (I)
- Número de entrevistas parciales (P)
- Número de rechazo o interrupción de la entrevista (R)
- Número de no contacto (NC)
- Número de otros casos (O)

Esto permitirá la entrega de tasas de respuesta, cooperación, rechazo y de contacto, indicadores que son importantes para ir analizando el desarrollo del levantamiento y ver la implementación de planes de contingencia en el caso que alguno de los indicadores sea crítico y mayor a lo esperado.

Con respecto al tamaño relativo de la muestra los lineamientos metodológicos de la OCDE recomiendan para la encuesta CDH una relación entre el tamaño total de la muestra y el tamaño total de la población mayor al 20%. La muestra original de 2.400 casos representa un 13,1% en relación al tamaño del marco muestral de 18.352. Este porcentaje logrado puede llevar a interpretaciones equívocas que es recomendable poner en perspectiva. Primero, en la presente versión, el tamaño del directorio aumentó en 7.510 personas respecto de la versión 2014 (n = 10.842) debido al esfuerzo desplegado en expandir el directorio como fue explicado en secciones anteriores. Segundo, la muestra efectiva aumentó desde 1666 a 2445 (ver sección 6). Es decir, se logró un aumento en el número de encuestas de un 46,7% (numerador), pero un aumento de registros de un 69,2% (denominador). Para lograr este resultado, en la presente versión se hizo un esfuerzo de contacto de 7.200 personas (2.400 de la muestra original y 4.800 de la muestra de reemplazo) lo que corresponde a un 39,2% del directorio.

# 4. Actividades previas a la aplicación de la encuesta

En esta sección se describen los procesos desarrollados para validar el cuestionario y los métodos de recolección de datos de contacto faltantes de personas en el marco muestral que provienen de LinkedIn.

## 4.1 Recolección de datos de contacto del Directorio de LinkedIn

Debido a que las fuentes de datos para la construcción del marco muestral provienen mayoritariamente de instituciones públicas ligadas a la educación superior y programas públicos ligados a becas de doctorado y actividades de I+D, la presencia de personas con grado de doctor que se desempeñan en el sector empresarial ha sido sub representada en sondeos anteriores. Para reducir este sesgo y aumentar la exhaustividad del marco muestral en el sector empresarial, una de las actividades solicitadas en las bases técnicas es la incorporación de la base de datos de LinkedIn en el marco de selección y la recolección de datos de contacto para este subgrupo. En el marco de selección hay 1.902 casos provenientes de la base de datos de LinkedIn originalmente sin datos de contacto, de ese total, 494 corresponden a personas que se desempeñan en el sector empresarial. La recolección de los datos de contacto para poder encuestar a estas personas se llevó a cabo mediante tres métodos y que fueron seleccionadas en la muestra original o la de reemplazo (para más detalles véase la sección 3.3.5).

- Contacto y solicitud de información a instituciones de educación superior donde cursaron el grado.
- Bola de nieve con doctores del marco muestral que hayan cursado el mismo grado en la misma institución.
- Búsqueda web de información de contacto disponibles en fuentes abiertas, como repositorios académicos, directorios de empresas y búsquedas generales en internet.

Del contacto a instituciones se comunicó a 219 instituciones, las que representaban a 406 doctores y doctoras. Se obtuvo respuesta de 22 instituciones lo que se obtuvo 22 datos de contacto.

A partir del método de Bola de Nieve[13], en el que se contactó a personas con grado de doctor que posiblemente conocieran a las personas de interés, solo se obtuvo una respuesta con los datos de contacto de un doctor.

El método de búsqueda web resultó en 47 datos de contacto a partir de búsquedas en repositorios académicos, páginas web del lugar de empleo y búsquedas generales en motores de búsqueda tradicional.

Adicionalmente a los casos actualizados, la contraparte hizo entrega de una base de datos de contacto de un total de 84 personas pertenecientes al directorio de LinkedIn, obtenidas a partir de datos administrativos del ministerio. Con esto se alcanzó **un total de 154 datos de contacto** recolectados de personas en LinkedIn ligados al sector empresarial. Estos contactos fueron incluidos

[13] El método de bola de nieve consiste en obtener información de sujetos que cumplen con los requisitos buscados a partir de la información entregada por otras personas contactadas previamente que las referencian o conocen.

en el marco muestral de acuerdo a lo descrito en la sección 3.3.4 y que se detalla en la siguiente tabla.

Tabla 12: Resultados de la recolección de datos de contacto del Directorio de LinkedIn vinculadas al sector empresarial

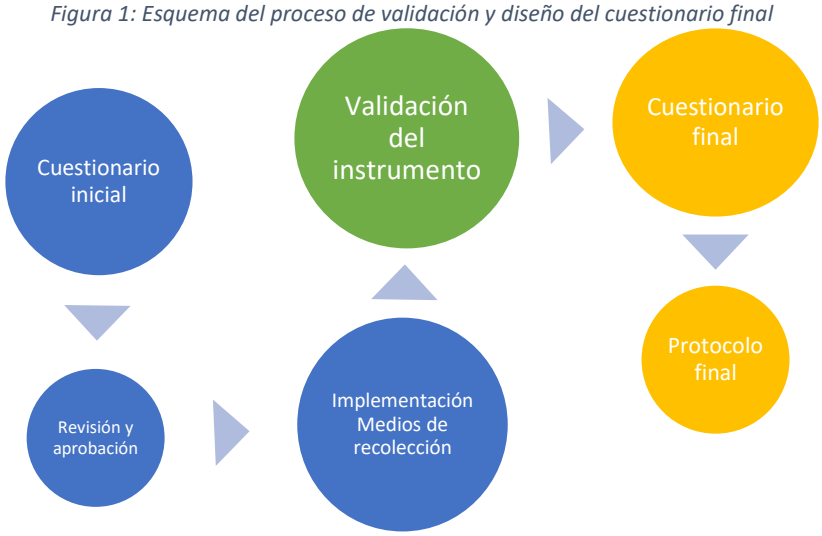
| Metodología | N |
|---|---|
| Contacto a instituciones | 22 |
| Bola de nieve | 1 |
| Búsqueda web | 47 |
| Minciencia | 84 |
| Total incorporado en la muestra | 154 |
| No incorporados en el marco muestral | 340 |
| Total | 494 |

## 4.2 Validación del instrumento y medio de recolección

Para la realización de la Encuesta de Trayectoria de Profesionales con Grado de Doctor en Chile (Encuesta CDH) fue necesario adaptar el cuestionario modelo propuesto por la OCDE a la realidad nacional y validar su fraseo, medios de recolección y protocolos de aplicación con el fin de garantizar la calidad de los datos recolectados.

En una primera instancia se realizó la adaptación del instrumento modelo conservando de forma integral su estructura y todas sus preguntas. Esta adaptación inicial del cuestionario fue revisada y aprobada por la contraparte técnica para su implementación en los medios de recolección propuestos y su posterior validación mediante la realización de un pretest. Dicho proceso se ilustra en la Figura 1, la etapa de validación del instrumento

*Figura 1: Esquema del proceso de validación y diseño del cuestionario final*




De los resultados del proceso de validación surgieron las consideraciones y recomendaciones para el diseño del cuestionario final a utilizar en el levantamiento de la Encuesta CDH en Chile en su versión 2019. A continuación, se detalla la metodología utilizada en el proceso de validación, incluyendo las fuentes de información, actividades y resultados.

### 4.2.1 Metodología de Validación del Cuestionario

La validación del cuestionario inicial contempla la integración de las percepciones, comentarios, interacciones y experiencias de la población objetivo mediante tres métodos: Las entrevistas cognitivas, el pretest vía telefónica (CATI) y el pretest vía web (CAWI).

La entrevista cognitiva es una encuesta especializada, centrada en descubrir la forma en que el informante piensa cuando lee/escucha las preguntas de un cuestionario. Las entrevistas se realizan a una pequeña muestra seleccionada de forma dirigida (en esta ocasión por DATAVOZ). El objetivo fundamental de esta técnica es descubrir cómo está interpretando el significado de las preguntas y las posibles respuestas, qué está pensando cuando delibera qué responder, cómo decide sus respuestas y qué significado tienen las mismas. Debido a las restricciones de movilidad producto de la crisis sanitaria por SARS-COV-2, las entrevistas se llevaron a cabo en formato online por medio de la plataforma de videoconferencia Zoom.

Los pretest (CATI y CAWI) son encuestas que se aplica a una muestra aleatoria simulando lo más fielmente posible el levantamiento de la Encuesta CDH. Este pilotaje permite observar la interacción entre los informantes y los medios de recolección, midiendo, por ejemplo, los tiempos de respuesta por pregunta y la duración total de la encuesta. Es de especial utilidad para evaluar cómo se distribuyen los recursos atencionales de los profesionales encuestados, es decir qué preguntas son más fáciles de comprender por los encuestados y cuáles requieren mayor atención y concentración de su parte.

*Figura 2: Metodología de validación de los instrumentos y sus participantes*

El tamaño de la muestra para la encuesta preliminar estuvo estipulado por las bases técnicas de la licitación del estudio y corresponde al 3% del tamaño de la muestra del levantamiento, es decir, 72 casos los que están distribuidos entre ambos medios de recolección de la siguiente manera:

- 30 casos para el levantamiento por vía telefónica (CATI)
- 42 casos para el levantamiento por vía web (CAWI)

Estos dos métodos de validación son considerados como actividades independientes y de la misma manera sus resultados se presentarán en secciones separadas.

### 4.2.2 Resultados de Validación del Cuestionario

En la presente sección se describen los principales aprendizajes obtenidos durante las actividades del proceso de validación con el objeto de que sirvan de insumos para futura aplicaciones de esta encuesta. Este registro es especialmente importante porque no todas las observaciones realizadas por los encuestados pudieron ser atendidas en la versión final del cuestionario debido a las restricciones de límite de tiempo de aplicación que debe cumplir el cuestionario para no desincentivar el abandono de la encuesta y/o el rechazo a participar de la misma.

*Entrevistas Cognitivas*

En la versión original, todos los entrevistados consideraron que la encuesta era excesivamente larga, por lo que reducir el número de preguntas y módulos se convirtió en una prioridad en la formulación del cuestionario definitivo.

Algunos entrevistados tuvieron comentarios respecto de la presentación de la encuesta. Específicamente, mostraron preocupación por el real anonimato de la encuesta. Al respecto, indicaron que en esa primera parte debía quedar completamente claro qué información de ellos se iba a compartir con el ministerio de forma innominada y cuál parte no. Adicionalmente los entrevistados indicaron que, dado que la encuesta era telefónica, las personas encuestadas tendrían mayor interés en responder.

También, en relación a la privacidad de los datos, la contraparte de la Subsecretaría de Ciencia manifestó su interés de que las respuestas a algunas preguntas pudieran ser utilizadas para ser cruzadas con otras bases de datos con el objetivo de estudiar de forma más amplia el sistema CTCI. Eso requería un consentimiento informado explícito por parte de los encuestados y surgió la duda sobre si eso debía hacerse al inicio del cuestionario o al final (en cuyo caso, se indicaría sobre qué preguntas específicas se pedía autorización). Esto fue consultado a los entrevistados quienes indicaron que preferían que esos consentimientos se hicieran al inicio de la encuesta y no al final.

Algunos entrevistados plantearon que la encuesta requería un importante esfuerzo cognitivo en algunas preguntas y que debido a que tenían poco tiempo, es probable que de haber sido seleccionados en la muestra a encuestar no la habrían podido terminar.

Los entrevistados también plantearon que la encuesta debería indagar también sobre la vinculación de los encuestados con estudiantes doctorales o de magister. De modo de saber cuánto se está compartiendo y formando conocimiento. También se señaló que se podría preguntar por la participación en conferencias internacionales, como indicador de la formación de redes.

También hubo reparos al fraseo de las preguntas, solicitando que se utilizara lenguaje inclusivo. Por ejemplo, se solicitó que el fraseo indicara “doctores y doctoras”.

Como se indicó anteriormente, la necesidad de limitar el cuestionario que era de una duración de aplicación muy larga (más sobre este punto en la sección del pretest) impedía acomodar todas estas inquietudes de los entrevistados.

El recuadro siguiente muestra algunos ejemplos de las inquietudes presentadas por los entrevistados para algunas preguntas específicas. Un reporte completo de los comentarios pregunta por pregunta se encuentra disponible en el informe parcial 3 en el marco de este proyecto.

*Tabla 13: Ejemplos de reacciones de los entrevistados*

MÓDULO: CARACTERÍSTICAS DE EDUCACIÓN (EDU)

Pregunta EDU_1: *¿En qué institución obtuvo su grado de Doctor?*

La pregunta no considera las posibilidades de doble grado. Tampoco considera que existen personas con dos doctorados. En este último caso ¿qué se responde? ¿El primero? ¿El último?

MÓDULO: ETAPA POSTDOCTORAL (ECR)

Pregunta ECR_2: *¿Cuáles de estas características se pueden aplicar a su postdoctorado?*

Una entrevistada indica que no se entiende la diferencia entre las opciones ECR_2_3 ("Es temporal") y ECR_2_8 ("Tiene una duración determinada")

MÓDULO: SITUACIÓN LABORAL (EMP)

Pregunta EMP_3A: *Por favor proporcione los siguientes detalles acerca de su empleo PRINCIPAL al 1 de diciembre de 2019*

En un caso se plantea la pregunta "¿Entonces es o no es anónima esta encuesta?"

Con respecto a la pregunta EMP_3A_32 sobre la Ocupación del empleo Principal, ninguno de los casos fue capaz de clasificarse de acuerdo al flujo de preguntas presentado.

En un caso, el entrevistado en la pregunta EMP_3A_32 selecciona, profesionales científicos e intelectuales, seleccionando como especificación en EMP_3A_34 administración pública y empresas, pero al no identificarse con ninguna especificación, pide cambiar su respuesta a profesionales en derecho, en ciencias sociales y culturales, donde nuevamente siente que no puede incluirse en ninguna categoría.

MÓDULO: MOVILIDAD INTERNACIONAL (MOB)

Pregunta MOB_1B: *Señale todos los países (incluyendo su país de origen) en los que ha estudiado, trabajado o realizado alguna investigación entre enero del 2010 y diciembre de 2019, indicando el periodo de estadía y/o residencia en cada uno de ellos (considere sólo estadías de más de tres meses). Completar en orden cronológico del más antiguo al más reciente.*

Dos entrevistados consideran que se pide mucha información, mucho detalle.

MÓDULO: EXPERIENCIA EN INVESTIGACIÓN (CAR)

Pregunta CAR_8_1: *¿Cuántos meses transcurrieron desde que completó su doctorado hasta que obtuvo su primer empleo en la carrera de investigador?*

Antes de responder, una entrevistada pregunta si el postdoctorado lo consideran o no como un empleo.

MÓDULO: CARACTERÍSTICAS PERSONALES (PER)

Pregunta PER_9: *La Agencia Nacional de Investigación y Desarrollo de Chile (ANID, ex CONICYT) se encuentra en proceso de la elaboración de una plataforma de laboral para doctores y doctoras. ¿Está usted dispuesto(a) a compartir con ANID algunas de sus respuestas?*

Fue un tema importante la privacidad entre los entrevistados, debido al nivel de especificidad de las preguntas. Si bien se leyó que datos serán compartidos, esto no fue suficiente para convencerlos.

La principal duda fue "¿para que ANID necesita estos datos tan específicos?".

Uno de los entrevistados sugirió que, a pesar de esta última pregunta, falta de todas maneras un consentimiento informado al inicio de la entrevista.

### ***Pretest vía telefónica (CATI)***

El levantamiento del pretest por vía telefónica fue llevado a cabo mediante el sistema **Gandia Integra CatiNet** el cual se encarga de la gestión de las llamadas por parte del personal del *call center*, el manejo y control automático de la muestra y la gestión del proceso de supervisión, entre otros. Todo el personal dedicado al levantamiento vía CATI fue capacitado el viernes 14 de agosto, previo al inicio del trabajo de campo.

*Resultados indicadores de respuesta pretest CATI*

El levantamiento del pretest CATI dio inicio el día 17 de agosto de 2020 y finalizó el día 19 de agosto de 2020, habiéndose recolectado un total de 30 encuestas completas. El equipo de levantamiento estuvo conformado por 5 encuestadoras que trabajaron bajo la supervisión del Jefe de Campo.

Se seleccionó una muestra original de 150 casos, de los cuales fueron contactados 130. El protocolo de contacto utilizado en el pretest consistió en llamar a los sujetos de la muestra seleccionada en orden aleatorio hasta completar el número requerido de 30 casos. Se realizaron entre dos y cuatro insistencias por número, en caso de no tener contestación o aplazar la encuesta para más adelante, se procedió a continuar con los llamados aleatorios.

De los 130 teléfonos contactados 75 contestaron las llamadas. De estos 75 casos, dos resultaron ser números de teléfono equivocados o números desactualizados, por ejemplo: casos donde la persona ya no trabajaba en el lugar. Otros 31 casos de los que contestaron las llamadas no tenían disponibilidad para responder la encuesta en el momento del contacto por lo que las encuestas fueron reagendadas si haber iniciado. Diez personas rechazaron la invitación a participar, dos respondieron de forma parcial y 30 lograron completar la encuesta.

El detalle de los resultados de los códigos de disposición final para la muestra seleccionada se muestra en la Tabla 14:

*Tabla 15: Detalle de códigos de disposición final pretest CATI*

| | Original | Intentos | CDF |
|---|---|---|---|
| **Selección** | **150** | | |
| Teléfonos no contactados | 20 | | |
| Teléfonos contactados | 130 | | |
| **Teléfonos contactados** | **130** | | |
| Averiados | 8 | 15 | NC |
| Comunica / Ocupado | 6 | 29 | NC |
| Buzón de voz | 20 | 80 | NC |
| No contesta | 21 | 81 | NC |
| Contesta | 75 | 173 | |
| **Contesta** | **75** | **173** | |
| Número equivocado | 2 | 6 | NC |
| Aplazadas sin iniciar | 31 | 71 | O |
| Encuestas no iniciadas | 10 | 20 | R |
| Encuestas parciales | 2 | 5 | P |
| **Encuestas completas** | **30** | **71** | I |

A partir del pretest se levantaron tasas de respuesta, cooperación, rechazo y contacto que fueron utilizadas para perfeccionar el diseño del trabajo de campo definitivo. Los que se resumen a continuación:

*Tabla 16: Cálculo de indicadores de respuesta AAPOR pretest CATI*

| Nombre de indicador | | Fórmula | Valor |
|---|---|---|---|
| *Tasa de respuesta* | Indicador 1 | $\frac{I}{(I+P)\ +\ (R+NC+O)}$ | 23% |
| | Indicador 2 | $\frac{(I+P)}{(I+P)\ +\ (R+NC+O)}$ | 25% |
| *Tasa de cooperación* | Indicador 1 | $\frac{I}{(I+P)+R+O}$ | 41% |
| | Indicador 2 | $\frac{(I+P)}{((I+P)+R+O))}$ | 44% |
| *Tasa de rechazo* | Indicador 1 | $\frac{R}{((I+P)+(R+NC+O))}$ | 8% |
| *Tasa de contacto* | Indicador 1 | $\frac{(I+P)+R+O}{(I+P)+R+O+NC}$ | 56% |

*<u>Otros resultados del pretest CATI</u>*

La distribución de las duraciones de las encuestas completas se concentró mayoritariamente entre los 36 y 46 minutos (ver gráfico de distribución de las encuestas). Se consideró como valor extremo a un solo caso con duración superior a los 65 minutos. Este caso corresponde una entrevista que no se pudo completar de corrido, retomándola en un momento posterior. Lo casos con valores extremos representan el 3,3% de las encuestas completas.

*Tabla 16: duración de encuestas completas CATI en el pretest*

| Duración | Promedio | Des. Est. | Mediana | N de casos |
|---|---|---|---|---|
| Con valores extremos | 43.3 | 8.7 | 43.6 | 30 |
| Sin valores extremos | 42.4 | 7.5 | 43.6 | 29 |

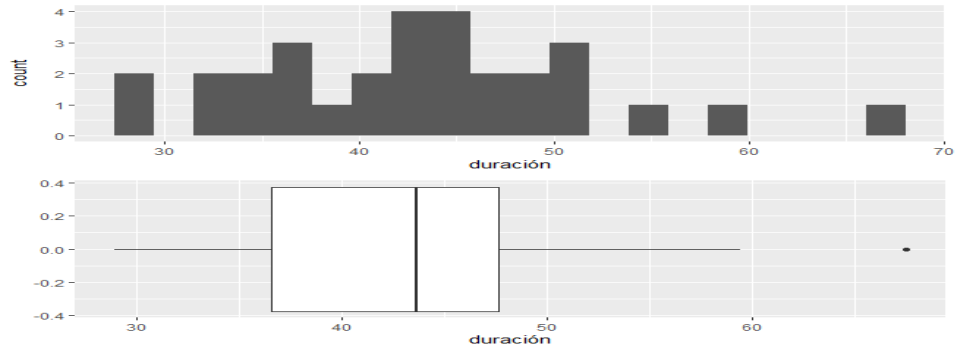



*Figura 3: Distribución de la duración de las encuestas CATI*

*Mesas de discusión de encuestadoras del pretest CATI*

Una vez finalizado el pretest CATI se realizó una mesa de discusión con las encuestadoras. El objetivo de estas entrevistas fue indagar en cómo las encuestadoras percibieron la actitud de los encuestados, qué preguntas consideraron difíciles de entender, cuáles resultaron incómodas. El equipo de levantamiento consiste en cinco encuestadoras con amplia experiencia en estudios de similares características. El método de evaluación del instrumento consistió en una reunión con modalidad de mesa de discusión para recopilar sus experiencias y aprendizajes. En la instancia se trataron tres temáticas: La percepción propia con respecto al instrumento, la percepción de los informantes con respecto al instrumento y aspectos del instrumento a mejorar.

## ***Pretest vía web (CAWI)***

El levantamiento por vía web consistió en un cuestionario auto aplicado mediante la plataforma para encuestas en línea **Survey Monkey**. La invitación a participar del pretest se envió por correo junto con una carta de respaldo del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación.

*Resultados indicadores AAPOR de respuesta pretest CAWI*

El proceso de levantamiento del pretest por el medio de recolección CAWI dio inicio el día 19 de agosto y finalizó el día 28 de agosto. En total se recolectaron 42 encuestas completas. En una primera etapa se seleccionó una muestra original de 200 casos. Se realizó un proceso de validación automática de las direcciones de correo de la muestra, resultando un total de 159 casos con correos válidos a los que se les envió la invitación a participar. De los 159 correos enviados, 118 correos fueron abiertos, de los cuales 67 personas abrieron el enlace hacia la encuesta, 63 iniciaron la encuesta y 38 la contestaron completa. Luego de realizar las cuatro insistencias estipuladas por el protocolo y en vista de que la cantidad de respuestas no aumentaba se procedió a la selección de una muestra de reemplazo de 50 casos. El protocolo de contacto utilizado fue el mismo que el para la muestra original. Los códigos de disposición final resultantes se muestran en la siguiente:

*Tabla 17: Detalle de códigos de disposición final pretest CAWI*

| | Original | Reemplazo | Total | CDF |
|---|---|---|---|---|
| **Selección** | **200** | **50** | **250** | |
| Correos inválidos | 41 | 14 | 55 | NC |
| Correos válidos | 159 | 36 | 195 | |
| **Correos válidos enviados** | **159** | **36** | **195** | |
| Correos no abiertos | 41 | 19 | 60 | R |
| Correos abiertos | 118 | 17 | 135 | |
| **Correos abiertos** | **118** | **17** | **135** | |
| Enlaces no abiertos | 51 | 12 | 63 | R |
| Enlaces abiertos | 67 | 5 | 72 | |
| **Enlace abierto** | **67** | **5** | **72** | |
| Encuestas no iniciadas | 4 | 1 | 5 | R |
| Encuestas parciales | 25 | 0 | 25 | P |
| **Encuestas completas** | **38** | **4** | **42** | I |

A partir del pretest CAWI se obtuvieron las siguientes tasas de respuesta, cooperación, rechazo y contacto:

*Tabla 18: Cálculo de indicadores de respuesta AAPOR pretest CAWI*

| Nombre de indicador | | Fórmula | Valor |
|---|---|---|---|
| *Tasa de respuesta* | Indicador 1 | $\frac{I}{(I+P)\ +\ (R+NC+O)}$ | 17% |
| | Indicador 2 | $\frac{(I+P)}{(I+P)\ +\ (R+NC+O)}$ | 27% |
| *Tasa de cooperación* | Indicador 1 | $\frac{I}{(I+P)+R+O}$ | 22% |
| | Indicador 2 | $\frac{(I+P)}{((I+P)+R+O))}$ | 34% |
| *Tasa de rechazo* | Indicador 1 | $\frac{R}{((I+P)+(R+NC+O))}$ | 51% |
| *Tasa de contacto* | Indicador 1 | $\frac{(I+P)+R+O}{(I+P)+R+O+NC}$ | 78% |

*<u>Otros resultados del pretest CAWI</u>*

La distribución de las duraciones de las encuestas completas se concentró mayoritariamente entre los 25 y 50 minutos. Se consideró valores extremos los casos con duraciones superiores a los 100 minutos. Estos casos probablemente corresponden a personas que comenzaron la encuesta y no la pudieron completar de corrido, retomándola en un momento posterior. Lo casos con valores extremos representan el 16,7% de las encuestas completas

Excluyendo los valores extremos se tiene que la duración promedio de la encuesta cuando se responde de corrido es de 33.2 minutos con una desviación estándar de 12.4 minutos.

*Tabla 19: Tabla de duración de encuestas completas CAWI en el pretest*

| Duración | Promedio | Des. Est. | Mediana | N de casos |
|---|---|---|---|---|
| Con valores extremos | 59,8 | 71,3 | 32,7 | 42 |
| Sin valores extremos | 33,2 | 12,4 | 31,7 | 35 |

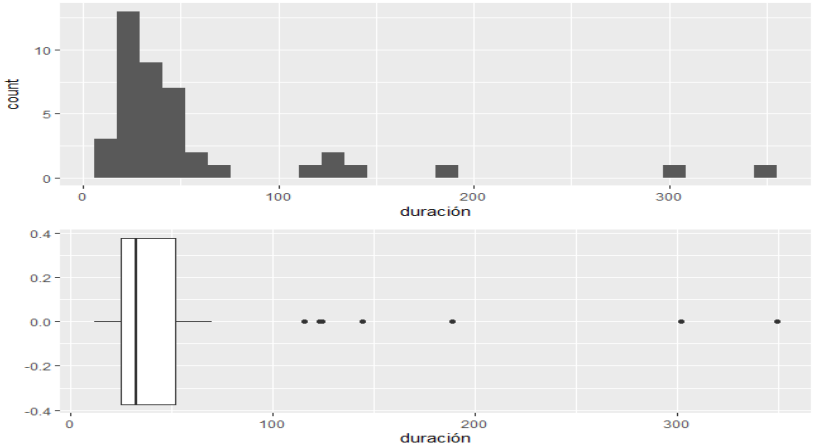



*Figura 4: Distribución de la duración de las encuestas*

## 4.3 Definición del Instrumento final

A partir de los aprendizajes logrados durante el proceso de validación del instrumento, del procesamiento de los insumos anteriores y del trabajo conjunto con la contraparte técnica, se llegó a **una versión final del cuestionario publicado en la web del Ministerio de Ciencia** (se entregaron más detalles a la contraparte en el "Informe Previo de Trabajo de Campo", producto entregado en el marco de este mismo contrato).

Debido a que las entrevistas cognitivas se llevaron a cabo antes del levantamiento de los pretest, algunos aprendizajes de la primera etapa de validación se tradujeron en mejoras al cuestionario previo a que se apliquen las pruebas del CATI y CAWI. Esto se dio específicamente con observaciones relevantes relacionadas con la privacidad de los datos y las preguntas sobre autorización de uso de información por parte del Ministerio de Ciencias, Tecnología, Conocimiento e Innovación en las entrevistas. Dichas observaciones se concretaron en el acuerdo con la Contraparte Técnica de incorporar modificaciones en la estructura del cuestionario para mejorar el aspecto de la percepción de privacidad y transparencia, del cuestionario.

Estas modificaciones consistieron en agregar un séptimo módulo sobre la autorización de uso de datos, en el que hay dos preguntas de autorización de uso de datos[14] además de datos de identificación (nombres y apellidos) y contacto (correo electrónico y número de teléfono). La incorporación de estas preguntas también significó algunos días de retraso en la programación. La última sugerencia respecto a este módulo fue la incorporación de la pregunta sobre el RUT, la cual se realizó el día de la capacitación y debió ser agregada a la programación el día 18 de agosto. Producto de este atraso no se logró validar completamente el medio de recolección en su programación final.

Por otra parte, los resultados del pretest CATI y CAWI y los comentarios realizados en las entrevistas cognitivas revelaron que la duración de la encuesta implementada era percibida como excesivamente larga, tanto por los encuestados como por el equipo encuestador. Debido al elevado tiempo de aplicación del cuestionario, una de las principales modificaciones que se le realizaron fue dejar afuera algunas preguntas no obligatorias con el fin de reducir el tiempo de aplicación a valores aceptables. Se mantuvieron todas las preguntas centrales necesarias para generar las estadísticas CDH y el reporte de tablas de resultados, además de algunas preguntas opcionales para el reporte a la OCDE pero de alto interés para el ministerio.

También se recibió desde la contraparte técnica la solicitud de agregar un nuevo módulo sobre brechas de género, el cual depende del rol del encuestado respecto a la toma de decisiones de contratación de personal. Debido a que el módulo de Género se recibió muy cerca de la fecha programada para el inicio del levantamiento no se puedo realizar nuevas pruebas para evaluar la duración de la encuesta.

14 Preguntas del módulo: autorización de uso de datos (DAT): **DAT_1_1** ¿Estaría usted dispuesto a que el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación pueda unir sus datos entregados en esta encuesta con otros registros administrativos con el único fin de realizar análisis internos, resguardando siempre el anonimato de los mismos?

## 4.4 Definición de protocolos de levantamiento

### 4.4.1 Antecedentes del proceso de definición de los protocolos de levantamiento

Al igual como ocurrió con el proceso de validación del cuestionario, los protocolos de levantamiento experimentaron varios cambios desde su definición inicial. Esto se debe en gran parte a los aprendizajes que se obtuvieron durante las actividades de validación del cuestionario. En primera instancia la propuesta técnica describía el protocolo de levantamiento como un proceso secuencial en el que el primer contacto se realizaba de manera telefónica y la recolección vía CAWI se reservaba para realizar seguimientos, insistencias y gestión de reemplazos.

Una vez que se dio inicio al proceso de validación se observó, a partir de los resultados del pretest, que el medio CAWI tiene una mayor tasa de contacto, y una menor tasa de respuesta que el medio CATI. A partir de estos resultados se concluye que utilizar el sistema CAWI como medio complementario y de soporte desaprovecha la capacidad de contacto masivo de este medio y que el mejor desempeño del medio CATI en términos de tasa de respuesta le da una ventaja con respecto al medio CAWI para su uso como contacto secundario y de seguimiento. Esto fundamenta la decisión de posicionar el levantamiento CAWI como método de recolección principal y el medio CATI como secundario, o sea, para ser aplicado una vez que se agoten las insistencias mediante el levantamiento CAWI.

Por otra parte, los resultados de la mesa de discusión del pretest realizado a 30 personas revelaron que la entrega de información sobre el estudio a los informantes, de forma previa al levantamiento telefónico tiene efectos positivos en la disposición a contestar la encuesta.

Por sugerencia de la contraparte técnica, se definió el envío de una primera carta de invitación que dé inicio al levantamiento en la que se proporcione información relevante sobre el estudio y que al mismo tiempo vaya acompañada con un enlace parapara responder la encuesta CAWI. De esta manera se aprovechan los beneficios de la entrega de información previa en conjunto con la capacidad de contacto masivo del sistema SurveyMonkey.

Este cambio de enfoque trajo consigo el desafío de coordinar la actualización de las bases de datos de ambos medios de recolección, lo cual es un aspecto clave para evitar intentos de contacto redundantes y respuestas duplicadas. Durante el proceso de transición se exploraron distintas alternativas para formalizar los mecanismos de coordinación entre CATI y CAWI, algunas de ellas se descartaron por su complejidad priorizando un modelo más flexible que permitiera a los doctores y doctoras responder por medio que le pareciera más cómodo. Algunas personas solicitaban el agendamiento de la entrevista durante los próximos días, mientras que otras manifestaban preferir contestar vía web.

A continuación, se describen brevemente los protocolos finales.

### 4.4.2 Protocolo aplicado en levantamiento CAWI

La construcción de la base de datos CAWI se realizó seleccionando los registros de la muestra seleccionada aleatoriamente con información de contacto que permitan el envío de la invitación para contestar vía web. La validación de los correos electrónicos se realizó de forma automática utilizando el software especializado Atomic Mail Verifier.

El contacto por vía CAWI contempló el envío de un correo de invitación inicial entregando información relevante sobre el estudio de Trayectoria de Doctores en Chile. Una vez recibido el correo, en caso de que la encuesta fuese iniciada y completada, el registro sería actualizado en la base de datos CAWI.

En el caso de que la encuesta no fuese respondida de forma completa se realizarían hasta 5 intentos de contacto por envío de recordatorios mediante correo electrónico. Esto quiere decir que sólo salen de la lista de correo aquellos que llegan hasta el final de la encuesta, y los recordatorios se siguen enviando en los siguientes casos:

- No abren el correo
- Abren el correo, no ingresan al enlace
- Abren el correo, ingresan al enlace, no inician la encuesta
- Abren el correo, ingresan al enlace, responden parcialmente

Estos recordatorios se realizaron con una frecuencia de 3 veces por semana: los días lunes, miércoles y viernes. La decisión de establecer un margen de dos días entre recordatorios en una misma semana se basó en la experiencia del estudio de Actualización de Directorio de Doctores. En dicho estudio las personas respondían mayoritariamente entre el primer y segundo día después de enviada la invitación (primer correo) y recordatorios (siguientes correos). Previo a cada envío de recordatorios se actualizó la base de datos de CAWI. Un ejemplo de carta de invitación se incluye en el Anexo 3.

La clasificación de los códigos de disposición final del sistema CAWI acordada con la contraparte técnica por medio del formato de los reportes de avance semanal se resume en la Tabla 20. Se consideran “Sin contacto” (NC) los casos en que los correos sean inválidos, o los que reciben el correo, pero no lo abren. Los casos considerados como “Rechazo” (R) corresponden a aquellos en que se dieron de baja (cancelar la suscripción a la encuesta), aquellos que abrieron el correo, pero no abrieron el enlace, y aquellos que abrieron el enlace, pero no iniciaron la encuesta. Las encuestas fuera de muestras se clasificaron como “Otros” (O) y las encuestas sin completar, como “Encuesta Parcial” (P). Las encuestas completas se clasifican como “Entrevista completa” (I).

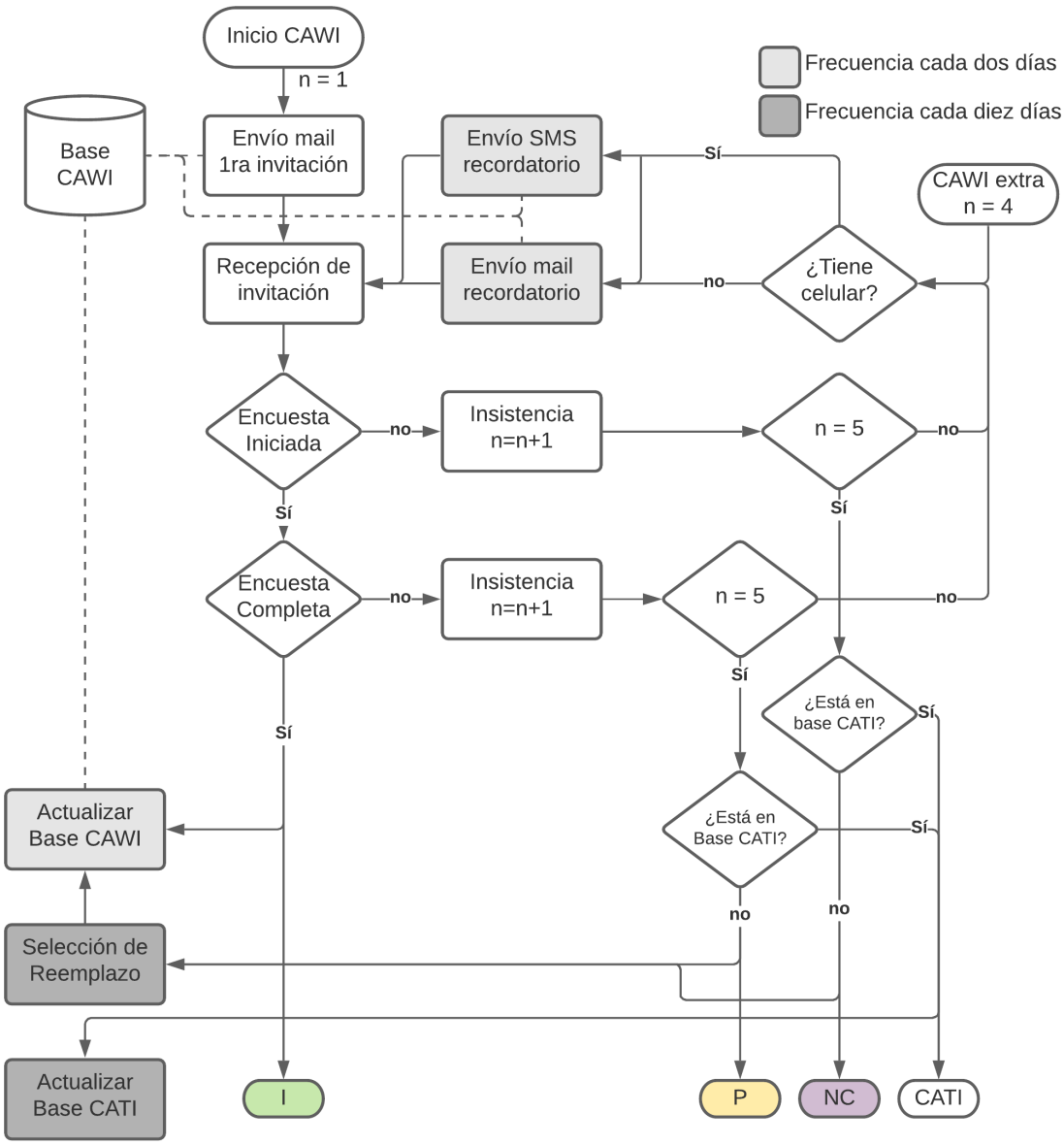
*Tabla 20: Códigos de disposición final CAWI*

| Tipo de contacto | CDF |
|---|---|
| Correos inválidos | NC |
| Dados de baja | R |
| Correos no abiertos | NC |
| Enlaces no abiertos | R |
| Encuestas no iniciadas | R |
| Encuestas parciales | P |

| | |
|---|---|
| Encuestas fuera de muestra | O |
| Encuestas dentro de muestra | I |

El siguiente diagrama sintetiza el flujo de toma de decisiones del protocolo de contacto CAWI. Es importante considerar que se deja en forma explícita en el diagrama la secuencialidad CAWI -> CATI en el proceso de contacto. Adicionalmente, se envió una carta de invitación cuyo formato se incluye en el Anexo 3.

*Figura 5: Esquema protocolo de contacto y levantamiento CAWI*



### 4.4.3 Protocolo aplicado en levantamiento CATI

La base de datos CATI se construyó seleccionando los casos de la muestra que tienen disponibilidad de datos de contacto telefónico válidos para la realización de la encuesta por medio del sistema CATI. La validación de los números de teléfono se realizó mediante el validador de números de teléfono incorporado a la central telefónica de Datavoz.

Una vez realizados los 5 intentos de contacto de la etapa CAWI, el equipo de levantamiento comenzó a realizar llamados asistido por el software Gandia Integra CatiNet, el cual gestiona el marcado automático, la selección de la muestra, los códigos de disposición final entre otras características.En caso de no lograr contacto, se consideraron hasta 5 llamados telefónicos.

De igual manera que para el levantamiento CAWI, la clasificación de los códigos de disposición final fue acordada con la contraparte técnica por medio de la tabla de reporte de avance semanal, tal como se indica en la Tabla 21. Se consideran como "Sin contacto" (NC) los casos de números de teléfonos averiados, siempre ocupados, no contesta, con contestador automático, número equivocado, y aquellos casos en que se solicite aplazamiento, pero no se pudo concretar la entrevista. Los casos en que la persona contactada se negó a participar se consideran "Rechazo" (R). Los casos en que la encuesta no se inició por otros motivos, por ejemplo, prefiere responder vía web, se clasifica como "Otros" (O). Las encuestas incompletas se clasifican como "Encuesta Parcial" (P) y las encuestas completas como "Entrevista completa" (I).

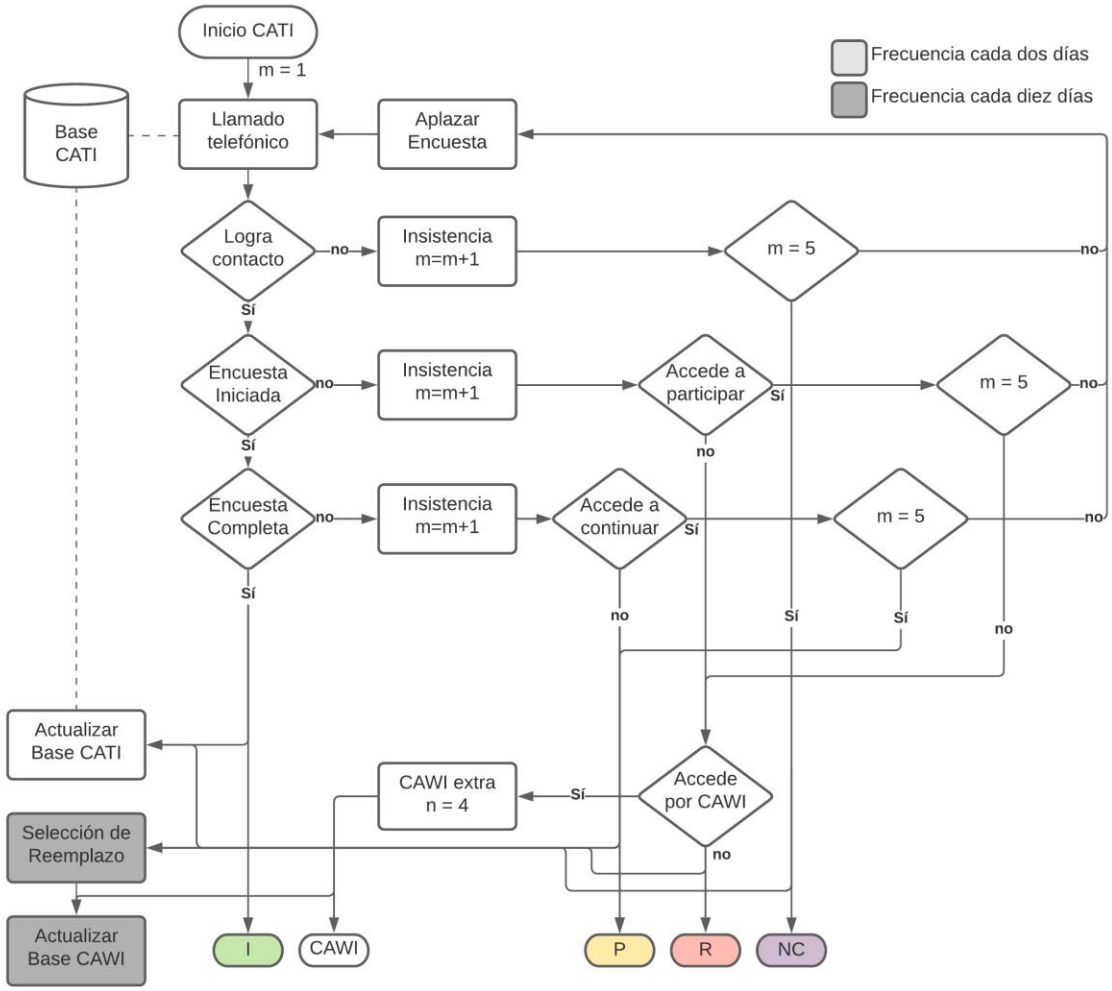
*Tabla 21: Códigos de disposición final CATI*

| Tipo de contacto | CDF |
|---|---|
| Averiados | NC |
| Comunica / Ocupado / No contesta | NC |
| Contestador | NC |
| Número equivocado | NC |
| Aplazadas sin iniciar | NC |
| Encuestas no iniciadas (rechazos) | R |
| Encuestas no iniciadas (otros) | O |
| Encuestas parciales | P |
| Encuestas dentro de muestra | I |

El siguiente diagrama resume el protocolo de contacto CATI.[15]

15 En la segunda línea del diagrama de flujo aparece el concepto "Aplazar". Este concepto significa que la llamada es pospuesta. En caso de que no se logre contacto el software Gandia gestiona el momento en que el número vuelve a aparecer en el sistema. En caso de que no se haya logrado contacto el tiempo mínimo de espera antes de un nuevo llamado es de un día. En caso de que se haya logrado contacto, el sistema permite agendar el día y hora de aplazamiento de la llamada, de esta manera el número aparecerá en el sistema el día y hora de la cita programada. Lo mismo ocurre si la encuesta quedó respondida parcialmente.

*Figura 6: Esquema protocolo de contacto y levantamiento CATI*



# 5. Resultados del trabajo de campo

## 5.1 Medidas de gestión operativa

El desarrollo de todo trabajo de campo requiere de medidas de coordinación para velar por la correcta aplicación de los protocolos, la continuidad de los avances y la calidad de los datos recolectados. En el caso de la encuesta CDH, debido a que el trabajo de campo considera la combinación de dos medios de recolección, estas medidas cobran una gran relevancia para mantener ambos sistemas actualizados y evitar situaciones indeseadas como insistencias a personas que ya han respondido, respuestas duplicadas, y registros erróneos de códigos de disposición final.

A continuación, se detallan las principales medidas gestión operativa que se llevaron a cabo durante el desarrollo del levantamiento.

*Actualización de bases de datos*

Los sistemas de recolección CAWI y CATI dependen de grandes bases de datos que se mantuvieron actualizadas en dos formas:

• Actualización interna: El estado de los contactos en la base se actualizó con sus respectivos códigos internos en cada sistema de forma independiente. Este tipo de actualización se encuentra integrado en ambos sistemas CAWI y CATI, y no requiere de mayor intervención. En el sistema CAWI la plataforma SurveyMonkey permite el envío de recordatorios sólo a personas que no hayan completado la encuesta aún. En el sistema CATI el sistema Gandia Catinet permite gestionar las llamadas ocupadas y las entrevistas aplazadas para su contacto posterior, y a la vez permite deshabilitar los contactos que han rechazado su participación, han completado el total de intentos de contacto o han completado la encuesta de forma exitosa.

• Actualización cruzada: El estado de los contactos en la base se actualiza incorporando información de emitidos por un sistema externo. Este tipo de actualización no ocurre de forma automática por lo que requiere de que el equipo de coordinación de levantamiento consolide la información y la integre a cada sistema periódicamente. Este tipo de actualización se lleva a cabo en dos sentidos: desde el sistema CATI al sistema CAWI, y desde el sistema CAWI al sistema CATI. La actualización en el sentido CATI a CAWI tiene un nivel de complejidad relativamente bajo debido a que el sistema CATI solo envía recordatorios con una frecuencia de dos a tres días y los llamados CATI solo se les da inicio una vez enviados los cinco correos de invitación y recordatorios. La actualización en el sentido contrario, CAWI a CATI, es más compleja y demandante ya que incluso una vez finalizado el envío de recordatorios CAWI, se siguen recibiendo respuestas tardías, y personas que, al ser contactadas telefónicamente, prefieren contestar vía web.

El proceso de actualización cruzada enfrentó dificultades de coordinación en el sentido de CAWI a CATI. Principalmente esto se debió a que las actualizaciones no son instantáneas y tienen tiempos de ejecución que generan diferencias entre ambos sistemas, dejando periodos donde hay respuestas que no son incorporadas a CATI hasta la siguiente actualización. Este problema se acentuó con mayor fuerza los días posteriores a feriados y fines de semana.

Las medidas tomadas para compensar estas dificultades incluyeron aumentar la frecuencia de las actualizaciones incluyendo los días sábado, y regularizar la hora en que se lleva a cabo el

proceso de actualización, velando que se haga lo más temprano posible para disminuir la cantidad de respuesta desfasadas.

*Reportes semanales*

El contenido de los reportes semanales fue acordado en conjunto con la contraparte técnica, e incluye reportes separados para el levantamiento CAWI y CATI, y también un reporte total que condensa los avances realizados mediante los dos medios de recolección de forma agregada.

La estructura del reporte CAWI y CATI fue básicamente el mismo y estuvo compuesto por tres tablas:

- Detalle de códigos de disposición final por tipo de muestra (original y reemplazo)
- Resumen de códigos de disposición final
- Cálculo de indicadores de respuesta AAPOR

Adicionalmente, el reporte contenía una sección de comentarios donde se registran las observaciones relevantes con respecto a los resultados obtenidos, el manejo de las muestras y la aplicación de los protocolos. La observación constante de los resultados del levantamiento permitió tomar decisiones de forma oportuna.

## 5.2 Decisión de expansión de la muestra

Los resultados observados en el primer reporte de avance, correspondiente al 26 de noviembre, muestran que el levantamiento CAWI registra una tasa de respuesta considerablemente superior a lo observado en el pretest y en otros estudios similares, alcanzando un 30%. Se hipotetiza que esta diferencia se debe a las diferencias en el tamaño de la muestra. A la fecha del primer reporte aún no se tenían resultados del levantamiento CATI.

En el segundo reporte, correspondiente al 2 de noviembre, se observó que la tasa de respuesta del levantamiento CAWI se mantenía estable sobre el 30% y a partir de los primeros resultados del levantamiento CATI se registró una tasa de respuesta de un 21%, un valor similar al observado en el levantamiento piloto (23%). Los comentarios recibidos durante la primera semana de levantamiento telefónico sugieren que, a pesar de la reducción del tiempo de ejecución del cuestionario, este se sigue percibiendo como extenso y los encuestados manifiestan en varias ocasiones poca disponibilidad de tiempo y preferencia por responder vía web.

Durante las dos primeras semanas de levantamiento se obtuvieron los siguientes aprendizajes: la tasa de respuesta del medio CAWI es mayor que el obtenido mediante CATI, la población objetivo tiene poca disponibilidad de tiempo lo que se traduce en una alta incidencia de agendamiento (aplazamiento de la entrevista a una fecha y hora posterior) y parte importante de los encuestados manifestaron preferencia por contestar vía web.

Tomando en cuenta los antecedentes recopilados durante las dos primeras semanas del levantamiento, se previó un importante aumento del tiempo de implementación respecto al planificado, producto del mayor tiempo requerido para obtener logros en el medio de recolección CATI. Para evitar los retrasos previstos se tomó la decisión, en conjunto con la contraparte técnica, de proyectar con base en los valores de las tasas de respuesta la cantidad de reemplazos necesarios

para cumplir con el objetivo de 2.400 casos y de esta manera reducir la cantidad de ciclos de reemplazo.

Utilizando los valores de las tasas de respuestas de los medios CAWI y CATI observados en los reportes de avance, se estimó que el tamaño de la selección de casos de reemplazo debiese ser igual a 4320 lo que equivale al 90% del total de casos disponibles en la muestra de reemplazo (N = 4800). En consecuencia, se adaptó la programación del trabajo en terreno para aplicar el protocolo de contacto a la muestra de reemplazo en paralelo a la muestra principal. De esta manera, mientras la muestra original aún se encontraba en su fase de levantamiento CATI, se dio inicio a la fase de levantamiento CAWI de la muestra de reemplazo proyectada.

Esta modificación se analizó desde el punto de vista estadístico y se concluyó que no generaría distorsiones en la probabilidad de selección de los encuestados por dos razones: La primera es que ambas muestran (original y reemplazo) fueron seleccionadas mediante el mismo proceso aleatorio; la segunda es que todos los casos serán sometidos al mismo protocolo por lo que esfuerzos de contacto son iguales para todos.

## 5.3 Seguimiento de las encuestas por modo de recolección

*Seguimiento de modalidad CAWI*

El levantamiento de la modalidad CAWI se dio inicio el día 16 de octubre y se mantuvo operativo hasta el día 10 de diciembre, fecha en que se cerró el enlace a la encuesta online.

El levantamiento inició con el contacto a la muestra original completando los 5 envíos (una primera invitación y cuatro recordatorios) el día 26 de octubre. Luego de esto no se enviaron nuevos recordatorios, pero se dejó abierto el enlace a la encuesta en caso de que los doctores contactados vía CAWI prefirieran contestar vía web.

El contacto a la muestra de reemplazo inició el día 6 de noviembre con el envío de la primera invitación y terminó con el envío del último recordatorio el día 17 de noviembre. Al igual que en el caso de la muestra original, una vez se hubiera completado el ciclo de envíos de recordatorios con la frecuencia establecida en el protocolo el enlace se dejó abierto por un periodo más largo para permitir a los doctores y doctoras que así lo prefirieran poder contestar vía web. Ambos enlaces se cerraron el día 10 de diciembre (para una comparación del levantamiento generado mediante CAWI y CATI, véase la sección 5.5).

Los resultados obtenidos hasta la fecha de cierre ascienden a 1909 encuestas completas por medio del levantamiento web, las que se componen por 693 respuestas provenientes de la muestra original y 1216 de la muestra de reemplazo.

Los códigos internos del sistema CAWI ubicados en la primera columna de la Tabla 22, corresponden a indicadores agregados que entrega el sistema SurveyMonkey combinados con información desagregada a partir de las encuestas completadas que permite cuantificar el número de encuestas fuera y dentro de la muestra, y que autorizan o no el uso de datos. Es importante recordar que el muestreo utilizado es aleatorio simple dentro de los estratos definidos por género (para detalles véase la sección 3.5). Por este motivo, dado que no todas las personas en el directorio tienen correcta información de contacto y que el tipo de muestreo protege el poder estadístico definiendo que toda persona seleccionada haya tenido la misma probabilidad de ser encuestada,

por construcción, el total de contactos posibles es menor al total seleccionado en la muestra original y la muestra de reemplazo.

*Tabla 22: Códigos de disposición final CAWI por tipo de muestra*

| Código Interno CAWI | Original | Reemplazo | Total | % | CDF |
|---|---|---|---|---|---|
| **Correos enviados** | **1943** | **3509** | **5419** | | |
| Correos inválidos | 33 | 97 | 130 | 2% | NC |
| Correos válidos | 1910 | 3412 | 5322 | 98% | |
| **Correos válidos enviados** | **1910** | **3412** | **5322** | **98%** | |
| Dados de baja | 9 | 44 | 53 | 2% | R |
| Correos no abiertos | 482 | 857 | 1341 | 25% | NC |
| Correos abiertos | 1419 | 2511 | 3928 | 74% | |
| **Correos abiertos** | **1419** | **2511** | **3928** | **74%** | |
| Enlaces no abiertos | 487 | 799 | 1286 | 33% | R |
| Enlaces abiertos | 930 | 1712 | 2642 | 67% | |
| **Enlaces abiertos** | **930** | **1712** | **2642** | **67%** | |
| Encuestas no iniciadas | 38 | 75 | 113 | 6% | R |
| Encuestas parciales | 173 | 377 | 550 | 21% | P |
| Encuestas completas | 719 | 1260 | 1979 | 75% | |
| **Encuestas completas** | **719** | **1260** | **1979** | **75%** | |
| Encuestas fuera de muestra | 28 | 51 | 79 | 4% | O |
| Encuestas dentro de muestra | 691 | 1209 | 1900 | 96% | I |
| **Encuestas dentro de muestra** | **693** | **1216** | **1909** | **96%** | |
| Autorizaron uso de datos | 513 | 859 | 1372 | 72% | |
| No autorizaron uso de datos | 178 | 350 | 528 | 28% | |

La Tabla 23 muestra el resumen de la información de acuerdo a los códigos de disposición final establecidos por la AAPOR para el posterior cálculo de los indicadores de respuesta.

*Tabla 23: Resumen códigos de disposición final CAWI*

| Código | Definición | Valor |
|---|---|---|
| *I* | entrevistas completas | 1909 |
| *P* | entrevistas parciales | 288 |
| *R* | rechazos y *break off* | 1738 |
| *NC* | Sin contacto (*no contact*) | 1484 |
| *O* | otros | 777 |
| | Total | 5419 |

*Seguimiento de modalidad CATI*

El levantamiento de la modalidad CATI se dio inicio el día 26 de octubre, fecha para la cual se había enviado el último recordatorio del levantamiento CAWI de la muestra original. Los llamados a la muestra original concluyeron el día 17 de noviembre

El día 18 de noviembre, una vez finalizado el contacto telefónico a la muestra original y el envío de correos a la muestra de reemplazo, se dio inicio el contacto telefónico a las personas en la muestra de reemplazo. Este proceso se extendió hasta el día 4 de diciembre.

Los resultados obtenidos hasta la fecha registrada ascienden a 224 encuestas completas pertenecientes a la muestra original. Al igual que en la modalidad CAWI, el sistema de Gandia Catinet posee códigos internos para la gestión operativa, los cuales combinados con información desagregada a partir de las encuestas completadas que permite cuantificar el número de encuestas fuera y dentro de la muestra, y que autorizan o no el uso de datos.

*Tabla 24: Códigos de disposición final CATI por tipo de muestra*

| Código Interno CATI | Original | Reemplazo | Total | % | CDF |
|---|---|---|---|---|---|
| **Teléfonos contactados** | **994** | **1991** | **2985** | | |
| Averiados | 89 | 78 | 167 | 6% | NC |
| Comunica / Ocupado / No contesta | 278 | 694 | 972 | 33% | NC |
| Contestador | 134 | 354 | 488 | 16% | NC |
| Contesta | 584 | 865 | 1449 | 49% | |
| **Contesta** | **584** | **865** | **1449** | **49%** | |
| Número equivocado | 29 | 20 | 49 | 3% | NC |
| Aplazadas sin iniciar | 0 | 0 | 0 | 0% | NC |
| Encuestas no iniciadas (rechazos) | 78 | 49 | 127 | 9% | R |
| Encuestas no iniciadas (otros) | 112 | 472 | 584 | 40% | O |
| Encuestas parciales | 17 | 7 | 24 | 2% | P |
| Encuestas completas | 257 | 317 | 574 | 40% | |
| **Encuestas completas** | **257** | **317** | **574** | **40%** | |
| Encuestas fuera de muestra | 12 | 10 | 22 | 4% | O |
| Encuestas dentro de muestra | 245 | 307 | 552 | 96% | I |
| **Encuestas dentro de muestra** | **245** | **307** | **552** | **96%** | |
| Autorizaron uso de datos | 220 | 277 | 497 | 90% | |
| No autorizaron uso de datos | 25 | 30 | 55 | 10% | |

La Tabla 25 muestra el resumen de los códigos de disposición final dispuestos por la AAPOR para el cálculo de los indicadores de respuesta.

*Tabla 25: Resumen códigos de disposición final CATI*

| Código | Definición | Valor |
|---|---|---|
| *I* | entrevistas completas | 552 |
| *P* | entrevistas parciales | 24 |
| *R* | rechazos y *break off* | 127 |
| *NC* | Sin contacto (*no contact*) | 1676 |
| *O* | otros | 584 |
| | Total | 2963 |

## 5.4 Códigos de disposición final e indicadores de respuesta

A partir de las respuestas al levantamiento CATI y CAWI se calculan los códigos de disposición final para el total de la muestra agregando ambos métodos de levantamiento. En la Tabla 26 se puede apreciar que el método CAWI, a pesar de tener una mayor tasa de respuesta, la cantidad de entrevistas parciales es también mucho mayor que en el caso CATI. Esto se puede explicar por el rol de los encuestadores y encuestadoras en mantener la atención del encuestado y/o en caso de ser necesario continuar y completar la encuesta en otro momento, seguimiento que en el sistema CAWI es algo que queda a decisión del encuestado.

*Tabla 26: Resumen de códigos de disposición final total*

| Código | Definición | CAWI | CATI | Total |
|---|---|---|---|---|
| *I* | entrevistas completas | 1.909 | 552 | 2.461 |
| *P* | entrevistas parciales | 288 | 24 | 312 |
| *R* | rechazos y *break off* | 1.738 | 127 | 1.865 |
| *NC* | Sin contacto (*no contact*) | 1.484 | 1.676 | 3.160 |
| *O* | otros | 77 | 584 | 661 |

De la misma manera que para los indicadores de respuesta de cada medio de recolección, el cálculo de los indicadores de respuesta de la total se realiza de acuerdo a las fórmulas de la AAPOR disponibles en el Anexo 2. Con respecto a los resultados obtenidos durante el pretest y durante el levantamiento, se observa que las tasas de respuesta obtenidas durante el levantamiento son considerablemente mayores que las obtenidas durante el pretest. En el caso del pretest CATI la hipótesis principal detrás de la diferencia es la extensión del cuestionario la que podría haber tenido un impacto negativo en la cantidad de entrevistas completas. También podría haber tenido un efecto la novedad del instrumento para el equipo de terreno, la familiarización del instrumento durante el levantamiento podría haber favorecido el desempeño de los encuestadores.

Con respecto a las diferencias en CAWI es posible que se deba a las diferencias de tamaño de las muestras, ya que en el pretest el logro proyectado era solo de 30 casos. Además, el instrumento utilizado durante el pretest, era también más extenso. Otro factor que pudo haber influido positivamente la tasa de respuesta en CAWI durante el levantamiento es el de los casos CATI que prefirieron contestar vía web, opción que no se encontraba disponible durante el pretest.

*Tabla 27: Indicadores de respuesta totales*

| Nombre de indicador | | Pretest | | Levantamiento | | Total |
|---|---|---|---|---|---|---|
| | | CAWI | CATI | CAWI | CATI | |
| *Tasa de respuesta* | Indicador 1 | 17% | 16% | 35% | 19% | 29% |
| | Indicador 2 | 27% | 17% | 40% | 19% | 33% |
| *Tasa de cooperación* | Indicador 1 | 22% | 43% | 48% | 43% | 46% |
| | Indicador 2 | 34% | 45% | 55% | 45% | 52% |
| *Tasa de rechazo* | Indicador 1 | 51% | 4% | 32% | 4% | 22% |
| *Tasa de contacto* | Indicador 1 | 78% | 38% | 73% | 43% | 63% |

Del total de respuestas, un 76% autorizaron el uso de datos para su cruce con otras fuentes de información internas del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, tal como se observa en la Tabla 28. Es posible notar además la considerable diferencia entre los resultados entre CATI y CAWI, siendo el medio telefónico el que logró un mayor porcentaje de autorizaciones (90%) que el medio web (72%). Respecto a esto cabe destacar la labor de los encuestadores y las encuestadoras en proporcionar información clara y construir la confianza necesaria.

*Tabla 28: Autorización de datos*

| Definición | Valor | CAWI | CATI | Valor(%) |
|---|---|---|---|---|
| Autorizaron uso de datos | 1.876 | 72% | 90% | 76% |
| No autorizaron uso de datos | 585 | 28% | 10% | 24% |
| Total | 2.461 | 100% | 100% | 100% |

## 5.5 Evolución del levantamiento

Desde el inicio del trabajo de campo con levantamiento vía CAWI es posible ver el efecto del envío de las invitaciones y recordatorios en la cantidad de respuesta diarias. Tal como muestra la **¡Error! No se encuentra el origen de la referencia.** las fechas de envío de recordatorios concentran u na gran cantidad de respuestas las que van decreciendo en el tiempo en la medida que se agota la muestra original. El mismo patrón se observa en el envío de invitaciones y recordatorios para la selección de la muestra de reemplazo.

La dinámica de las respuestas CATI es mucho más estable pues depende de los esfuerzos de contacto realizados por el equipo de levantamiento, los cuales se distribuyen de forma más homogénea en el tiempo, con excepción de las interrupciones observadas en los fines de semana.

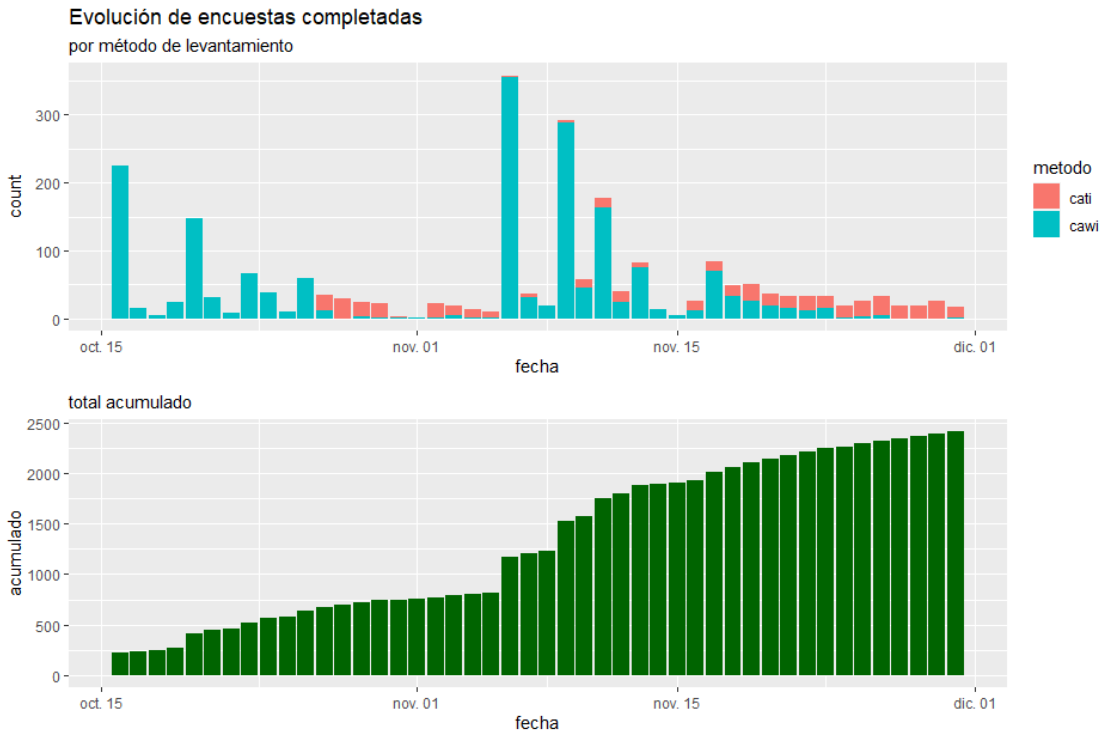



*Figura 7: Evolución de encuestas completadas*

Se puede observar también que una vez finalizado el envío de recordatorios y dado inicio al método CATI la cantidad de respuestas CAWI no llega a cero. Esto se puede deber a las personas que al ser contactadas por vía telefónica manifiestan su preferencia de responder vía web. En el gráfico del total acumulado se puede observar que el envío de recordatorio genera un aumento escalonado que luego se sostiene y tiende a estabilizarse.

## 5.6 Supervisión y acompañamiento

*Supervisión modalidad CATI*

La supervisión del medio de recolección CATI consistió en el contacto vía telefónica de las personas correspondientes al 20,3% de las encuestas realizadas de cada encuestador (112 controles en total) y encuestadora para verificar la aplicación de la encuesta y corroborar los datos registrados en 5 preguntas del cuestionario.

El proceso de supervisión se dio inicio una vez que se completó el total de llamados correspondientes a la muestra original, lo cual ocurrió en la tercera semana de noviembre. Estos controles de supervisión se distribuyen entre los encuestadores de forma proporcional a las encuestas realizadas por cada uno.

Los resultados resumidos en la Tabla 29 muestran que la totalidad de los casos supervisados fueron respondidos de forma efectiva y las respuestas obtenidas coinciden con las registradas en la base de datos. Adicionalmente se observa que valoran con un alto puntaje los ítems de información entregada y el trato del encuestador (a).

*Tabla 29: Resumen supervisión llamadas telefónicas*

| Pregunta | Opciones | Número | Porcentaje |
|---|---|---|---|
| P1. ¿El número de teléfono corresponde a la persona encuestada? | Sí | 112 | 100% |
| | No | 0 | 0% |
| P2. ¿La persona encuestada está disponible para responder preguntas de supervisión? | Sí | 112 | 100% |
| | No | 0 | 0% |
| P3. ¿Fue usted contactado/a para un estudio sobre trayectoria de profesionales con grado de doctor? | Sí, respondí la encuesta completa | 112 | 100% |
| | Sí, pero respondí parcialmente | 0 | 0% |
| | Sí, pero no la contesté | 0 | 0% |
| | No, no fui contactado | 0 | 0% |
| EDU_8_1 | Coincide con la respuesta | 112 | 100% |
| | No coincide con la respuesta | 0 | 0% |
| ECR_1_1 | Coincide con la respuesta | 112 | 100% |
| | No coincide con la respuesta | 0 | 0% |
| EMP_1_1 | Coincide con la respuesta | 112 | 100% |
| | No coincide con la respuesta | 0 | 0% |
| MOB_1A_1 | Coincide con la respuesta | 112 | 100% |
| | No coincide con la respuesta | 0 | 0% |
| PER_6_1 | Coincide con la respuesta | 112 | 100% |
| | No coincide con la respuesta | 0 | 0% |
| Por último, me gustaría hacerle algunas preguntas acerca de la labor del encuestador. De 1 a 7 ¿Cómo evaluaría Ud.? | Información entregada: Claridad de objetivos del estudio | 6,98 | (promedio) |
| | Trato encuestador | 6,91 | (promedio) |

*Acompañamiento CAWI*

El acompañamiento de los participantes de la encuesta por el medio de recolección CAWI se hizo mediante correo electrónico dando respuesta a las inquietudes recibidas en la casilla de dedicada al estudio, desde donde se realiza además el envío de invitaciones: encuestas.trayectoria.phd@datavoz.cl. En total se recibieron a alrededor de 60 correos al día 4 de diciembre.

Los problemas e inquietudes más comunes se pueden agrupar en las siguientes categorías de saturación:

- **La persona presenta problemas con la auto aplicación de preguntas**: Las personas encuestadas presentan problemas para responder una pregunta en particular. Varios perciben que la encuesta se "quedó pegada" o que "no permite avanzar". En los casos que no dan mayor información al respecto, se les invita a compartir más detalles para una mejor atención. Una vez recopilados los antecedentes se les guía para poder continuar la encuesta. Los casos resueltos en su mayoría corresponden a situaciones en que hay un ítem de la pregunta que queda sin responder, sin ser notado por los participantes. Adicionalmente a la respuesta se les informa que pueden continuar respondiendo la encuesta desde la última pregunta respondida sin perder el avance siempre que se haga desde el mismo computador y navegador.

- **La persona presenta problemas con la comprensión de preguntas**: Las personas encuestadas presentan dudas sobre cómo entender ciertos conceptos, principalmente el de año de

referencia. Esto ocurre mayoritariamente en las preguntas de “Situación Laboral”, donde a pesar de haber leído las notas aún persisten dudas sobre si deben o no responder con los datos de su empleo actual en el 2020. En estos casos se responde dando mayores antecedentes sobre el año de referencia y porqué es necesario preguntar de esta manera. Además, se les informa que pueden continuar respondiendo la encuesta desde la última pregunta respondida sin perder el avance siempre que se haga desde el mismo computador y navegador.

• **La persona ya contestó la encuesta y le siguen llegando invitaciones**: Estos casos son originados en su totalidad por los desfases en la actualización de las bases de datos. La forma de actuar esta situación es tomar los datos de la persona, verificar la información y actualizar la base de datos correspondiente. Una vez realizada la actualización se le agradece al encuestado su participación y se le notifica las medidas tomadas en su situación particular.

• **La persona notifica no pertenecer a la población objetivo**: En estos casos se toma la información personal del encuestado y actualiza la base de datos para incorporar la información.

• **La persona no desea participar del estudio:** Al igual que en el caso anterior se toma la información personal del encuestado y actualiza la base de datos para incorporar la información. Se registra como “Rechazo” (R). Las causas de rechazo son diversas, algunas de ellas manifiestan descontento con la gestión del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación.

• **La persona no desea participar del estudio y manifiesta preocupación por la privacidad y el uso de los datos:** Estos son casos muy poco frecuentes, se abordan de la misma manera que en los casos de rechazo.

# 6. Aspectos metodológicos asociados a la elaboración de la base de datos

La encuesta CDH en Chile forma parte de un esfuerzo internacional promocionado por la OCDE para comprender y caracterizar a la población con grado de doctor, así como sus trayectorias laborales. Con esta información se espera mejorar los insumos utilizados para formular políticas públicas que fortalezcan la formación y desarrollo de personas con capital humano avanzado y consecuentemente las capacidades de realizar actividades I+D+i en el país. En las secciones precedentes se han explicado las distintas etapas del proceso de generación de la información que contiene la encuesta. En esta sección se revisan algunas de las estadísticas principales que resultaron del proceso anterior.

No es el objetivo de esta sección hacer un análisis exhaustivo de los datos ni testear hipótesis respecto de los patrones que se observen. Si no más bien, entregar al lector una guía base de la información contenida en la encuesta que sirva de insumos para investigaciones más focalizadas. Adicionalmente a lo reportado en esta sección, como parte del informe final se generaron tablas de resultados siguiendo los lineamientos OCDE y las bases de datos de la encuesta; ambos insumos permitirán a los usuarios el hacer comparaciones de la realidad chilena con la de los otros países donde se ha aplicado la encuesta.

A continuación, se presentan algunas estadísticas descriptivas

## 6.1 Procesamiento y validación de la Base de Datos

### 6.1.1 Homologación y deduplicación

El levantamiento se llevó a cabo combinando dos medios de recolección. El medio de recolección principal consistió en una encuesta auto-aplicada implementada en la plataforma CAWI SurveyMonkey. El medio de recolección secundario consistió en la aplicación de la encuesta por el equipo de encuestadores y encuestadoras del centro de llamado de Datavoz, con asistencia de la plataforma Gandia CatiNET. Las principales diferencias en la estructura de la base de datos de salida de ambos medios de recolección se pueden resumir en los siguientes aspectos:

- Registro de estampa de tiempo
- Preguntas de selección múltiple
- Preguntas de con opción de entrada de texto (“Otros, especifique”)
- Etiquetas de las respuestas

*Identificación de casos repetidos*

Se identificaron 15 casos en que la persona encuestada, habiendo respondido la encuesta mediante un medio de recolección, respondió también mediante el otro medio:

- 3 respondieron primero mediante CAWI y luego CATI

- 12 respondieron primero mediante CATI y luego CAWI

*Criterio de selección del caso*

Las personas que respondieron dos veces no lo hicieron exactamente de la misma manera en ambos medios. Se pudo observar leves diferencias en las respuestas. Se tomó la decisión de conservar las respuestas correspondientes al medio de recolección CAWI. Cabe hacer notar que las diferencias no se referían a variables sociodemográficas o a cambios significativos en las respuestas, sino más bien diferencias de grado en algunas escalas *lickert* contenidas en la encuesta (por ejemplo, entre estar "totalmente de acuerdo" y estar "moderadamente de acuerdo").

*Tamaño de muestra efectiva lograda*

A partir del proceso de homologación se obtuvo una base de datos con 2461 respuestas. Luego del proceso de deduplicación de las encuestas repetidas el tamaño final de la muestra lograda asciende a **2.445 respuestas**.

### 6.1.2 Validación de la Base de Datos

*Malla de validación*

Se realizaron alrededor de 100 evaluaciones de condiciones lógicas de saltos y consistencia, para detectar anomalías en las respuestas. Se descubrieron anomalías en 65 de ellas, la gran mayoría debido a errores de ingreso en los datos y se concentraron principalmente en casos correspondientes al levantamiento CAWI. Se estima que estos casos responden a una característica propia de las encuestas CAWI: Al poder navegar por la encuesta de forma libre (avanzar y retroceder) los doctores y doctoras podían contestar una pregunta de filtro avanzar en una dirección y retroceder.

*Acciones correctivas*

Se eliminaron las opciones inconsistentes con la última trayectoria registrada en la encuesta. Otros errores de ingreso de datos como rangos de fechas inconsistentes se subsanaron en su mayoría recurriendo a los datos disponibles en el directorio.

## 6.1 Ponderadores y factores de expansión

### 6.1.1 Cálculo de ponderadores

El cálculo de ponderadores tiene por objeto restituir a cada género su peso real en el universo nacional de la población de personas con grado de doctor residentes en Chile. Los ponderadores se calculan proyectando la muestra lograda a lo que debiera haber sido una muestra proporcional al peso poblacional de cada segmento que sirve de base para la ponderación.

Este cálculo debe hacerse para cada uno de los géneros: masculino y femenino. Para cada estrato *i*, se considera un ponderador de la siguiente manera:

$$Ponderador_i = \frac{\%Muestra_i}{\%Universo_i}$$

### 6.1.2 Cálculo de factores de expansión

Los factores de expansión se utilizan cuando se desea conocer el valor de la población de un resultado obtenido en la muestra efectiva. La fórmula para el cálculo de los coeficientes de expansión no difiere según si se trata de la muestra auto ponderada o desproporcionada.

Para obtener cada coeficiente se aplica la siguiente fórmula:

$$Expansor_i = \frac{Universo_i}{MuestraPonderada_i}$$

### 6.1.3 Valores calculados de ponderadores y factores de expansión

Los valores de los ponderadores se resumen en la Tabla 30:

*Tabla 310: Valores de ponderadores y factores de expansión calculados*

| Estratos | Muestra | Universo | % Muestra | % Universo | Ponderadores* | Expansores* |
|---|---|---|---|---|---|---|
| Femenino | 907 | 6.731 | 37,1% | 36,7% | 0,9887 | 7,5059 |
| Masculino | 1.538 | 11.621 | 62,9% | 63,3% | 1,0067 | 7,5059 |
| Total general | 2.445 | 18.352 | 100,0% | 100,0% | | |

*/ Los valores son redondeados aquí a cuatro decimales para evitar números excesivamente largos en la tabla.

## 6.2 Error muestral final

El diseño muestral implementado corresponde a un muestreo aleatorio estratificado el cual consistió en llevar a cabo un muestreo aleatorio simple para cada estrato que compone la muestra. De acuerdo al diseño muestral definido el cálculo del error muestral efectivo se obtiene a partir de la aplicación de la siguiente formula:[16]

$$error.muestral.efectivo = Z_{1-\alpha/2} * \sqrt{\frac{p \cdot (1-p)}{n} \cdot \left(1 - \frac{n}{N}\right) \cdot deff}$$

Donde se considera:

- Un nivel de confianza $(1-\alpha)$ de 95%
- Corrección por universo finito $\left(1 - \frac{n}{N}\right)$, en que "N" es el tamaño de universo y "n" el tamaño de la muestra
- Varianza máxima considerando la estimación de una proporción $(p = 0{,}5)$

Con respecto a la estimación del efecto del diseño muestral *deff,* luego de una evaluación de los resultados obtenidos en la muestra efectiva se concluyó que es posible aproximar su valor a 1,0

16 En que $Z_{1-\alpha/2}$ =1,96 (percentil 97,5 de la distribución normal estándar al considerar un nivel de confianza del 95%)

sin generar sesgo estadístico. Esta decisión se fundamenta en los antecedes del diseño muestral donde cada estrato fue calculado directamente en función de la proporción observada de cada estrato en la población, descartando el uso de cuotas o sobre muestra.

Los resultados expuestos en la Tabla 31 muestra una diferencia de apenas un 0,4% en la proporción de la muestra de ambos estratos en comparación con los valores correspondiente al universo. En consecuencia, el valor calculado para los ponderadores es en ambos casos prácticamente 1, lo que sustenta la idea de que por las consideraciones del diseño y la condiciones en las que se llevó a cabo el levantamiento el proceso se puede aproximar a un muestreo aleatorio simple.

Aplicando las consideraciones descritas en la sección anterior se observa que los valores obtenidos del cálculo de los errores muestrales efectivos se mantuvieron por debajo del 3,1% lo que responde al diseño muestral descrito en la sección 0

*Tabla 31: Distribución de la muestra, universo y error muestral por género*

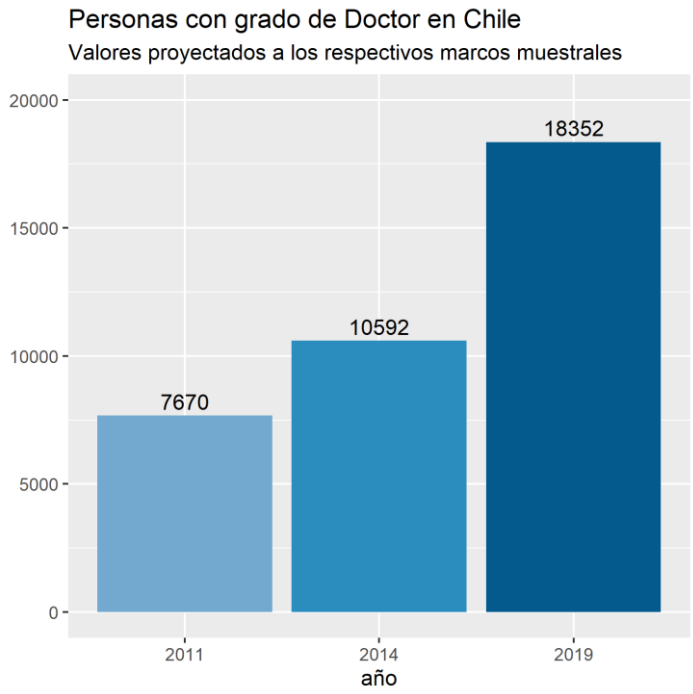
| Estratos | Muestra | Universo | % Muestra | % Universo | Error muestral estimado |
|---|---|---|---|---|---|
| Femenino | 907 | 6.731 | 37,1% | 36,7% | 3,03% |
| Masculino | 1.538 | 11.621 | 62,9% | 63,3% | 2,33% |
| Total general | 2.445 | 18.352 | 100,0% | 100,0% | 1,85% |

# 7. Análisis de Resultados

### 7.1.1 Crecimiento de la población de personas con grado de doctor

En total se lograron 2.445 encuestas completas las que se proyectan sobre una población de 18.352 personas con grado de doctor. Respecto a la versión anterior de la encuesta CDH, se acentúa la presencia de una población adulta joven, especialmente en el rango etario de 35 a 44 años que alcanza un 41,9% aumentando respecto a la encuesta CDH anterior (38,8%). Lo mismo ocurre en el rango etario de 45 a 54 años, con un aumento de 5,4%. Esto podría verse explicado por los esfuerzos públicos para financiar estudios doctorales y, por ende, generar nuevas cohortes de personas con grado de doctor más amplios que los anteriores.

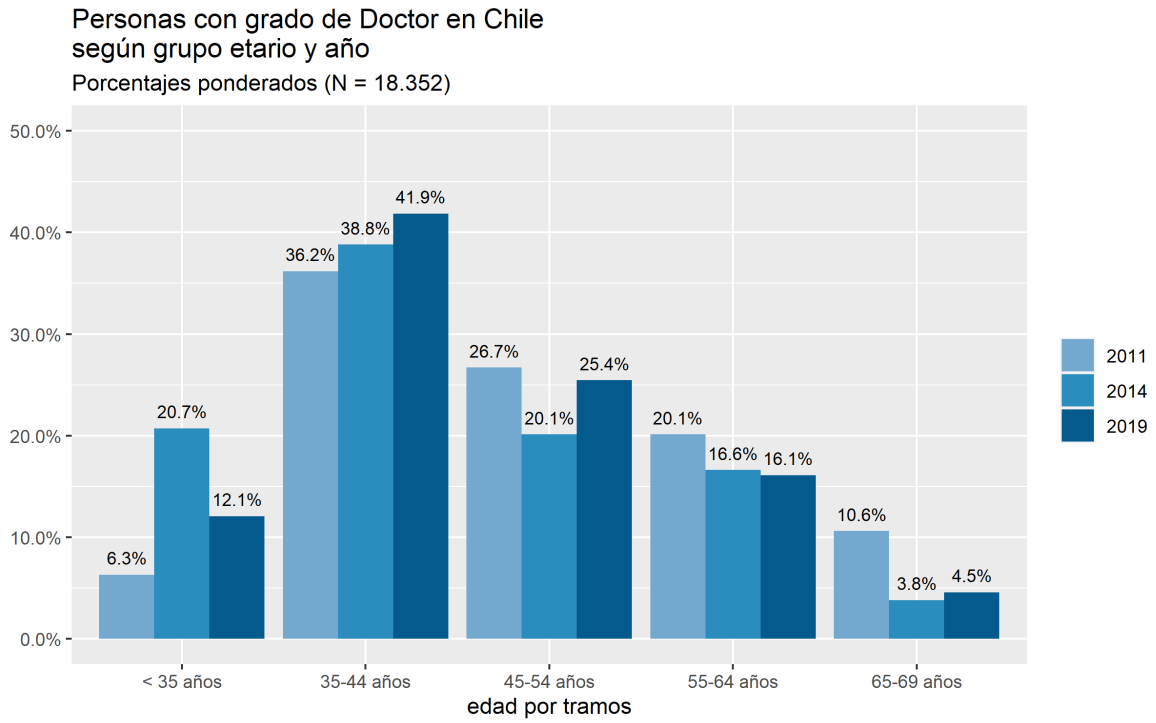
*Figura 8: Marco muestral encuestas CDH en Chile*




Sin embargo, la situación es distinta en el rango etario de menores de 35 años, que representa un 12,1% de la población, en contraste con el 20,7% correspondiente a la medición del 2014, notándose una desaceleración. Todas las diferencias entre rangos etarios son estadísticamente significativas[17], salvo por el rango etario entre 55 y 64 años.

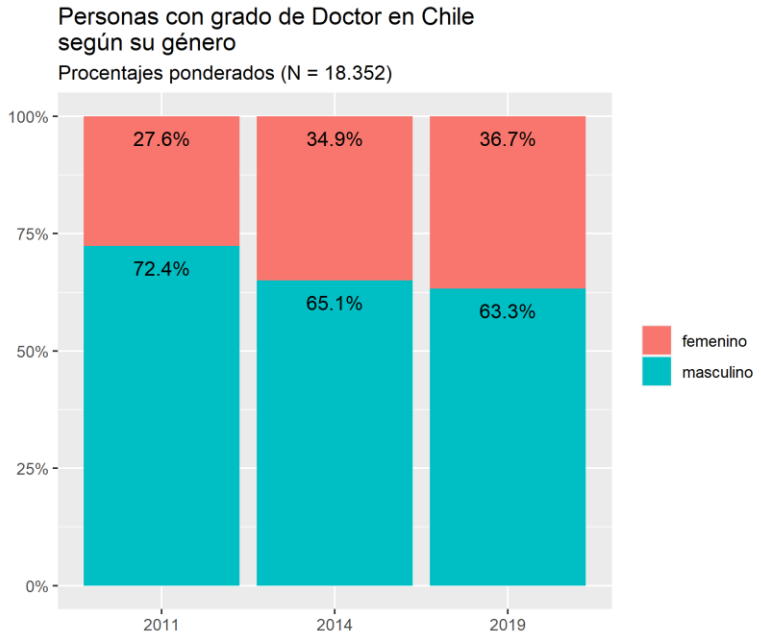
[17] Se obtuvo un $p < 0.005$ para cada una de ellas de acuerdo al test de diferencia de proporciones entre dos muestras, lo que permite rechazar la hipótesis nula de que las proporciones poblacionales sean iguales.

*Figura 9: Distribución personas con doctorado por tramo de edad últimas 3 encuestas CDH en Chile*




En cuanto al género, del total de personas encuestadas, un 36,7% fueron del género femenino y un 63,3% del género masculino. Si bien la proporción de mujeres es menor, su proporción en el total sigue creciendo en relación con lo informado en las dos mediciones previas de la encuesta CDH.

*Figura 10: Distribución por género marco muestral doctorados últimas 3 encuestas CDH en Chile*



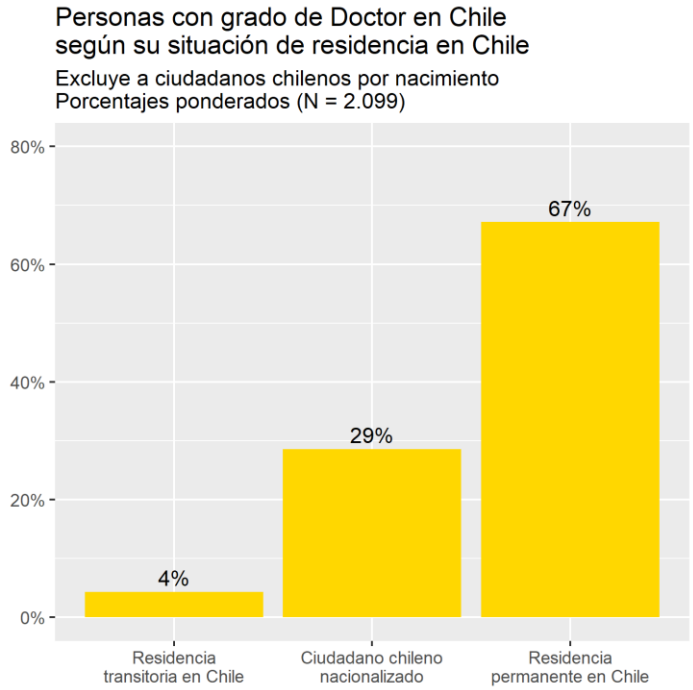
En relación con el tiempo transcurrido desde el inicio del programa y la obtención del grado de doctor, en torno al 80% de los encuestados demoró seis o menos años en completar sus estudios doctorales.

Por otro lado, en esta medición el porcentaje con nacionalidad chilena por nacimiento (89%) es ligeramente superior con el registrado en la encuesta CDH anterior (87%). A los que debe agregarse un 2% de personas nacionalizadas.

*Tabla 32: PER_4_1 ¿Cuál es su situación de residencia en Chile?*

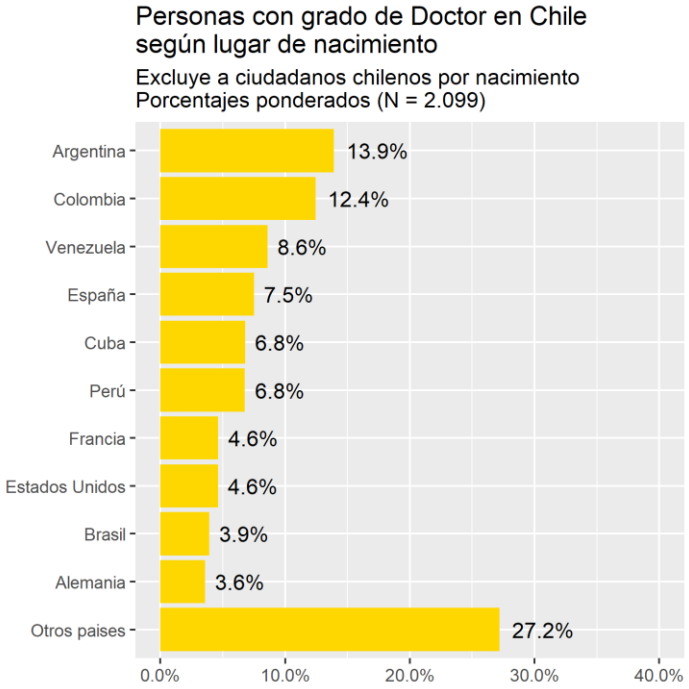
| Situación de residencia | Muestra | Proyección | Porcentaje |
|---|---|---|---|
| Ciudadano chileno por nacimiento | 2.170 | 16.290 | 88,8% |
| Ciudadano chileno nacionalizado | 60 | 449 | 2,4% |
| Residencia permanente en Chile | 203 | 1.522 | 8,3% |
| Residencia transitoria en Chile | 12 | 90 | 0,5% |
| Total | 2.445 | 18.352 | 100,0% |

*Figura 11: Distribución por situación de residencia personas con grado de doctor en Chile 2019*




Además, se puede observar en los gráficos que las personas con grado de Doctor residentes en Chile nacidas en el extranjero provienen mayoritariamente de países hispanohablantes, seguidos en menor proporción por Francia, Estados Unidos, Brasil y Alemania.

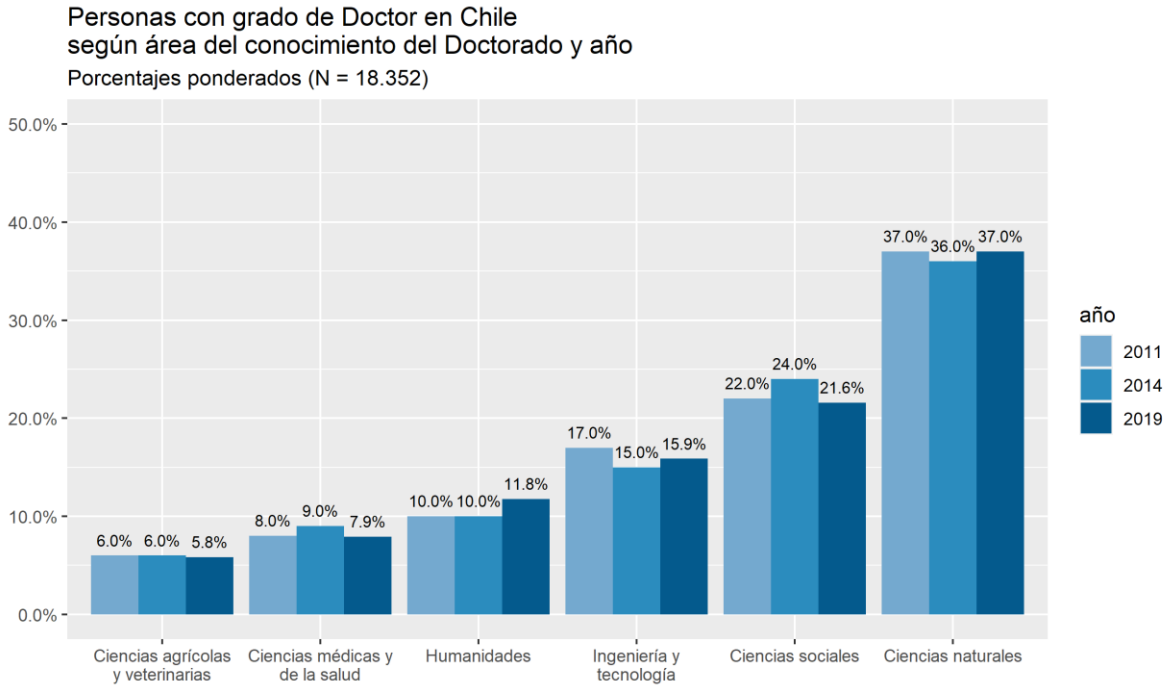
*Figura 12: Distribución porcentual por país de nacimiento en personas con grado de doctor en Chile 2019*



### 7.1.2 Estudios cursados

La distribución de la población de doctores en las distintas áreas del conocimiento se concentra mayoritariamente entre las Ciencias Naturales y las Ciencias Sociales, representando casi el 60% de las personas con grado de Doctor. Como se observar en el gráfico siguiente, las proporciones de personas con grado de doctor entre áreas de conocimiento el 2019 es similar a las observadas en mediciones anteriores.

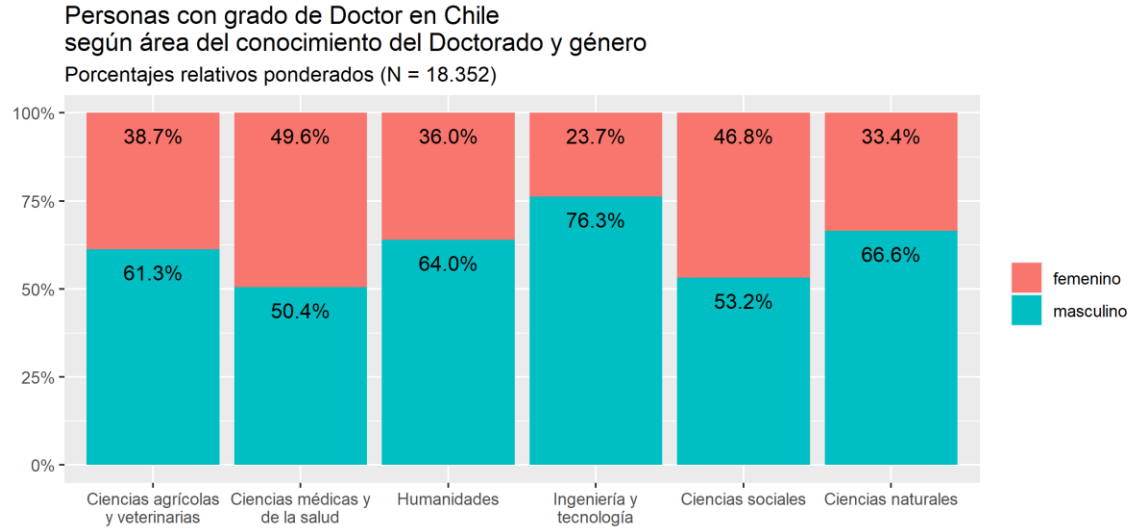
*Figura 13: Distribución por área de conocimiento del doctorado personas con grado de doctor en Chile 2019*




Al observar por cada área del conocimiento qué proporción hay por cada género, se constata un incremento de la proporción de mujeres en el área de Ciencias Médicas y de la Salud respecto a la medición anterior, avanzando de un 46% a casi un 50%. También se observan aumentos significativos en Ciencias Sociales de un 39% a un 46,8% y Ciencias Naturales de un 30% a un 33,4%. Por otro lado, en las áreas de Ingeniería y Tecnología, Humanidades la tendencia va en el sentido contrario bajando significativamente de un 26% en 2015 a un 23,7% en 2019. Todas las diferencias de distribución de la participación de género femenino según área de conocimiento entre ambas ediciones de la encuesta son estadísticamente significativas[18].

[18] Se obtuvo un $p < 0.001$ para cada una de ellas de acuerdo al test de diferencia de proporciones entre dos muestras, lo que permite rechazar la hipótesis nula de que las proporciones poblacionales sean iguales.

*Figura 14: Distribución por género según área del conocimiento doctorado personas con grado de doctor en Chile 2019*



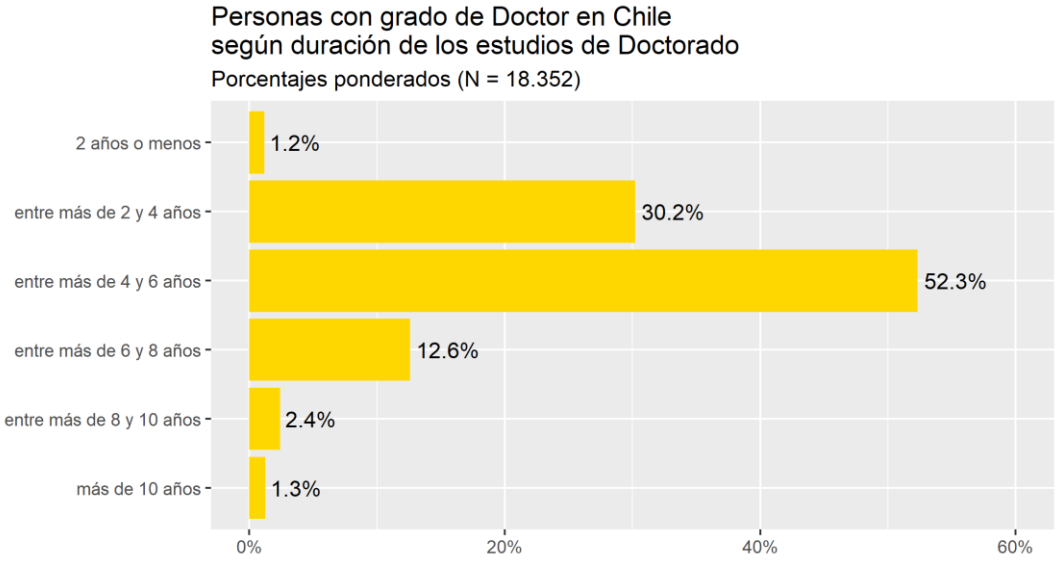

Complementariamente, si vemos cómo se distribuyen los hombres y las mujeres entre las distintas áreas de conocimiento, se puede constatar (ver Tabla 33) que las mujeres aparecen más vinculadas a Ciencias Sociales que los hombres (27,6% versus 18%).

*Tabla 34: Distribución del área del conocimiento por género*

| EDU_2_1 | Femenino | | Masculino | | Otro | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Ciencias naturales | 2.279 | 33,8% | 4.503 | 38,8% | 8 | 50,0% |
| Ingeniería y tecnología | 683 | 10,1% | 2.229 | 19,2% | 0 | 0,0% |
| Ciencias médicas y de la salud | 720 | 10,7% | 733 | 6,3% | 0 | 0,0% |
| Ciencias agrícolas y veterinarias | 416 | 6,2% | 657 | 5,7% | 0 | 0,0% |
| Ciencias sociales | 1.863 | 27,6% | 2.093 | 18,0% | 8 | 50,0% |
| Humanidades | 779 | 11,6% | 1.383 | 11,9% | 0 | 0,0% |
| Total | 6.739 | 100,0% | 11.598 | 100,0% | 15 | 100,0% |

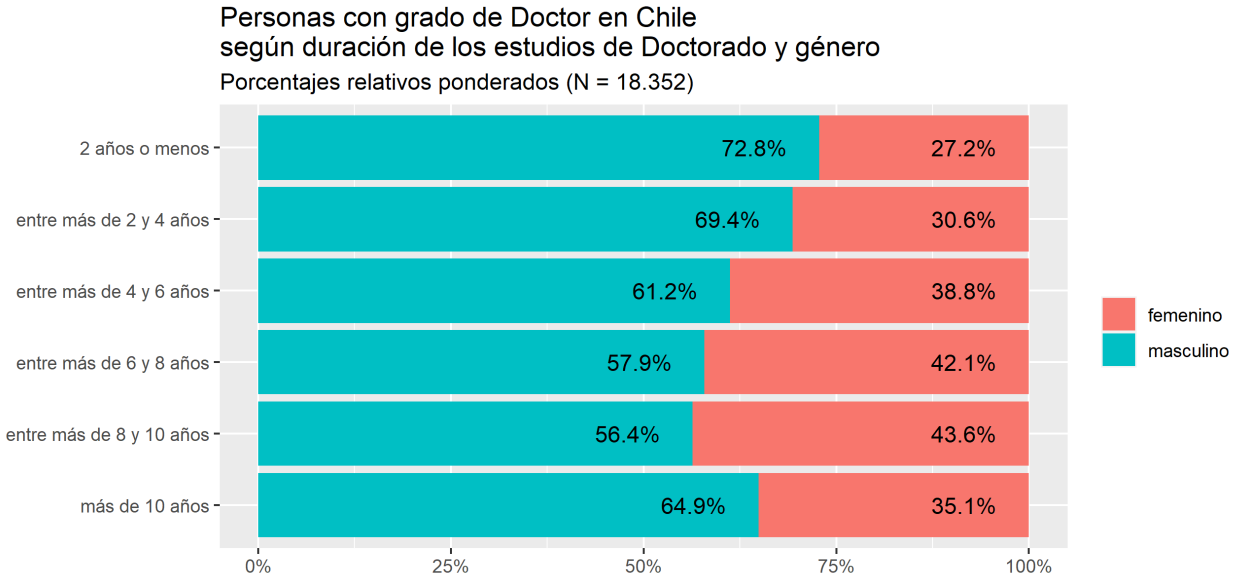
Respecto a la duración de los estudios de doctorado se observa que un 82,5% de los doctores y doctoras tardan entre 2 y 6 años en obtener el grado.

*Figura 15: Distribución porcentual por duración estudios personas con grado de doctor en Chile 2019*




Al desagregar los datos anteriores por género, se observa que las mujeres están participando en mayor proporción en los rangos de mayor duración.[19] Como se muestra en el gráfico y tabla siguientes, en ambos géneros, la mayoría se demora entre 4 a 6 años (55,3% las mujeres y 50,6% los hombres); no obstante, la distribución se inclina más hacia menores tiempos de estudio en los hombres que en las mujeres. Ello no sólo significa que las mujeres demoren en promedio más tiempo, sino también que – todo lo demás constante – los procesos de inserción laboral en contextos de alta competencia podrían ser también más difíciles para mujeres que para hombres (véase sección 6.3.3). Al respecto, cabe señalar que sería pertinente indagar en los factores sociales y/o económicos que pueden estar afectando a la población femenina en la duración de sus estudios.

*Figura 16: Distribución por género según duración estudios doctorado en personas con grado de doctor en Chile 2019*




[19] Diferencias estadísticamente significativas entre hombres y mujeres a un valor p < 0.001.

*Tabla 35: Duración del doctorado por género*

| | Femenino | | Masculino | | Otro | |
|---|---|---|---|---|---|---|
| Duración del doctorado | N | % | N | % | N | % |
| 2 años o menos | 59 | 0,9% | 159 | 1,4% | 0 | 0,0% |
| entre más de 2 y 4 años | 1.700 | 25,2% | 3.838 | 33,1% | 8 | 50,0% |
| entre más de 4 y 6 años | 3.726 | 55,3% | 5.871 | 50,6% | 8 | 50,0% |
| entre más de 6 y 8 años | 980 | 14,5% | 1.330 | 11,5% | 0 | 0,0% |
| entre más de 8 y 10 años | 193 | 2,9% | 249 | 2,1% | 0 | 0,0% |
| más de 10 años | 82 | 1,2% | 151 | 1,3% | 0 | 0,0% |
| Total | 6.739 | 100,0% | 11.598 | 100,0% | 15 | 100,0% |

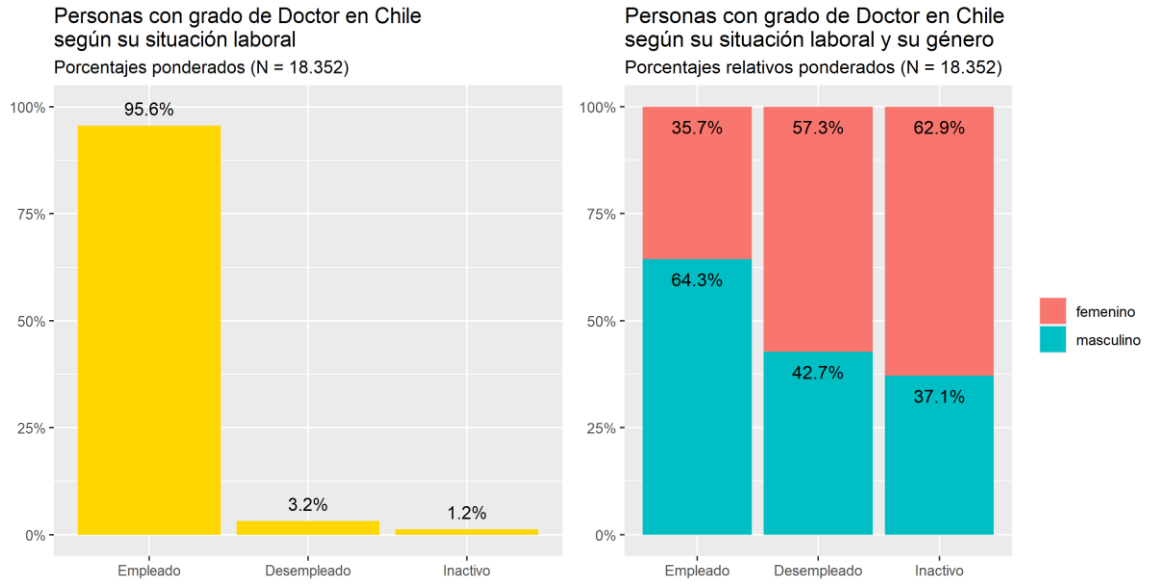
### 7.1.3 Inserción laboral

Al ser consultados sobre su situación laboral, el 96% declara estar empleado, un 3% desempleado y un 1% estar inactivo. Estas proporciones dan cuenta de una inserción laboral significativamente superior al promedio de la población nacional (un patrón similar fue observado en la encuesta CDH anterior, 94%).

*Tabla 36: EMP_1_1 Al 1 de diciembre de 2019 usted se encontraba…*

| Situación laboral | Muestra | Proyección | Porcentaje |
|---|---|---|---|
| Empleado | 2.337 | 17.545 | 95,6% |
| Desempleado | 78 | 583 | 3,2% |
| Inactivo | 30 | 224 | 1,2% |
| Total | 2.445 | 2.445 | 18.352 |

*Figura 17: Situación laboral personas con grado de doctor en Chile y desagregación por género*



Al consultarse sobre el tipo de empleo principal, en su mayoría los encuestados declaran participar en el sistema de educación superior. Al respecto, es importante recordar (ver sección 3) que uno de los desafíos más importantes de la caracterización de personas con grado de doctor en Chile es que las principales fuentes de información para contactarlos están vinculadas al proceso educativo (universidades y fuentes de becas). Es con ese antecedente que debe considerarse el análisis de estos datos en lógica comparativa. En esta medición, un 7,1% de los encuestados manifiestan estar vinculados al mundo empresarial. En la versión 2014 ese porcentaje fue de un 3,8%. Es decir, prácticamente se duplicó la proporción de personas vinculadas a la industria. Todo lo cual sugiere que, probablemente, la presente muestra es más representativa de la composición de la inserción laboral de los y las doctores en el mundo de la empresa.

*Tabla 37: EMP_3A_8 Sector al que corresponde su empleo PRINCIPAL*

| Empleo principal | Muestra | Proyección | Porcentaje |
|---|---|---|---|
| Empresarial | 174 | 1.308 | 7,1% |
| Administración pública | 90 | 676 | 3,7% |
| Educación Superior | 1.963 | 14.736 | 80,3% |
| Otro sector de educación | 25 | 187 | 1,0% |
| Institución privada sin fines de lucro | 85 | 637 | 3,5% |
| Desempleado o Inactivo | 108 | 807 | 4,4% |
| Total | 2.445 | 18.352 | 100,0% |

Por otro lado, 2.337 de los 2.445 encuestados indican cuál es la ocupación principal en la que se desempeñan. De ellos, 2.095 (89,6%) señala ocupar cargos profesionales científicos e intelectuales lo que da cuenta de una inserción laboral consistente con la formación doctoral. En cuanto a los niveles gerenciales, hay 197 personas (8,4%) que declaran estar en ocupaciones de dirección o gerencia.

Todo lo anterior permite proyectar a nivel poblacional que en Chile hay 15.728 doctores y doctoras desempeñándose como científicos e intelectuales y otras 1.480 en cargos de dirección o gerenciales.

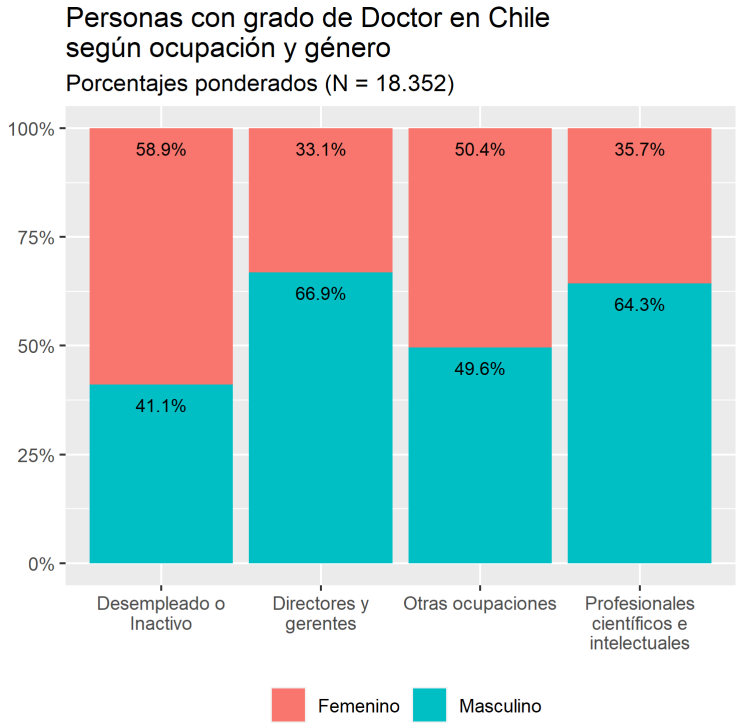
*Tabla 38: EMP_3A_32 Selecciones la categoría que mejor describa su Ocupación del empleo PRINCIPAL*

| Categoría de empleo | Muestra | Proyección | Porcentaje |
|---|---|---|---|
| Directores y gerentes | 197 | 1.480 | 8,1% |
| Profesionales científicos e intelectuales | 2.095 | 15.728 | 85,7% |
| Técnicos y profesionales de nivel medio | 26 | 195 | 1,1% |
| Personal de apoyo administrativo | 10 | 75 | 0,4% |
| Trabajadores de los servicios y vendedores de comercios y mercados | 1 | 7 | 0,0% |
| Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros | 1 | 8 | 0,0% |
| Operadores de instalaciones y máquinas y ensambladores | 1 | 7 | 0,0% |
| Ocupaciones elementales | 4 | 30 | 0,2% |
| Ocupaciones militares | 2 | 15 | 0,1% |

| | | | |
|---|---|---|---|
| Desempleado o Inactivo | 108 | 807 | 4,4% |
| Total | 2.445 | 18.352 | 100,0% |

Ahora bien, los mismos datos anteriores mirados desde una perspectiva de género muestran patrones diferentes entre hombres y mujeres. En la población inactiva es mayor la proporción de mujeres (58,9%) y lo contrario sucede en las ocupaciones de dirección y gerencia (33,1%) y profesionales científicos e intelectuales (35,7%).

*Figura 18: Distribución por género según ocupación principal personas con grado de doctor en Chile 2019*




Entre quienes indicaron estar vinculados a una institución de educación superior como empleo principal, 1.966 identificaron la categoría de universidad según la clasificación propuesta (ver tabla). De este modo, el 65,6% indica estar vinculado a una universidad pertenecientes al CRUCH y un 13,9% a otras universidades fuera del CRUCH.

*Tabla 39: EMP_3A_9 Tipo de institución de educación superior*

| EMP_3A_9 | Muestra | Proyección | Porcentaje |
|---|---|---|---|
| Universidad pública o tradicional (CRUCH) | 1.604 | 12.041 | 65,6% |
| Universidad no perteneciente al CRUCH | 339 | 2.546 | 13,9% |
| Instituto Profesional | 5 | 38 | 0,2% |
| Centro de formación técnica | 1 | 7 | 0,0% |
| Otro | 14 | 105 | 0,6% |
| Desempleado o Inactivo | 482 | 3.616 | 19,7% |
| Total | 2.445 | 18.352 | 100,0% |

*6.3.3.1 Inserción laboral en la investigación*

Como uno de los objetivos principales que el Estado persigue al invertir en capital humano avanzado es potenciar la investigación en el país, es importante indagar en las variables que podrían estar contribuyendo o limitando a ese propósito. Para ello es importante mirar la inserción laboral de las personas con grado de doctor en búsqueda de potenciales patrones e hipótesis. En esta subsección, consideramos diferencias relacionadas con la inserción laboral de personas que obtuvieron su grado de doctor a distintos rangos de edad y entre personas que difirieron en el tiempo que demoraron en obtener dicho grado.

En términos generales, el desarrollo de una carrera de investigación podría verse potenciada mientras antes las personas con grado de doctor empiezan a involucrase en las actividades afines. Ello significa que, mientras menor sea la edad de egreso, menor el tiempo de formación doctoral y, por otro lado, mientras antes sea la inserción en procesos de investigación más experiencia y capital humano se acumulará y, por ende, ello debería reflejarse de algún modo en la inserción laboral de las personas con grado de doctor.

Los datos de esta encuesta parecen estar en sintonía con esas premisas. Primero, en términos generales, la población encuestada presenta altos niveles de empleo en comparación con el promedio nacional (17.545 de los 18.352, es decir 95,6%). Estos porcentajes son elevados para todos los rangos de edad al egreso. Aunque disminuyen ligeramente mientras mayor es la edad a la que se obtuvo el grado.

*Tabla 40: Rango de edad a la que se obtuvo el grado y situación laboral*

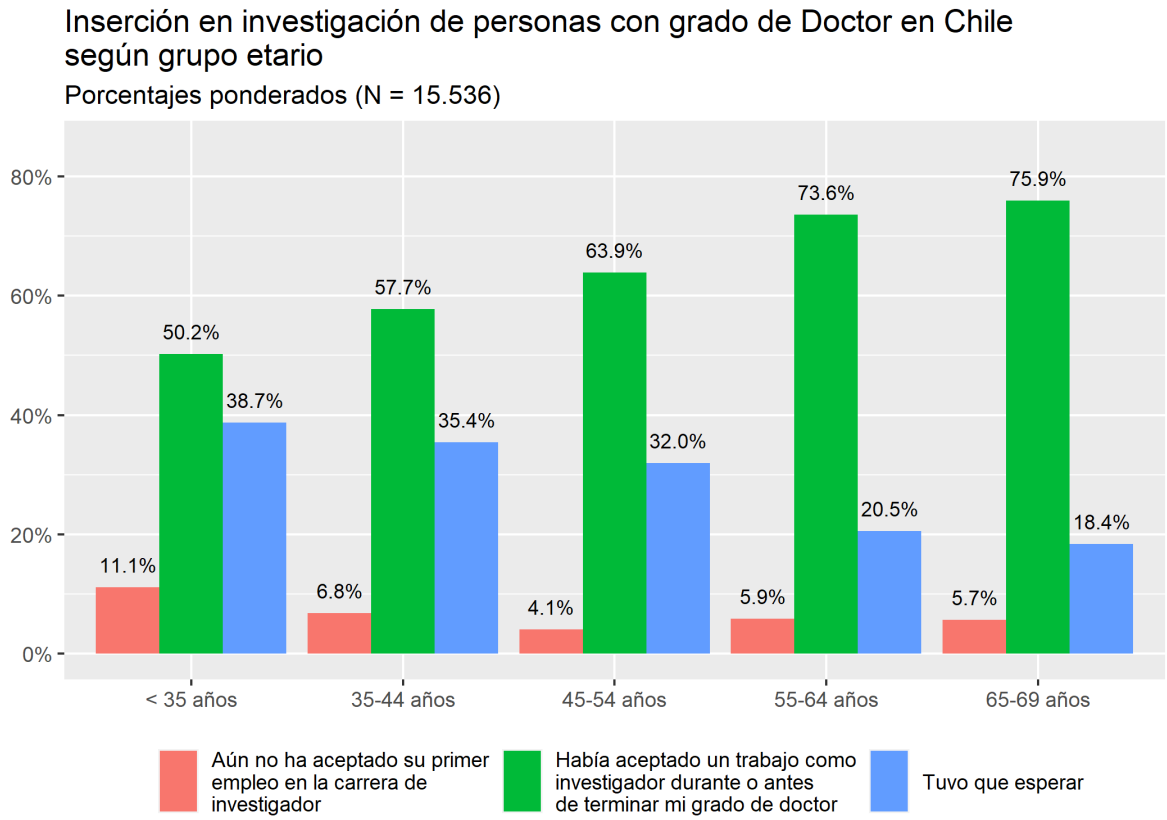
| Rango de edad a la que se obtuvo el grado | Empleado | | Desempleado | | Inactivo | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| < 35 años | 2.033 | 11,6% | 150 | 25,6% | 30 | 13,2% |
| 35-44 años | 7.346 | 41,9% | 262 | 44,9% | 75 | 33,4% |
| 45-54 años | 4.542 | 25,9% | 105 | 18,0% | 22 | 10,0% |
| 55-64 años | 2.879 | 16,4% | 52 | 9,0% | 22 | 10,0% |
| 65-69 años | 745 | 4,2% | 15 | 2,6% | 75 | 33,4% |
| Total | 17.545 | 100,0% | 583 | 100,0% | 224 | 100,0% |

Por otro lado, al proyectar los datos a nivel del universo de personas con grado de doctor, el 81,7% del total (aproximadamente 15 mil doctores) han aceptado alguna vez un trabajo como investigador o con alguna carga de investigación. Y lo relevante en materia de fomento de la investigación es que 12 mil de ellos aceptaron y se integraron a trabajos con un componente de investigación antes de terminar sus estudios de doctorado. Sólo 6.300 (un 34,4%) ha tenido que esperar hasta después del egreso para conseguir un trabajo vinculado a la investigación.

Por último, al cruzar estos datos por el rango de edad al egreso se constatan algunos patrones relevantes. Como se muestra en el siguiente gráfico, si segmentamos a los encuestados en cinco rangos de edad al momento de la obtención del título de doctor (<35, 35-44, 45-54, 55-64 y 65-69), se constata que, en todos los grupos mayores de 30 años, más de la mitad se insertó en trabajos de

investigación desde antes del egreso. No obstante, para las cohortes más jóvenes al momento del egreso esos procesos han sido significativamente[20] más lentos y la espera fue mayor.

*Figura 19: Inserción laboral en la investigación personas con grado de doctor en Chile 2019*




Entonces, como se indicó anteriormente, dado que el perfil etario de los doctores se ha movido hacia un aumento en los rangos de menor edad, estos datos podrían estar reflejando las dificultades de inserción laboral que enfrentan las nuevas cohortes de personas con grado de doctor. Por otro lado, dicha cohorte es la que en mayor medida se ha involucrado en trabajos de investigación puesto que sólo un 3,8% declara no haber aceptado aún un trabajo afín. En cambio, la proporción que indican no haber participado de trabajos de investigación en los grupos etarios mayores crece, siendo el máximo el grupo de mayores de 60 años con un 27,3%.

Ahora bien, no todas las personas que indicaron haber aceptado un trabajo de investigación alguna vez y que están trabajando en la actualidad, se mantienen en trabajos de investigación. Como se muestra en la tabla siguiente, de las 15 mil personas que han participado en investigación (81,7%), sólo 13.767 estaría realizando este tipo de labores al momento de la encuesta. Siendo mayor la proporción en cohortes más jóvenes que trabajan en investigación vis à vis las que no lo hacen.

[20] Todas las diferencias por rango de edad son significativas con $p < 0.001$.

*Tabla 41: Inserción laboral actual de los doctores según edad de egreso y ocupación vinculada a investigación.*

| Rango de edad a la que se obtuvo el grado | Trabaja en Investigación | | No trabaja en investigación | | Desempleado o Inactivo | | Total |
|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | |
| < 35 años | 1.688 | 12,3% | 345 | 9,1% | 179 | 22,2% | 2.212 |
| 35-44 años | 5.729 | 41,6% | 1.616 | 42,8% | 337 | 41,7% | 7.682 |
| 45-54 años | 3.575 | 26,0% | 967 | 25,6% | 127 | 15,7% | 4.669 |
| 55-64 años | 2.255 | 16,4% | 624 | 16,5% | 75 | 9,3% | 2.954 |
| 65-69 años | 520 | 3,8% | 225 | 6,0% | 90 | 11,1% | 835 |
| Total | 13.767 | 100,0% | 3.778 | 100,0% | 807 | 100,0% | 18.352 |

Al segmentar a los encuestados por el tiempo que demoraron en obtener su grado de doctor y cruzar la información con la declaración que hicieron sobre si estaban o no haciendo trabajo de investigación, lo que debería suceder es que hay mayor inserción en investigación mientras menor sea el tiempo de duración de los estudios. La tabla siguiente muestra evidencia en esa dirección. Mientras los doctores que demoraron 4 años o menos en terminar sus estudios están sobre un 70% trabajando en investigación, aquellos que demoraron 8 años o más tienen proporciones de trabajo en investigación inferiores al 63%.

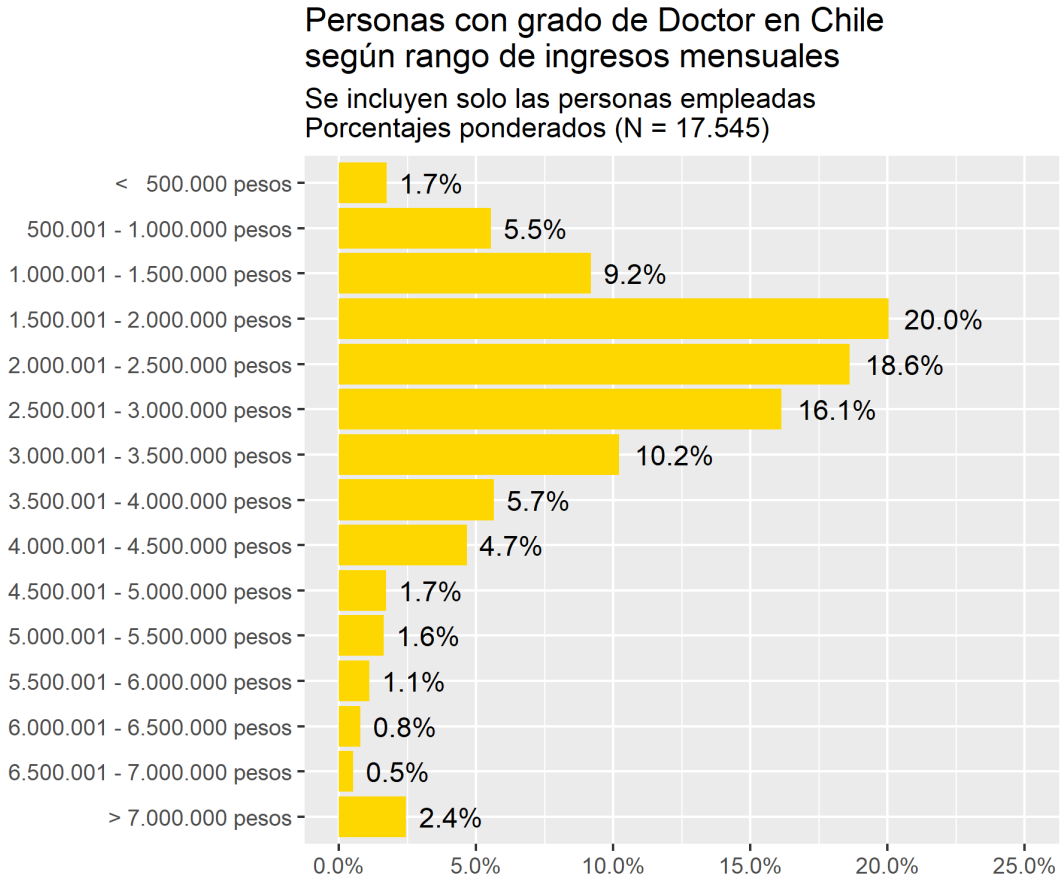
*Tabla 42: Duración de estudios doctorales hasta obtener el grado e inserción laboral en investigación*

| Duración de estudios doctorales hasta obtener el grado | Trabaja en Investigación | | No trabaja en investigación | | Desempleado o Inactivo | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| 2 años o menos | 173 | 1,3% | 30 | 0,8% | 15 | 1,9% |
| entre más de 2 y 4 años | 4.270 | 31,0% | 1.088 | 28,8% | 187 | 23,2% |
| entre más de 4 y 6 años | 7.268 | 52,8% | 1.925 | 51,0% | 411 | 50,9% |
| entre más de 6 y 8 años | 1.636 | 11,9% | 517 | 13,7% | 157 | 19,4% |
| entre más de 8 y 10 años | 278 | 2,0% | 142 | 3,8% | 22 | 2,8% |
| más de 10 años | 143 | 1,0% | 75 | 2,0% | 15 | 1,9% |
| Total | 13.767 | 100,0% | 3.778 | 100,0% | 807 | 100,0% |

### *6.3.3.2 Remuneraciones declaradas*

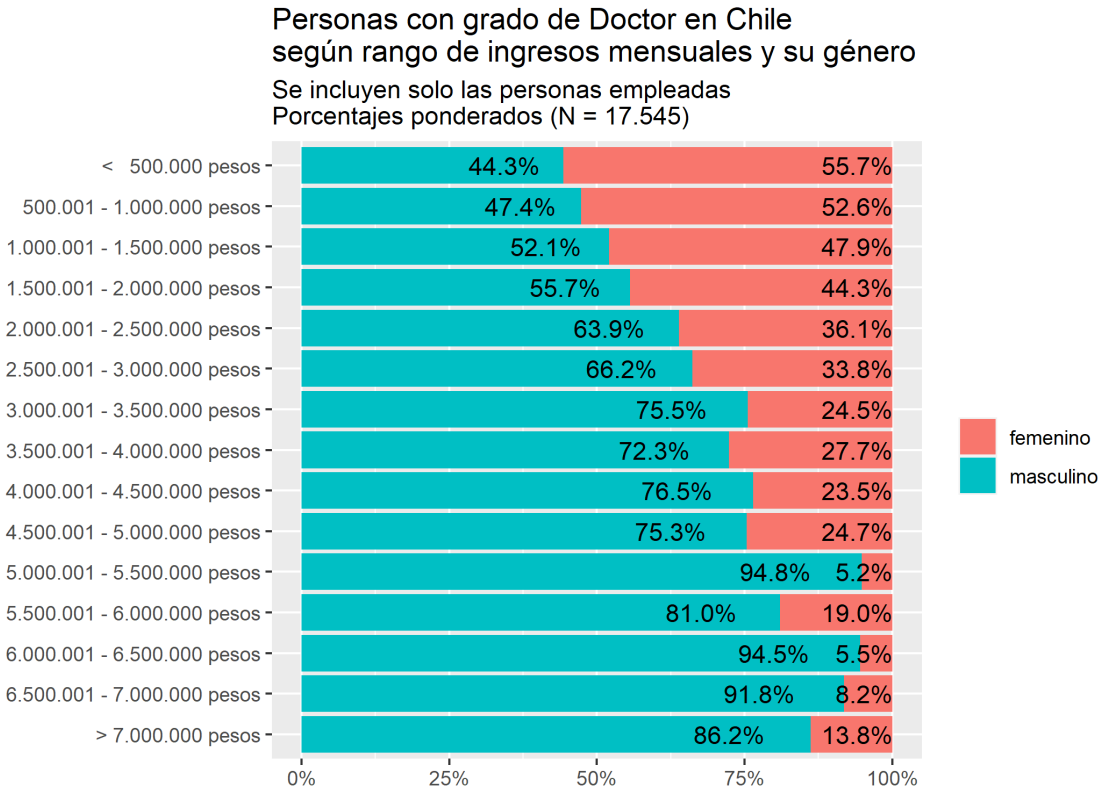
Como se consignó anteriormente, en las etapas cualitativas de preparación de la encuesta algunos de los entrevistados mostraron aprehensiones respecto a dar a conocer sus ingresos. Por este motivo, se procedió a establecer rangos de ingresos en vez de montos específicos (como es en la versión original del cuestionario OCDE). Los resultados muestran que existe alta dispersión en las remuneraciones declaradas con una larga cola hacia rangos de ingresos superiores; no obstante, dos tercios del total de los ingresos mensuales está concentrado en el rango entre 1,5 y 3,5 millones de pesos (65%).

*Figura 20: Distribución por ingreso bruto mensual de las personas con grado de doctor en Chile 2019*




Al comparar la distribución de las remuneraciones de hombres y mujeres, se constata inmediatamente que los ingresos de las mujeres están relativamente más inclinados hacia remuneraciones menores que los hombres.

*Figura 21: Distribución por género según tramo de ingresos brutos mensuales personas con grado de doctor en Chile 2019*




Imputando el salario promedio de cada rango, es posible estimar el promedio ponderado del ingreso anual y mensual por género. En las categorías límites (menor a 500.000 pesos y más de 7.000.000) se imputó el límite mismo.[21] En línea con la evidencia del gráfico anterior, se observa una diferencia superior a los 600 mil pesos entre el ingreso mensual percibido por mujeres y hombres.

*Tabla 43: EMP_2_1 y PER_1_1 Ingresos Promedios Brutos, anuales y mensuales*

| Género | Ingreso anual bruto (en miles de pesos) | Ingreso mensual bruto (en pesos) |
|---|---|---|
| Femenino | 26.619 | 2.218.268 |
| Masculino | 33.865 | 2.822.122 |
| Total ponderado | 31.252 | 2.604.300 |

No obstante, para poder diseñar políticas públicas eficaces que permitan acortar esas brechas es necesario que futuros estudios indaguen en los detalles que puedan explicarlas.

Al comparar las distribuciones de la inserción laboral por tipo de ocupaciones, nuevamente, los hombres aparecen relativamente más representados en cargos de gerencia. Pero también en cargos profesionales y científicos.

[21] No se corrige en términos de horas trabajadas, lo que podría ser una limitante del análisis.

*Tabla 44: Distribución de las ocupaciones por género*

| EMP_3A_32 | Femenino | | Masculino | | Otro | | Total |
|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | |
| Directores y gerentes | 490 | 7,3% | 990 | 8,5% | 0 | 0,0% | 1.480 |
| Profesionales científicos e intelectuales | 5.603 | 83,1% | 10.109 | 87,2% | 15 | 100,0% | 15.727 |
| Técnicos y profesionales de nivel medio | 89 | 1,3% | 106 | 0,9% | 0 | 0,0% | 195 |
| Personal de apoyo administrativo | 45 | 0,7% | 30 | 0,3% | 0 | 0,0% | 75 |
| Trabajadores de los servicios y vendedores de comercios y mercados | 7 | 0,1% | 0 | 0,0% | 0 | 0,0% | 7 |
| Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros | 0 | 0,0% | 8 | 0,1% | 0 | 0,0% | 8 |
| Oficiales, operarios y artesanos de artes mecánicas y de otros oficios | 0 | 0,0% | 0 | 0,0% | 0 | 0,0% | 0 |
| Operadores de instalaciones y máquinas y ensambladores | 7 | 0,1% | 0 | 0,0% | 0 | 0,0% | 7 |
| Ocupaciones elementales | 22 | 0,3% | 8 | 0,1% | 0 | 0,0% | 30 |
| Ocupaciones militares | 0 | 0,0% | 15 | 0,1% | 0 | 0,0% | 15 |
| Desempleado o Inactivo | 475 | 7,0% | 332 | 2,9% | 0 | 0,0% | 807 |
| Total | 6.739 | 100,0% | 11.598 | 100,0% | 15 | 100,0% | 18.352 |

Los resultados anteriores muestran algunos patrones generales que arroja la encuesta CDH que ameritarían investigaciones más detalladas. Es de esperar que estas indagaciones contribuyan a futuros testeos de hipótesis sobre las distintas dinámicas de participación de las personas con grado de doctor en el mercado laboral chileno y su contribución o no al desarrollo de una masa crítica de personas impulsando procesos de investigación, desarrollo e innovación en el país. En la sección siguiente se revisan algunos patrones adicionales entorno a la realización de postdoctorados.

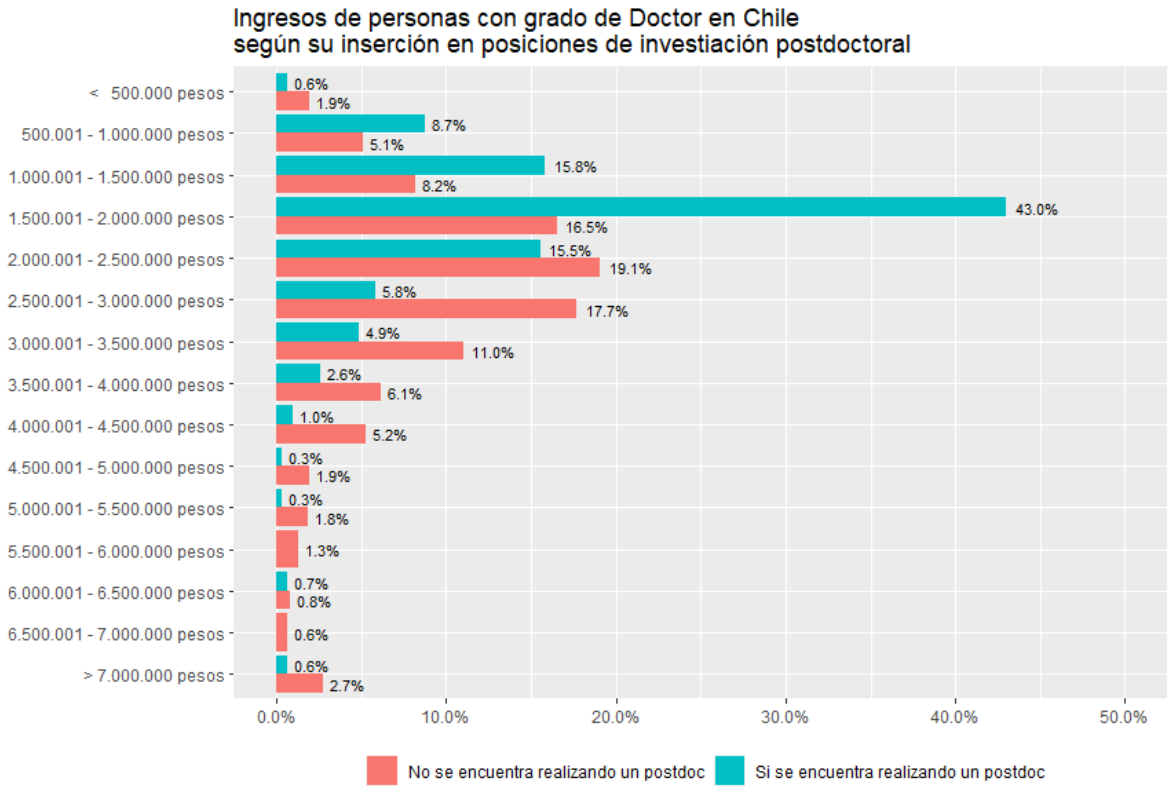
### 7.1.4 Similitudes y diferencias entre las personas que realizaron o no realizaron postdoctorados

Como se indicó en las etapas previas de la elaboración del cuestionario, existen distintas aproximaciones en los encuestados a la hora de concebir los postdoctorados. Para algunos es una experiencia de trabajo más y para otros es una etapa formativa en sí misma y, por lo tanto, aunque no es conducente a la obtención de un grado podría ser percibido como el último escalafón en el proceso de inversión en capital humano.

Lo relevante de este punto es que ello podría estar incidiendo, primero, en el deseo o no de desarrollar postdoctorados mostrando diferencias entre distintas cohortes etarias y, segundo, impactando en las remuneraciones que reciben las personas que han obtenido recientemente su título de doctor.

Los datos son consistentes con estas ideas generales. Primero, como se puede ver en el gráfico siguiente, las personas que están realizando postdoctorados señalan tener ingresos mensuales menores que las que no lo realizaron.

*Figura 22: Distribución por tramos de ingresos brutos mensuales en Chile según situación postdoctoral personas con grado de doctor en Chile 2019*




Consideramos que existen dos líneas de interpretación en cierto modo complementarias para estos patrones. Una primera línea explicativa puede ser que el trabajo de postdoctorado se ha convertido en los hechos en una forma de inserción laboral que ha bajado los costos de contratación de personas con grado de doctor. Y una segunda línea se focaliza en que la práctica de realizar postdoctorados es relativamente reciente y, por ende, no estuvo disponible a gran escala entre personas cuyos títulos fueron obtenidos hace más de una década. Y, por ende, el haber hecho un posdoctorado podría estar vinculado a tener menor edad, experiencia y, en consecuencia, estar en una etapa en la que los rangos de remuneraciones son aún bajos en relación al valor esperado respecto del conjunto de las personas con grado de doctor.

Para visualizar si la edad cumple un rol en este punto, en la siguiente tabla presentamos la distribución de las personas según rango de edad y realización o no de postdoctorados. Como puede verse, dos de cada tres personas que están realizando un postdoctorado es menor de 39 años. En

cambio, en el grupo de personas que no están realizando postdoctorado, las distribuciones por rango etario son menos concentradas.

*Tabla 45: Distribución de rango de edades e inserción postdoctoral*

| Rango de edad | Está realizando postdoc | | Está realizando postdoc | |
|---|---|---|---|---|
| | Número | Porcentaje | Número | Porcentaje |
| < 35 años | 712 | 29,7% | 1.501 | 9,4% |
| 35-44 años | 1.236 | 51,5% | 6.446 | 40,4% |
| 45-54 años | 263 | 11,0% | 4.406 | 27,6% |
| 55-64 años | 166 | 6,9% | 2.788 | 17,5% |
| 65-69 años | 23 | 0,9% | 812 | 5,1% |
| Total | 2.399 | 100,0% | 15.953 | 100,0% |

Por otro lado, existe también factores de oferta que inciden en la realización de posdoctorados. Primero, como puede verse en la tabla siguiente, la participación en postdoctorados se concentra mayormente en algunas disciplinas. El 51,5% de las personas que están en un posdoctorado están vinculadas al área de las ciencias naturales. En cambio, sólo el 34,8% de las personas que no realizaron un postdoctorado están vinculadas a esas áreas del conocimiento.

*Tabla 46: Distribución de áreas del conocimiento e inserción postdoctoral*

| Campo de ciencia y tecnología | Está realizando postdoc | | Está realizando postdoc | |
|---|---|---|---|---|
| | Número | Porcentaje | Número | Porcentaje |
| Ciencias naturales | 1.237 | 51,5% | 5.553 | 34,8% |
| Ingeniería y tecnología | 301 | 12,5% | 2.611 | 16,4% |
| Ciencias médicas y de la salud | 225 | 9,4% | 1.227 | 7,7% |
| Ciencias agrícolas y veterinarias | 127 | 5,3% | 946 | 5,9% |
| Ciencias sociales | 284 | 11,8% | 3.679 | 23,1% |
| Humanidades | 225 | 9,4% | 1.937 | 12,1% |
| Total | 2.399 | 100,0% | 15.953 | 100,0% |

Y segundo, hay diferencias relevantes en el acceso a posdoctorados según zonas geográficas. Específicamente, la mayoría de las personas que se encuentran en postdoctorados se ubican en la Región Metropolitana (52,8%) mientras que dicha proporción entre los que no están haciendo un postdoctorado cae a 48,8%

*Tabla 47: Distribución de zona de residencia e inserción postdoctoral*

| Zona | Si realizó postdoc | | No realizó postdoc | |
|---|---|---|---|---|
| | Número | Porcentaje | Número | Porcentaje |
| Región Metropolitana | 1.266 | 52,8% | 7.791 | 48,8% |
| Otras Regiones | 998 | 41,6% | 7.422 | 46,5% |
| Desempleado, Inactivo o empleo principal fuera de Chile | 135 | 5,6% | 740 | 4,6% |
| Total | 2.399 | 100,0% | 15.953 | 100,0% |

Por último, es mayor la proporción de mujeres que declaran estar realizando un postdoctorado al 1 de diciembre de 2019. Específicamente, un 15,3% versus un 11,9% de los hombres. Una posible consecuencia que es recomendable evaluar con más detalle en futuras

investigaciones es que estos resultados abren una hipótesis según la que la realización de postdoctorados podría no estar necesariamente potenciando la obtención de mayores ingresos, sino más bien siendo un medio de inserción laboral que baja costos de acceso.

### 7.1.5 Similitudes y diferencias relevantes entre zonas geográficas

A nivel geográfico, al agrupar las respuestas según residencia en la Región Metropolitana versus otra región del país, se pueden observar algunas diferencias relevantes por género y por tipo y variedad de la inserción laboral.

En cuanto al género, se puede observar una mayor proporción relativa de mujeres con grado de doctor en la Región Metropolitana (39%) que en otras regiones (33,3%).[22]

*Tabla 48: Distribución de género por zona de residencia*

| Género | Región Metropolitana | | Otras Regiones | | Desempleado, Inactivo o empleo principal fuera de Chile | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Masculino | 5.553 | 61,3% | 5.644 | 67,0% | 400 | |
| Femenino | 3.496 | 38,6% | 2.768 | 32,9% | 475 | |
| Otro | 8 | 0,1% | 8 | 0,1% | 0 | |
| Total | 9.057 | 100,0% | 8.420 | 100,0% | 875 | |

Por otro lado, existe una mayor variedad de opciones de inserción laboral en la Región Metropolitana que puede constatarse en las siguientes tablas. Primero, existe un porcentaje mayor de personas en la capital vis a vis otras regiones que optan por cargos de administración pública (4,7% versus 2,9%) y empresarial (9,8% versus 5%).[23]

*Tabla 49: Distribución de sector al que pertenece la institución donde trabaja por zona de residencia*

| Sector al que corresponde su empleo principal | Región Metropolitana | | Otras Regiones | | Desempleado, Inactivo o empleo principal fuera de Chile | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Empresarial | 887 | 9,8% | 421 | 5,0% | 0 | 0,0% |
| Administración pública | 427 | 4,7% | 248 | 2,9% | 0 | 0,0% |
| Educación Superior | 7.368 | 81,4% | 7.315 | 86,9% | 53 | 6,0% |
| Otro sector de educación | 67 | 0,7% | 113 | 1,3% | 8 | 0,9% |
| Institución privada sin fines de lucro | 307 | 3,4% | 322 | 3,8% | 8 | 0,9% |
| Desempleado o Inactivo | 0 | 0,0% | 0 | 0,0% | 807 | 92,2% |
| Total | 9.057 | 100,0% | 8.420 | 100,0% | 875 | 100,0% |

22 Esta diferencia es estadísticamente significativa con un valor $p < 0.001$.
23 Esta diferencia también es estadísticamente significativa con un valor $p < 0.001$.

Por otro lado, hay también una mayor proporción, estadísticamente significativa de personas en dirección y gerencia en la Región Metropolitana que en otras regiones (9,7% versus 7,1%, respectivamente).[24] Por último, considerando que la gran mayoría de las personas con grado de doctor están vinculadas a instituciones de educación superior, existe en la Región Metropolitana una mayor proporción en universidades fuera del CRUCH (20,5% versus 8,0%).[25]

*Tabla 50: Distribución del tipo de ocupación por zona de residencia*

| Ocupación | Región Metropolitana | | Otras Regiones | | Desempleado, Inactivo o empleo principal fuera de Chile | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Directores y gerentes | 878 | 9,7% | 602 | 7,1% | 0 | 0,0% |
| Profesionales científicos e intelectuales | 7.947 | 87,7% | 7.713 | 91,6% | 68 | 7,8% |
| Técnicos y profesionales de nivel medio | 113 | 1,2% | 82 | 1,0% | 0 | 0,0% |
| Personal de apoyo administrativo | 67 | 0,7% | 8 | 0,1% | 0 | 0,0% |
| Trabajadores de los servicios y vendedores de comercios y mercados | 7 | 0,1% | 0 | 0,0% | 0 | 0,0% |
| Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros | 0 | 0,0% | 8 | 0,1% | 0 | 0,0% |
| Operadores de instalaciones y máquinas y ensambladores | 7 | 0,1% | 0 | 0,0% | 0 | 0,0% |
| Ocupaciones elementales | 22 | 0,2% | 7 | 0,1% | 0 | 0,0% |
| Ocupaciones militares | 15 | 0,2% | 0 | 0,0% | 0 | 0,0% |
| Desempleado o Inactivo | 0 | 0,0% | 0 | 0,0% | 807 | 92,2% |
| Total | 9.057 | 100,0% | 8.420 | 100,0% | 875 | 100,0% |

*Tabla 51: Tipo de institución de educación superior donde trabaja por zona de residencia*

| Tipo de institución de educación superior | Región Metropolitana | | Otras Regiones | | Desempleado, Inactivo o empleo principal fuera de Chile | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Universidad pública o tradicional (CRUCH) | 5.439 | 60,1% | 6.586 | 78,2% | 15 | 1,7% |
| Universidad no perteneciente al CRUCH | 1.854 | 20,5% | 677 | 8,0% | 15 | 1,7% |
| Instituto Profesional | 23 | 0,2% | 15 | 0,2% | 0 | 0,0% |
| Centro de formación técnica | 0 | 0,0% | 7 | 0,1% | 0 | 0,0% |
| Otro | 52 | 0,6% | 30 | 0,4% | 23 | 2,6% |

24 Con un valor $p < 0.001$.

25 Esta diferencia es estadísticamente significativa con un valor $p < 0.001$.

| | | | | | | |
|---|---|---|---|---|---|---|
| Desempleado, Inactivo o no trabaja en Educación Superior | 1.689 | 18,6% | 1.104 | 13,1% | 823 | 94,0% |
| Total | 9.057 | 100,0% | 8.420 | 100,0% | 875 | 100,0% |

Todo lo anterior incide en ingresos distribuidos hacia tramos superiores en la Región Metropolitana en comparación a las otras regiones. Siguiendo el mismo procedimiento de imputación de ingresos usado para los géneros, el ingreso bruto promedio en la región metropolitana se ubica en 2,64 millones mensuales y en 2,22 millones para las otras regiones.[26]

*Tabla 52: Rango de remuneraciones por zona de residencia*

| Zona | Ingreso anual bruto (en miles de pesos) | Ingreso mensual bruto (en pesos) |
|---|---|---|
| Otras regiones | 26.673 | 2.222.742 |
| Metropolitana | 31.758 | 2.646.527 |
| Total ponderado | 29.308 | 2.442.361 |

En suma, en concordancia con los objetivos del levantamiento de información encargado en la presente licitación, la encuesta CDH contiene una rica variedad de información relevante para caracterizar a la población residente en Chile con grado de doctor. En estas líneas hemos querido destacar algunos de esos resultados y motivar algunas hipótesis plausibles sobre dinámicas que podrían estar a la base de los patrones observables.

El análisis de estos datos y su comparación con los resultados reportados en las versiones anteriores de la encuesta y con los generados en los otros países donde se realiza el levantamiento constituyen un insumo relevante para el diseño de políticas públicas dirigidas al fomento de la investigación, el desarrollo y la innovación.

# 8. Aprendizajes y recomendaciones para futuros levantamientos

## 8.1 Aprendizajes sobre la población objetivo

- **Disponibilidad y organización del tiempo:**

El ejercicio profesional de los doctores y doctoras a menudo requiere la coordinación un elevado número de actividades diversas. En el contexto de la crisis sanitaria por SARS-CoV-2 esta realidad se ha intensificado debido a las medidas de distanciamiento social y teletrabajo. Para futuros estudios se recomienda explorar a priori formas de adaptarse a esta realidad e incluir la organización del tiempo dentro de los parámetros del levantamiento.

26 Esta diferencia entre ingresos agrupados en estas dos zonas geográficas es estadísticamente significativa con un valor $p < 0.001$.

- **Disposición e Interés en participar:**

Durante el levantamiento llamó la atención del equipo de terreno a cargo del levantamiento CATI la disposición favorable de los doctores y las doctoras a participar del estudio. En comparación con otros estudios, el interés mostrado por los doctores y doctoras se tradujo en una tasa de respuesta considerablemente mayor a lo que – en experiencia de Datavoz – es una tasa de respuesta habitual para estudios de este tipo (en porcentajes que oscilan entre 10 y 15%), y que se observó de forma independiente en ambos medios de recolección. Se reportó también la predominancia de un trato amable y cordial por parte de las personas encuestadas, lo que facilitó en gran medida el levantamiento CATI.

## 8.2 Aprendizajes sobre los medios de recolección

### 8.2.1 Aprendizajes de levantamiento CAWI

- **Invitaciones y Recordatorios**

El envío de recordatorios periódicos tiene un efecto positivo en la recolección de datos. La mayor cantidad de respuestas se obtiene el mismo día del envío o el siguiente, al tercer día la cantidad de respuestas cae significativamente. Para futuros levantamientos se recomienda enviar recordatorios con una frecuencia de dos a tres días.

- **Flexibilidad y organización del tiempo**

La duración de las encuestas recolectadas mediante CAWI muestran una alta dispersión lo que sugiere heterogeneidad en la disponibilidad de tiempo. Algunas personas contestan rápidamente, en tiempos menores a media hora, mientras otras pueden tardar hasta un par de días o más. Esto se debe a la capacidad de la plataforma de almacenar los avances parciales de la encuesta permitiendo retomar más adelante.

### 8.2.2 Aprendizajes de levantamiento CAWI

- **Agendamiento y planificación**

La organización del tiempo es un aspecto muy importante al trabajar con la población de doctoras y doctores. La baja disponibilidad de ratos libres es probablemente el mayor obstáculo para la realización de las encuestas telefónicas de forma espontánea. El agendamiento del próximo llamado, acordando la fecha y la hora con el encuestado, tuvo mejores resultados en comparación con la gestión automática de las llamadas de forma aleatoria.

- **Confianza y autorización de datos**

El acompañamiento por parte del encuestador o encuestadora tiene beneficios como mantener la atención del encuestado en la tarea de responder la encuesta, reduce la cantidad respuestas erróneas, pero además genera un vínculo de confianza en el que el encuestado puede expresar sus dudas y opiniones. Esto se ve reflejado en un tasa de autorización de datos del 90% en comparación con el 72% logrado en encuestas autoaplicadas.

### 8.2.3 Aprendizajes de la coordinación entre CAWI y CATI

Como se comentó en la sección 4.4.1 las modificaciones realizadas a los protocolos tomaron un foco más centrado en el encuestado abriendo las oportunidades para que los doctores y las doctoras pudieran escoger más fácilmente el medio por el cual quisieran contestar. Inevitablemente a raíz de estas medidas que admitían mayor flexibilidad se produjeron situaciones puntuales como el contacto a personas que ya hubieran respondido la encuesta, o recordatorios a personas que ya tenían agendada una entrevista. Para estos casos se respondió vía web agradeciendo el interés y su participación, y ofreciendo las respectivas disculpas. La respuesta fue positiva.

## 8.3 Recomendaciones para futuros levantamientos

Uno de los aprendizajes logrados que tiene mayor relevancia para futuros levantamientos dirigidos a la población de personas con grado de doctor es que una oferta mixta de medios de recolección que combine levantamiento telefónico y web, tiene un gran potencial de adaptarse a las necesidades de flexibilidad horaria de las personas encuestadas.

Sin embargo, aprovechar esta modalidad requiere de un gran esfuerzo de coordinación e infraestructura para el desarrollo del levantamiento de forma armoniosa. Se recomienda explorar soluciones de plataformas que integren ambos medios de recolección en una misma plataforma que centralice la base de datos de contacto, facilite las medidas de gestión operativa y coordine los esfuerzos de la central telefónica y plataforma online.

De forma alternativa, mientras no exista un producto comercial capaz prestar esas utilidades se recomienda el desarrollo de un modelo organizacional capaz de enfrentar el desafío de coordinación entre múltiples plataformas.

# 9. ANEXO 1 Cuestionario Aplicado

El cuestionario aplicado se encuentra publicado en la web del Ministerio de Ciencia.

# 10. ANEXO 2 Cálculo de indicadores AAPOR

*Fórmula de indicadores de respuesta, cooperación, rechazo y contacto, según AAPOR*

| Nombre de indicador | | *Fórmula* |
|---|---|---|
| *Tasa de respuesta* | Indicador 1 | $\frac{I}{(I+P)\ +\ (R+NC+O)}$ |
| | Indicador 2 | $\frac{(I+P)}{(I+P)\ +\ (R+NC+O)}$ |
| *Tasa de cooperación* | Indicador 1 | $\frac{I}{(I+P)+R+O}$ |
| | Indicador 2 | $\frac{(I+P)}{((I+P)+R+O))}$ |
| *Tasa de rechazo* | Indicador 1 | $\frac{R}{((I+P)+(R+NC+O))}$ |
| *Tasa de contacto* | Indicador 1 | $\frac{(I+P)+R+O}{(I+P)+R+O+NC}$ |
| | Indicador 2 | $\frac{(I+P)+R+O}{(\mathrm{I}+\mathrm{P})+\mathrm{R}+\mathrm{O}+\mathrm{NC}}$ |

# 11. ANEXO 3 Modelo de carta de invitación

Santiago, 7 de agosto de 2020

Estimado(a),

El Ministerio de Ciencia, Tecnología, Conocimiento e Innovación se encuentra realizando la tercera versión de la encuesta "TRAYECTORIA DE PROFESIONALES CON GRADO DE DOCTOR EN CHILE (CDH 2019)". El objetivo de esta encuesta es recopilar y analizar información respecto a la trayectoria laboral y profesional de las personas con grado de doctor (PhD) residentes en Chile. La idea es que esta información sirva como un insumo para el diseño y evaluación de estrategias de políticas públicas en estas materias.

Su realización se hace en concordancia con el proyecto internacional que analiza las trayectorias de profesionales con grado de doctor "*Career of Doctorate Holders (CDH)*" de la Dirección de Ciencia, Tecnología e Industria de la OCDE, el Instituto de Estadísticas de la UNESCO y la Oficina de Estadísticas de la Comisión Europea (EUROSTAT).

Dado lo anterior, la empresa Datavoz ha obtenido la licitación pública ID 1098710-1-LE20 para el levantamiento de la encuesta. En esta primera etapa, se está realizando una prueba piloto para la validación del cuestionario que será utilizado.

Al respecto, solicitamos su colaboración en el estudio contestando esta encuesta por vía web, la cual se encuentra disponible a través del enlace proporcionado en el correo adjunto. Ésta tiene una duración aproximada de 20 minutos.

Cabe señalar que los datos que se recopilen por medio de esta consulta serán de uso exclusivo del Ministerio en el marco del estudio anteriormente citado.

Para mayor información, puede contactar a Diego Armas (darmas@minciencia.gob.cl) o Heinz Doebbel (hdoebbel@minciencia.gob.cl), profesionales de la Subsecretaría de Ciencia, Tecnología, Conocimiento e Innovación.

Sin otro particular, y agradeciendo su valiosa colaboración, se despide atentamente,

María José Bravo Pizarro
Jefa División Estudios y Estadísticas
Subsecretaría de Ciencia, Tecnología, Conocimiento e Innovación